# Polar codes for empirical coordination over noisy channels with strictly causal encoding

Giulia Cervia, Laura Luzzi, Mael Le Treust, Matthieu R Bloch

# Polar codes for empirical coordination over noisy channels with strictly causal encoding

Giulia CERVIA[1], Laura LUZZI[1], Maël LE TREUST[1], Matthieu R. BLOCH[2]

[1]ETIS UMR 8051, Université Paris Seine, Université Cergy-Pontoise, ENSEA, CNRS, Cergy, France.

[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia

{giulia.cervia, laura.luzzi, mael.le-treust}@ensea.fr, matthieu.bloch@ece.gatech.edu

**Résumé –** Dans ce travail, nous proposons un schéma de codage basé sur les codes polaires pour la coordination empirique d'appareils autonomes. Nous considérons un réseau simple composé de deux nœuds reliés par un lien bruité, et nous cherchons à coordonner les signaux en entrée et en sortie du canal, avec la source et sa reconstruction. Lorsque l'encodeur est strictement causal, nous montrons que les codes polaires atteignent la région optimale de coordination empirique, à condition que les deux nœuds partagent une source aléatoire, dont le débit est asymptotiquement négligeable.

**Abstract –** In this paper, we propose a coding scheme based on polar codes for empirical coordination of autonomous devices. We consider a two-node network with a noisy link in which the input and output signals have to be coordinated with the source and the reconstruction. In the case of strictly causal encoding, we show that polar codes achieve the empirical coordination region, provided that a vanishing rate of common randomness is available.

## 1 Introduction

In decentralized networks of connected objects, such as wireless sensors, medical and wearable devices, smart energy meters, home appliances, and self-driving cars, devices sense their environment and choose their actions in order to achieve a general objective. It is essential that these devices, considered as autonomous decision-makers, cooperate and coordinate their actions to induce a global behavior, represented by a utility function to be maximized.

Within the framework of information theory, two different metrics have been proposed to measure the level of coordination : *empirical coordination* requires the joint histogram of the actions to approach a target distribution, while *strong coordination* requires the joint distribution of actions to converge in total variation to an i.i.d. target distribution [1].

We consider a two-node network with an information source and a noisy channel in which the input and output signals should be empirically coordinated with the source and the reconstruction. In [2], the authors provide a characterization of the *coordination region* when the encoder is strictly causal. Inspired by the binning technique using polar codes in [3], we propose an explicit coding scheme that achieves a subset of the coordination region in [2] by turning the argument of [4] into an explicit polar coding proof. The scenario in which both the encoder and the decoder are non-causal has already been considered for empirical coordination with polar codes [5]. Here, we focus on the setting in which the encoder is strictly causal.

In this paper, we only achieve a subset of the coordination region because of the use of binary polar codes, but the whole region can be achieved using non-binary polar codes.

The remainder of the paper is organized as follows. Section 2 introduces the notation, describes the model under investigation and states the main achievability result. Section 3 details the proposed coordination scheme using polar codes. Finally, Section 4 proves the main result.

## 2 Problem statement

### 2.1 Notation

We define $[a, b]$ as the set of the integers between $a$ and $b$. For $n = 2^m$, $m \in \mathbb{N}$, we note $G_n := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m}$ the source polarization transform defined in [6]. Given $X^{1:n} := (X^1, \dots, X^n)$ a random vector, we note $X^{1:j}$ the first $j$ components of $X^{1:n}$ and $X[A]$, where $A \subset [1, n]$, the components $X^j$ such that $j \in A$. We note $\mathbb{V}(\cdot, \cdot)$ and $\mathbb{D}(\cdot \| \cdot)$ the variational distance and the Kullback-Leibler divergence between two distributions, respectively.

### 2.2 System model and main result

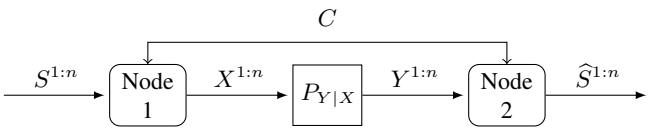


FIGURE 1 – Coordination of signals and actions for a two-node network with a noisy channel.

