Polar codes for empirical coordination over noisy channels with strictly causal encoding
Giulia Cervia, Laura Luzzi, Mael Le Treust, Matthieu R Bloch

To cite this version:
Giulia Cervia, Laura Luzzi, Mael Le Treust, Matthieu R Bloch. Polar codes for empirical coordination over noisy channels with strictly causal encoding. XXVIème colloque GRETSI, Sep 2017, Juan-Les-Pins, France. hal-01718146

HAL Id: hal-01718146
https://hal.science/hal-01718146
Submitted on 27 Feb 2018

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

Giulia CERVIA\textsuperscript{1}, Laura LUZZI\textsuperscript{1}, Maël LE TREUST\textsuperscript{1}, Matthieu R. BLOCH\textsuperscript{2}

\textsuperscript{1}ETIS UMR 8051, Université Paris Seine, Université Cergy-Pontoise, ENSEA, CNRS, Cergy, France.
\textsuperscript{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 $X^{1:n} := (X^1, \ldots, X^n)$ 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

![Diagram](image1.png)

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 \( C \in C_n \) (Figure 1). Node 1 observes an i.i.d. sequence of actions \( S_1^{1:n} \in S^n \) with discrete probability distribution \( P_S \). Node 1 then selects a signal \( X_1^{1:n} \) such that \( X_1^i = f_1(S_1^{1:i-1}, C) \), where \( f_1 \in \{ f_1 \}_{n=1}^n \). \( f_1 : S_1^{1:n-1} \times C_n \to X_1 \) is the strictly causal encoder. The signal \( X_1^{1:n} \) is transmitted over a discrete memoryless channel with transition probability \( P_{Y_1|X} \). Upon receiving \( Y_1^{1:n} \in Y^n \), Node 2 selects an action \( \hat{S}_1^{1:n} = g^n(Y_1^{1:n}, C) \), where \( g^n : Y^n \times C_n \to \hat{S}_1^{1:n} \) is the non-causal encoder. 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 two agents' outputs.

Definition 1. A distribution \( P_{SXY} \) is achievable if for all \( \varepsilon > 0 \) there exists a sequence of codes \( \{(f^n, g^n)\}_{n=1}^\infty \) such that
\[
\lim_{n \to \infty} \mathbb{P}\{ T_{S_1^{1:n}, Y_1^{1:n}} - n, P_{SXY} \geq \varepsilon \} = 0,
\]
where \( T_{S_1^{1:n}, Y_1^{1:n}} = g^n(s, x, y, s) \) is the empirical distribution of the tuple \( (S_1^{1:n}, X_1^{1:n}, Y_1^{1:n}, S_1^{1:n}) \) induced by the code.

The empirical coordination region \( C \) is the set of achievable distributions \( P_{SXY} \).

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 \( C \) is given by
\[
C := \left\{ P_{SXY} : \begin{array}{c}
P_{SXY} = P_S P_X P_{Y|X} P_{S|XY} \\ P_{SXY|UY} = P_S P_{X|S} P_{Y|S} P_{U|Y} \\ I(X, U; S) \leq I(X; U; Y) \\ \|U\| \leq |S||X||Y| + 1 \end{array} \right\}
\]

Remark 3. By the chain rule, we have
- \( I(X, U; S) = I(U; S|x) + I(X; S) = I(U; S|X) \) since \( S_1^{1:n} \) and \( X_1^{1:n} \) are independent.
- \( I(X, U; Y) = I(U; Y|x) + I(X; Y) = I(X; Y) \) because of the Markov chain \( U \rightarrow X \rightarrow Y \).

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

Theorem 4. For all \( P_{SXY} \in C \) such that \( \|U\| = 0, 1 \), 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 \( C \). The proof can be generalized to the case where \( \|U\| \) is a prime number using non-binary polar codes.