We consider two agents, Node 1 and Node 2, who have access to a shared randomness source $C \in \mathcal{C}_n$ (Figure 1). Node 1 observes an i.i.d. sequence of actions $S^{1:n} \in \mathcal{S}^n$ with discrete probability distribution $P_S$. Node 1 then selects a signal $X^{1:n}$ such that $X^i = f_i(S^{1:i-1}, C)$, where $f^n = \{f_i\}_{i=1}^n$, $f_i : \mathcal{S}^{i-1} \times \mathcal{C}_n \to \mathcal{X}$ is the strictly causal encoder. The signal $X^{1:n}$ is transmitted over a discrete memoryless channel with transition probability $P_{Y|X}$. Upon receiving $Y^{1:n} \in \mathcal{Y}^n$, Node 2 selects an action $\widehat{S}^{1:n} = g^n(Y^{1:n}, C)$, where $g^n : \mathcal{Y}^n \times \mathcal{C}_n \to \widehat{\mathcal{S}}^n$ is the non-causal decoder. For block length $n$, the pair $(f^n, g^n)$ constitutes a code. Node 1 and Node 2 wish to coordinate in order to obtain a joint distribution of actions and signals that is close to a target distribution $P_{SXY\widehat{S}}$. We focus on the empirical coordination metric defined in [1].

**Definition 1.** *A distribution $P_{SXY\widehat{S}}$ is achievable if for all $\varepsilon > 0$ there exists a sequence of codes $\{(f^n, g^n)\}_{n\in\mathbb{N}}$ such that*

$$\lim_{n\to\infty} \mathbb{P}\left\{\mathbb{V}\left(T_{S^{1:n}X^{1:n}Y^{1:n}\widehat{S}^{1:n}}, P_{SXY\widehat{S}}\right) > \varepsilon\right\} = 0,$$

*where $T_{S^{1:n}X^{1:n}Y^{1:n}\widehat{S}^{1:n}}(s, x, y, \hat{s})$ is the empirical distribution of the tuple $(S^{1:n}, X^{1:n}, Y^{1:n}, \widehat{S}^{1:n})$ induced by the code.*

*The empirical coordination region $\mathcal{C}$ is the set of achievable distributions $P_{SXY\widehat{S}}$.*

**Theorem 2** (Strictly causal encoder [2])**.** *Let $P_S$ and $P_{Y|X}$ be the given source and channel parameters. When the encoder is strictly causal, the coordination region $\mathcal{C}$ is given by*

$$\mathcal{C} := \left\{ \begin{array}{c} P_{SXY\widehat{S}} : \; P_{SXY\widehat{S}} = P_S P_X P_{Y|X} P_{\widehat{S}|SXY} \\ \exists\, U \text{ taking values in } \mathcal{U} \\ P_{SXYU\widehat{S}} = P_S P_X P_{U|XS} P_{Y|X} P_{\widehat{S}|UY} \\ I(X, U; S) \le I(X, U; Y) \\ |\mathcal{U}| \le |\mathcal{S}||\mathcal{X}||\mathcal{Y}||\widehat{\mathcal{S}}| + 1 \end{array} \right\} \quad (1)$$

**Remark 3.** *By the chain rule, we have*

- $I(X, U; S) = I(U; S|X) + I(X; S) = I(U; S|X)$ *since $S^{1:n}$ and $X^{1:n}$ are independent ;*
- $I(X, U; Y) = I(U; Y|X) + I(X; Y) = I(X; Y)$ *because of the Markov chain $U - X - Y$.*

*Hence the condition $I(X, U; S) \le I(X, U; Y)$ in* (1) *becomes $I(U; S|X) \le I(X; Y)$.*

**Theorem 4.** *For all $P_{SXY\widehat{S}} \in \mathcal{C}$ such that $\mathcal{U} = \{0, 1\}$, there exists an explicit polar coding scheme that achieves empirical coordination with vanishing rate of common randomness.*

**Remark 5.** *Since $U$ is binary we only achieve a subset of $\mathcal{C}$. The proof can be generalized to the case where $|\mathcal{U}|$ is a prime number using non-binary polar codes.*

# 3 Polar coding scheme

Consider the random vectors $S^{1:n}$, $U^{1:n}$, $X^{1:n}$, $Y^{1:n}$ and $\widehat{S}^{1:n}$ generated i.i.d. according to $P_{SXUY\widehat{S}}$ that factorize as in (1) with the same mutual information and cardinality constraints.

**Polarize $X$** Let $Z^{1:n} = X^{1:n}G_n$ be the polarization of $X^{1:n}$, where $G_n$ is the source polarization transform. For some $0 < \beta < 1/2$, let $\delta_n := 2^{-n^\beta}$ and define the very high and high entropy sets :

$$\begin{aligned} \mathcal{V}_X &:= \left\{j \in [1, n] : H(Z^j|Z^{1:j-1}) > 1 - \delta_n\right\}, \\ \mathcal{H}_X &:= \left\{j \in [1, n] : H(Z^j|Z^{1:j-1}) > \delta_n\right\}, \\ \mathcal{H}_{X|Y} &:= \left\{j \in [1, n] : H(Z^j|Z^{1:j-1}Y^{1:n}) > \delta_n\right\}. \end{aligned} \quad (2)$$

Partition the set $[1, n]$ into four disjoint sets :

$$A_1 := \mathcal{V}_X \cap \mathcal{H}_{X|Y}, \quad A_2 := \mathcal{V}_X \cap \mathcal{H}^c_{X|Y},$$
$$A_3 := \mathcal{V}^c_X \cap \mathcal{H}_{X|Y}, \quad A_4 := \mathcal{V}^c_X \cap \mathcal{H}^c_{X|Y}.$$

**Remark 6.** *We have :*

- $\mathcal{V}_X \subset \mathcal{H}_X$ *and* $\lim_{n\to\infty} \dfrac{|\mathcal{H}_X \setminus \mathcal{V}_X|}{n} = 0$ *[6],*
- $A_1 \cup A_2 = \mathcal{V}_X$ *and* $\lim_{n\to\infty} \dfrac{|\mathcal{V}_X|}{n} = H(X)$ *[7],*
- $A_1 \cup A_3 = \mathcal{H}_{X|Y}$ *and* $\lim_{n\to\infty} \dfrac{|\mathcal{H}_{X|Y}|}{n} = H(X|Y)$ *[6].*

*Since* $\lim_{n\to\infty} \dfrac{|A_2| - |A_3|}{n} = H(X) - H(X|Y) = I(X; Y) \ge 0$ *this implies directly that for $n$ large enough $|A_2| \ge |A_3|$.*

**Polarize $U$** Let $V^{1:n} = U^{1:n}G_n$ be the polarization of $U^{1:n}$ and define :

$$\begin{aligned} \mathcal{V}_{U|XS} &:= \left\{j \in [1, n] : H(V^j|V^{1:j-1}X^{1:n}S^{1:n}) > 1 - \delta_n\right\}, \\ \mathcal{H}_{U|XS} &:= \left\{j \in [1, n] : H(V^j|V^{1:j-1}X^{1:n}S^{1:n}) > \delta_n\right\}, \\ \mathcal{H}_{U|X} &:= \left\{j \in [1, n] : H(V^j|V^{1:j-1}X^{1:n}) > \delta_n\right\}. \end{aligned} \quad (3)$$

Partition the set $[1, n]$ into four disjoint sets :

$$B_1 := \mathcal{V}_{U|XS} \cap \mathcal{H}_{U|X} = \mathcal{V}_{U|XS}, \quad B_2 := \mathcal{V}_{U|XS} \cap \mathcal{H}^c_{U|X} = \emptyset,$$
$$B_3 := \mathcal{V}^c_{U|XS} \cap \mathcal{H}_{U|X}, \quad B_4 := \mathcal{V}^c_{U|XS} \cap \mathcal{H}^c_{U|X} = \mathcal{H}^c_{U|X}.$$