3 Polar coding scheme

Consider the random vectors \( S_1^{1:n}, U_1^{1:n}, X_1^{1:n}, Y_1^{1:n} \) and \( \hat{S}_1^{1:n} \) generated i.i.d. according to \( P_{SXY} \) that factorize as in (1) with the same mutual information and cardinality constraints.

Polarize \( X \) Let \( Z_1^{1:n} = X_1^{1:n} G_n \) be the polarization of \( X_1^{1:n} \), 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:
\[
\begin{align*}
V_X &= \{ j \in [1, n] : H(Z_j | Z_{j-1}^j) > 1 - \delta_n \}, \\
H_X &= \{ j \in [1, n] : H(Z_j | Z_{j-1}^j) > \delta_n \}, \\
H_{X|Y} &= \{ j \in [1, n] : H(Z_j | Z_{j-1}^j, Y_1^n) > \delta_n \}.
\end{align*}
\]

Partition the set \( [1, n] \) into four disjoint sets:
\[
A_1 := V_X \cap H_{X|Y}, \quad A_2 := V_X \cap H_{X|Y}, \\
A_3 := V_X \cap H_{X|Y}, \quad A_4 := V_X \cap H_{X|Y}.
\]

Remark 6. We have:
- \( V_X \subset H_X \) and \( \lim_{n \to \infty} \frac{|H_X \setminus V_X|}{n} = 0 \) [6].
- \( A_1 \cup A_2 = V_X \) and \( \lim_{n \to \infty} \frac{|V_X|}{n} = H(X) \) [7].
- \( A_1 \cup A_3 = H_{X|Y} \) and \( \lim_{n \to \infty} \frac{|H_{X|Y}|}{n} = H(X|Y) \) [6].

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

Polarize \( U \) Let \( V_1^{1:n} = U_1^{1:n} G_n \) be the polarization of \( U_1^{1:n} \) and define:
\[
\begin{align*}
V_{U|XS} &= \{ j \in [1, n] : H(V_j | V_{j-1}^j X_1^n S_1^n) > 1 - \delta_n \}, \\
H_{U|XS} &= \{ j \in [1, n] : H(V_j | V_{j-1}^j X_1^n S_1^n) > \delta_n \}, \\
H_{U|X} &= \{ j \in [1, n] : H(V_j | V_{j-1}^j X_1^n) > \delta_n \}.
\end{align*}
\]

Partition the set \( [1, n] \) into four disjoint sets:
\[
B_1 := V_{U|XS} \cap H_{U|X} = V_{U|XS}, \quad B_2 := V_{U|XS} \cap H_{U|X} = \emptyset, \\
B_3 := V_{U|XS} \cap H_{U|X}, \quad B_4 := V_{U|XS} \cap H_{U|X} = H_{U|X}.
\]

Remark 7. We have:
- \( V_{U|XS} \subset H_{U|XS} \) and \( \lim_{n \to \infty} \frac{|H_{U|XS} \setminus V_{U|XS}|}{n} = 0 \) [6].
- \( B_1 = V_{U|XS} \) and \( \lim_{n \to \infty} \frac{|V_{U|XS}|}{n} = H(U|XS) \) [7].
- \( B_4 = H_{U|X} \) and \( \lim_{n \to \infty} \frac{|V_{U|XS}|}{n} = 1 - H(U|X) \) [7].
- \( B_3 \cup B_4 = V_{U|XS} \) and \( \lim_{n \to \infty} \frac{|V_{U|XS}|}{n} = 1 - H(U|XS) \) [7].