**Remark 7.** *We have :*

- $\mathcal{V}_{U|XS} \subset \mathcal{H}_{U|XS}$ *and* $\lim_{n\to\infty} \dfrac{|\mathcal{H}_{U|XS} \setminus \mathcal{V}_{U|XS}|}{n} = 0$ *[6],*
- $B_1 = \mathcal{V}_{U|XS}$ *and* $\lim_{n\to\infty} \dfrac{|\mathcal{V}_{U|XS}|}{n} = H(U|XS)$ *[7],*
- $B_4 = \mathcal{H}^c_{U|X}$ *and* $\lim_{n\to\infty} \dfrac{|\mathcal{V}^c_{U|X}|}{n} = 1 - H(U|X)$ *[7],*
- $B_3 \cup B_4 = \mathcal{V}^c_{U|XS}$ *and* $\lim_{n\to\infty} \dfrac{|\mathcal{V}^c_{U|XS}|}{n} = 1 - H(U|XS)$ *[7].*

*Note that $H(U|X) - H(U|XS) = I(X, U; S) \ge 0$ and $|B_3|/n$ tends to $I(X, U; S)$. Since $I(X, U; Y) = I(X; Y)$, the inequality $I(X, U; S) \le I(X, U; Y)$ implies directly that for $n$ large enough $|B_3| \le |A_2| - |A_3|$.*

**Encoding** We use a chaining construction over multiple blocks. The encoder observes $(S_0^{1:n}, S_1^{1:n}, \ldots, S_k^{1:n})$, where $S_0^{1:n}$ is a uniform random sequence and $S_i^{1:n}$ for $i \in [1, k]$ are $k$ blocks of the source. It then generates for each block $i \in [1, k]$ random variables $\widetilde{Z}_i^{1:n}$ and $\widetilde{V}_i^{1:n}$ following the procedure described in Algorithm 1. In particular, the chaining construction proceeds as follows. The bits in $A_1 \subset \mathcal{V}_X$ and $B_1 \subset \mathcal{V}_{U|XS}$ are chosen with uniform probability using

**Algorithm 1:** Encoding algorithm at Node 1

**Input** : $(S_0^{1:n}, \dots, S_k^{1:n})$, local randomness (uniform random bits) $M$ and common randomness $C = (C_1, K_1, C_2, K_2)$ shared with Node 2 :
- $C_1$ of size $|A_1|$ and $K_1$ of size $|A_3|$ ;
- $C_2$ of size $|B_1|$ and $K_2$ of size $|B_3|$.

**Output**: $(\widetilde{Z}_1^{1:n}, \dots, \widetilde{Z}_k^{1:n})$, $(\widetilde{V}_1^{1:n}, \dots, \widetilde{V}_k^{1:n})$

**if** $i = 1$ **then**

$\widetilde{Z}_1[A_1] \longleftarrow C_1 \qquad \widetilde{Z}_1[A_2] \longleftarrow M$

**for** $j \in A_3 \cup A_4$ **do**

Successively draw the bits $\widetilde{Z}_i^j$ according to

$$P_{Z^j|Z^{i:j-1}}\left(\widetilde{Z}_i^j \mid \widetilde{Z}_i^{i:j-1}\right) \tag{4}$$

$\widetilde{V}_1[B_1] \longleftarrow C_2$

**for** $j \in B_3 \cup B_4$ **do**

Given $S_1^{1:n}$, successively draw the bits $\widetilde{V}_1^j$ according to

$$P_{V^j|V^{1:j-1}X^{1:n}S^{1:n}}\left(\widetilde{V}_i^j \mid \widetilde{V}_i^{1:j-1}\widetilde{X}_i^n S_{i-1}^n\right) \tag{5}$$

**for** $i = 2, \dots, k$ **do**

$\widetilde{Z}_i[A_1] \longleftarrow C_1 \quad \widetilde{Z}_i[A_2'] \longleftarrow M$

$\widetilde{Z}_i[B_3'] \longleftarrow \widetilde{V}_{i-1}[B_3] \oplus K_2 \quad \widetilde{Z}_i[A_3'] \longleftarrow \widetilde{Z}_{i-1}[A_3] \oplus K_1$

**for** $j \in A_3 \cup A_4$ **do**

Successively draw the bits $\widetilde{Z}_i^j$ according to (4)

$\widetilde{V}_i[B_1] \longleftarrow C_2$

**for** $j \in B_3 \cup B_4$ **do**

Successively draw the bits $\widetilde{V}_i^j$ according to (5).

uniform randomness sources $(C_1, C_2)$ shared with Node 2, and their value is reused over all blocks. In the first block the bits in $A_2 \subset \mathcal{V}_X$ are chosen with uniform probability using a local randomness source $M$. The bits in $A_3 \cup A_4$ and $B_3 \cup B_4$ are generated according to the previous bits using successive cancellation encoding [6]. Note that it is possible to sample efficiently from $P_{Z^j|Z^{1:j-1}}$ and $P_{V^j|V^{1:j-1}X^{1:n}S^{1:n}}$ (given $S^{1:n}$ and $X^{1:n}$) respectively [6].

From the second block, let $A_3'$ and $B_3'$ be two disjoint subsets of $A_2$ such that $|A_3'| = |A_3|$ and $|B_3'| = |B_3|$. The existence of those disjoint subsets is guaranteed by Remark 6 and Remark 7. The bits of $A_3$ and $B_3$ in block $i$ are used as $A_3'$ and $B_3'$ in block $i+1$ using one time pads with keys $K_1$ and $K_2$ respectively. Thanks to the Crypto Lemma [8, Lemma 3.1], if we choose $K_1$ of size $|A_3|$ and $K_2$ of size $|B_3|$ to be uniform random keys, the bits in $A_3'$ and $B_3'$ in the block $i+1$ are uniform. The bits in $A_2' := A_2 \setminus (A_3' \cup B_3')$ are chosen with uniform probability using the local randomness source $M$.

The encoder then computes $\widetilde{X}_i^{1:n} = \widetilde{Z}_i^{1:n} G_n$ for $i = 1, \dots, k$ and sends it over the channel. We use an extra $(k+1)$-th block to send a version of $\widetilde{Z}_k[A_3]$ encoded with a good channel code as in [5, Section III.B].

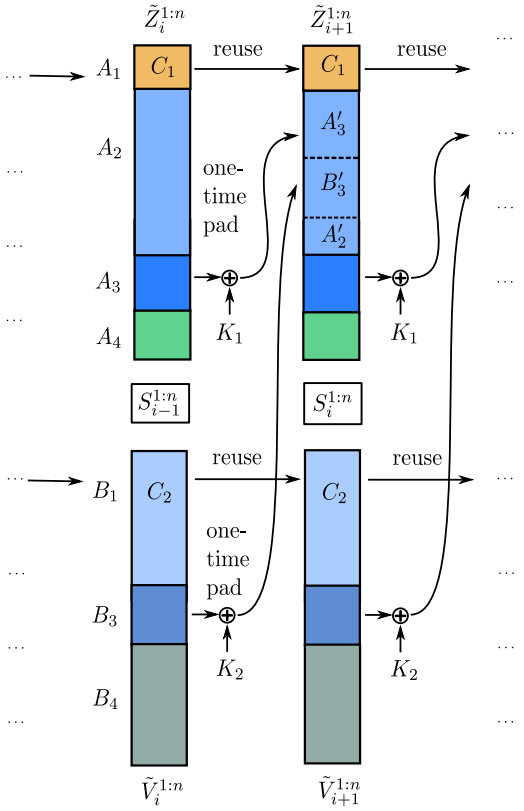


FIGURE 2 – Chaining construction for block Markov encoding

**Algorithm 2:** Decoding algorithm at Node 2

**Input** : $(Y_1^{1:n}, \dots, Y_{k+1}^{1:n})$, $C = (C_1, K_1, C_2, K_2)$ common randomness shared with Node 1

**Output**: $(\widehat{Z}_1^{1:n}, \dots, \widehat{Z}_k^{1:n})$, $(\widehat{V}_1^{1:n}, \dots, \widehat{V}_k^{1:n})$

**for** $i = k, \dots, 1$ **do**

$\widehat{Z}_i[A_1] \longleftarrow C_1 \qquad \widehat{V}_i[B_1] \longleftarrow C_2$

**if** $i = k$ **then**

$\widehat{Z}_k[A_3] \longleftarrow Y_{k+1}^{1:n} \qquad \widehat{V}_k[A_3] \longleftarrow Y_{k+1}^{1:n}$

**else**

$\widehat{Z}_i[A_3] \leftarrow \widehat{Z}_{i+1}[A_3'] \oplus K_1$

$\widehat{V}_i[B_3] \leftarrow \widehat{Z}_{i+1}[B_3'] \oplus K_2$

**for** $j \in A_2 \cup A_4$ **do**

Successively draw the bits according to

$$\widehat{Z}_i^j = \begin{cases} 0 & \text{if } L_n(Y_i^{1:n}, Z_i^{1:j-1}) \geq 1 \\ 1 & \text{else} \end{cases}$$

$$L_n(Y_i^{1:n}, Z_i^{1:j-1}) = \frac{P_{Z_i^j|Z_i^{1:j-1}Y_i^{1:n}}\left(0 \mid \widehat{Z}_i^{1:j-1}Y_i^{1:n}\right)}{P_{Z_i^j|Z_i^{1:j-1}Y_i^{1:n}}\left(1 \mid \widehat{Z}_i^{1:j-1}Y_i^{1:n}\right)}$$

**for** $j \in B_4$ **do**

Successively draw the bits according to

$$\widehat{V}_i^j = \begin{cases} 0 & \text{if } L_n(X_{i+1}^{1:n}, V_i^{1:j-1}) \geq 1 \\ 1 & \text{else} \end{cases}$$

**Decoding** The decoder observes $(Y_1^{1:n}, \dots, Y_{k+1}^{1:n})$ and the $(k+1)$-th block allows it to decode in reverse order. In block $i \in [1,k]$, the decoder has access to $\widehat{Z}_i[A_1 \cup A_3] = \widehat{Z}_i[\mathcal{H}_{X|Y}]$ and $\widehat{V}_i[B_1 \cup B_3] = \widehat{V}_i[\mathcal{H}_{U|X}]$ : the bits in $A_1$ and $B_1$ correspond to shared randomness $(C_1, C_2)$, in block $i \in [1, k-1]$ the bits in $A_3$ and $B_3$ are obtained by successfully recovering $A_2$ in block $i+1$ and in block $k$ they are recovered from $Y_{k+1}^{1:n}$ as in [5, Section III.C]. For each block $i = k, \dots, 1$ the decoder recovers the estimates $\widehat{Z}_i^{1:n}$ and $\widehat{V}_i^{1:n}$ using Algorithm 2. From ${Y_i}^{1:n}$ and $\widehat{Z}_i[A_1 \cup A_3]$ the successive cancellation decoder can retrieve $\widehat{Z}_i[A_2 \cup A_4]$ and therefore $\widehat{V}_i[B_4]$. Note that, as shown in [6, Theorem 3], $\widehat{V}^{1:n}$ is equal to $\widehat{V}^{1:n}$ with high probability. The decoder computes $\widehat{U}_i^{1:n} = \widehat{V}_i^{1:n} G_n$. It then generates $\widehat{S}_i^{1:n}$ symbol by symbol using : $P_{\widehat{S}_i^j | \widehat{U}_i^j Y_i^j}(s|u,y) = P_{\widehat{S}|UY}(s|u,y)$.

**Remark 8.** *The rate of common randomness is negligible, since :*

$$\lim_{\substack{n\to\infty \\ k\to\infty}} \frac{|A_1| + |A_3| + |B_1| + |B_3|}{kn} = \lim_{\substack{n\to\infty \\ k\to\infty}} \frac{|\mathcal{V}_{X|Y}| + |\mathcal{H}_{U|X}|}{nk}$$
$$= \lim_{k\to\infty} \frac{H(X|Y) + H(U|X)}{k} = 0.$$

# 4 Proof of Theorem 4

Given $\varepsilon > 0$, we want to prove that :

$$\lim_{n\to\infty} \mathbb{P}\left\{ \mathbb{V}\left( T_{S_{1:k+1}^{1:n} X_{1:k+1}^{1:n} Y_{1:k+1}^{1:n} \widehat{S}_{1:k+1}^{1:n}}, P_{SXY\widehat{S}} \right) > \varepsilon \right\} = 0.$$

This requires a few steps :
**1.** $\forall i \in [1,k]$, $\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_{S_i^{1:n} \widetilde{X}_i^{1:n} \widetilde{U}_i^{1:n}}, P_{SXU}) > \varepsilon\} = 0$;
**2.** $\forall i \in [1,k]$, $\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_{S_i^{1:n} \widetilde{X}_i^{1:n} \widetilde{U}_i^{1:n} Y_i^{1:n}}, P_{SXUY}) > \varepsilon\} = 0$;
**3.** $\forall i \in [1,k]$, $\lim_{n\to\infty} \mathbb{P}\{\mathbb{V}(T_{S_i^{1:n} \widetilde{X}_i^{1:n} \widetilde{U}_i^{1:n} Y_i^{1:n} \widehat{S}_i^{1:n}}, P_{SXUY\widehat{S}}) > \varepsilon\} = 0$;
**4.** Convergence in each block implies overall convergence ;
**5.** The theorem follows from the fact that

$$\mathbb{V}\left( T_{S_{1:k+1}^{1:n} \widetilde{X}_{1:k+1}^{1:n} Y_{1:k+1}^{1:n} \widehat{S}_{1:k+1}^{1:n}}, P_{SXY\widehat{S}} \right) \leq$$
$$\mathbb{V}(T_{S_{1:k+1}^{1:n} \widetilde{X}_{1:k+1}^{1:n} \widetilde{U}_{1:k+1}^{1:n} Y_{1:k+1}^{1:n} \widehat{S}_{1:k+1}^{1:n}}, P_{SXUY\widehat{S}}).$$

Note that since the steps 2 to 5 have already been proved in [5, Section IV], we only need to prove the first step. For all $\varepsilon_0 > 0$, we define

$$\mathcal{T}_{\varepsilon_0}(P_{SXU}) := \left\{ (\mathbf{s},\mathbf{x},\mathbf{u}) \middle| \mathbb{V}\left(P_{SXU}, T_{(\mathbf{s},\mathbf{x},\mathbf{u})}\right) \leq \varepsilon_0 \right\}$$

Observe that for the i.i.d. distribution, we have $\lim_{n\to\infty} \mathbb{P}\left\{(\mathbf{s},\mathbf{x},\mathbf{u}) \in \mathcal{T}_{\varepsilon_0}(P_{SXU})\right\} = 1$.

Let $i \in [1,k]$, we have :

$$\mathbb{P}\left\{ \mathbb{V}\left( T_{S_i^{1:n} X_i^{1:n} U_i^{1:n}}, P_{SXU} \right) > \varepsilon_0 \right\}$$
$$= \sum_{\mathbf{s},\mathbf{x},\mathbf{u}} P_{S_i^{1:n} \widetilde{X}_i^{1:n} \widetilde{U}_i^{1:n}}(\mathbf{s},\mathbf{x},\mathbf{u}) \mathbb{1}\left\{(\mathbf{s},\mathbf{x},\mathbf{u}) \notin \mathcal{T}_{\varepsilon_0}(P_{SXU})\right\}$$
$$= \sum_{\mathbf{s},\mathbf{x},\mathbf{u}} (P_{S_i^{1:n} \widetilde{X}_i^{1:n} \widetilde{U}_i^{1:n}}(\mathbf{s},\mathbf{x},\mathbf{u}) - P_{S^{1:n} X^{1:n} U^{1:n}}(\mathbf{s},\mathbf{x},\mathbf{u})$$
$$+ P_{S^{1:n} X^{1:n} U^{1:n}}(\mathbf{s},\mathbf{x},\mathbf{u})) \mathbb{1}\left\{(\mathbf{s},\mathbf{x},\mathbf{u}) \notin \mathcal{T}_{\varepsilon_0}(P_{SXU})\right\}$$
$$\leq \mathbb{V}(P_{S^{1:n} \widetilde{X}^{1:n} \widetilde{U}^{1:n}}, P_{S^{1:n} X^{1:n} U^{1:n}}) + \mathbb{P}\{(\mathbf{s},\mathbf{x},\mathbf{u}) \notin \mathcal{T}_{\varepsilon_0}(P_{SXU})\}$$

which tends to 0 thanks to a typicality argument and the following result.

**Lemma 9.** *For $i \in [1,k]$, let $\delta_n = 2^{-n^\beta}$ where $0 < \beta < 1/2$,*

$$\mathbb{V}\left( P_{S_i^{1:n} \widetilde{X}_i^{1:n} \widetilde{U}_i^{1:n}}, P_{S^{1:n} X^{1:n} U^{1:n}} \right) \leq 2\sqrt{\log 2}\sqrt{n\delta_n}.$$

*Proof.* By the chain rule, we have

$$\mathbb{D}\left( P_{S^{1:n} X^{1:n} U^{1:n}} \middle\| P_{S_i^{1:n} \widetilde{X}_i^{1:n} \widetilde{U}_i^{1:n}} \right) \tag{6}$$
$$= \mathbb{D}\left( P_{X^{1:n}|S^{1:n}} \middle\| P_{\widetilde{X}_i^{1:n}|S_i^{1:n}} \middle| P_{S^{1:n}} \right)$$
$$+ \mathbb{D}\left( P_{U^{1:n}|X^{1:n}S^{1:n}} \middle\| P_{\widetilde{U}_i^{1:n}|\widetilde{X}_i^{1:n}S_i^{1:n}} \middle| P_{X^{1:n}S^{1:n}} \right)$$

We call $D_1$ and $D_2$ the first and the second term. Then :

$$D_1 \overset{(a)}{=} \mathbb{D}\left( P_{X^{1:n}} \middle\| P_{\widetilde{X}_i^{1:n}} \right) \overset{(b)}{=} \mathbb{D}\left( P_{Z^{1:n}} \middle\| P_{\widetilde{Z}_i^{1:n}} \right) \tag{7}$$
$$\overset{(c)}{=} \sum_{j=1}^{n} \mathbb{D}\left( P_{Z_i^j|Z_i^{1:j-1}} \middle\| P_{\widetilde{Z}_i^j|\widetilde{Z}_i^{1:j-1}} \middle| P_{Z_i^{1:j-1}} \right)$$
$$\overset{(d)}{=} \sum_{j\in A_1\cup A_2} \mathbb{D}\left( P_{Z_i^j|Z_i^{1:j-1}} \middle\| P_{\widetilde{Z}_i^j|\widetilde{Z}_i^{1:j-1}} \middle| P_{Z_i^{1:j-1}} \right)$$
$$\overset{(e)}{=} \sum_{j\in A_1\cup A_2} \left(1 - H\left(Z_i^j | Z_i^{1:j-1}\right)\right) \overset{(f)}{<} n|\mathcal{V}_X| \leq n\delta_n$$

where $(a)$ comes from the fact that $X$ is independent of $S$, $(b)$ from the invertibility of $G_n$, $(c)$ from the chain rule, $(d)$ from (4), $(e)$ from the fact that the conditional distribution $P_{\widetilde{Z}_i^j|\widetilde{Z}_i^{1:j-1}}$ is uniform for $j$ in $A_1$ and $A_2$ and $(f)$ from Definition (2).

Similarly, $D_2 < n\delta_n$. Then $D_1 + D_2 < 2n\delta_n$ and the proof is completed using Pinsker's inequality.

□

# References


[1] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.

[2] P. Cuff and C. Schieler, "Hybrid codes needed for coordination over the point-to-point channel," in *Proc. of Allerton Conference on Communication, Control and Computing*, 2011, pp. 235–239.

[3] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages : A random binning analogy," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2410–2429, 2016.

[4] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Capacity-distortion trade-off in channels with state," in *Proc. of Allerton Conference on Communication, Control and Computing*, 2010, pp. 1311–1318.

[5] G. Cervia, L. Luzzi, M. R. Bloch, and M. L. Treust, "Polar coding for empirical coordination of signals and actions over noisy channels," in *Proc. of IEEE Information Theory Workshop*, 2016, pp. 81–85.

[6] E. Arıkan, "Source polarization," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 899–903.

[7] R. A. Chou, M. R. Bloch, and E. Abbe, "Polar coding for secret-key generation," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6213–6237, 2015.

[8] M. Bloch and J. Barros, *Physical-layer security : from information theory to security engineering*. Cambridge University Press, 2011.