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

Encoding We use a chaining construction over multiple blocks. The encoder observes \( \{S_1^{1:n}, S_2^{1:n}, \ldots, S_k^{1:n}\} \), where \( S_0^{1:n} \) is a uniform random sequence and \( S_1^{1:n} \) for \( i \in [1, k] \) are blocks of the source. It then generates for each block \( i \in [1, k] \) random variables \( \tilde{Z}_i^{1:n} \) and \( \tilde{V}_i^{1:n} \) following the procedure described in Algorithm 1. In particular, the chaining construction proceeds as follows. The bits in \( A_1 \subset V_X \) and \( B_1 \subset V_{U|XS} \) are chosen with uniform probability using
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 \(A_2 \subseteq \mathcal{V}_X\) 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 \(P_{Z_{i+j-1}^{i+j-1}X_{i+j}S_{i+j}^n}\) and \(P_{Y_{i+j-1}^{i+j-1}X_{i+j}S_{i+j}^n}\) (given \(S_{i}^n\) and \(X_{i}^{1:n}\)) respectively [6].

From the second block, let \(A'_2\) and \(B'_2\) be two disjoint subsets of \(A_2\) such that \(|A'_2| = |A_3|\) and \(|B'_2| = |B_3|\). 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 \(A'_2\) and \(B'_2\) 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 \(A'_2\) and \(B'_2\) in the block \(i+1\) are uniform. The bits in \(A_2 := A_2 \setminus (A'_2 \cup B'_2)\) are chosen with uniform probability using the local randomness source \(M\).

The encoder then computes \(X_{i}^{1:n} = \tilde{Z}_{i}^{1:n}G_n\) for \(i = 1, \ldots, k\) and sends it over the channel. We use an extra \((k+1)\)-th block to send a version of \(\tilde{Z}_k[A_3]\) encoded with a good channel code as in [5, Section III.B].
Convergence in each block implies overall convergence; this requires a few steps:

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

\[
\lim_{k \to \infty} \frac{|A_1| + |A_3| + |B_1| + |B_3|}{kn} = \lim_{k \to \infty} \frac{|Y_X| + |H_U|}{nk} = 0.
\]

4 Proof of Theorem 4

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

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

This requires a few steps:

1. \( \forall i \in [1,k] \), \( \lim_{n \to \infty} P\left\{ \mathbb{V}(T_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{SXY\bar{S}}) > \varepsilon \right\} = 0; \)
2. \( \forall i \in [1,k] \), \( \lim_{n \to \infty} P\left\{ \mathbb{V}(T_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{SXY\bar{S}}) > \varepsilon \right\} = 0; \)
3. \( \forall i \in [1,k] \), \( \lim_{n \to \infty} P\left\{ \mathbb{V}(T_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{SXY\bar{S}}) > \varepsilon \right\} = 0; \)
4. Convergence in each block implies overall convergence;
5. The theorem follows from the fact that:

\[
\mathbb{V}(T_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{SXY\bar{S}}) \leq \mathbb{V}(T_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{SXY\bar{S}}).
\]

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:

\[
\mathcal{T}_{\varepsilon_0}(P_{SXY\bar{S}}) := \{(s, x, u) | \mathbb{V}(P_{SXY\bar{S}}, T_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}) \leq \varepsilon_0 \}
\]

Observe that for the i.i.d. distribution, we have:

\[
\lim_{n \to \infty} P\left\{ (s, x, u) \in \mathcal{T}_{\varepsilon_0}(P_{SXY\bar{S}}) \right\} = 1.
\]

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

\[
P\left\{ \mathbb{V}(T_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{SXY\bar{S}}) > \varepsilon_0 \right\} = \sum_{s, x, u} P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}(s, x, u) \mathbb{1}\{(s, x, u) \notin \mathcal{T}_{\varepsilon_0}(P_{SXY\bar{S}})\}
\]

\[
= \sum_{s, x, u} \left( P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}(s, x, u) - P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}(s, x, u) \right)
+ P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}(s, x, u) \mathbb{1}\{(s, x, u) \notin \mathcal{T}_{\varepsilon_0}(P_{SXY\bar{S}})\}
\]

\[
\leq \mathbb{V}(P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}} + \mathbb{P}\{(s, x, u) \notin \mathcal{T}_{\varepsilon_0}(P_{SXY\bar{S}})\})
\]

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

Lemma 9. For \( i \in [1, k] \), let \( \delta_n \leq 2^{-n^a} \) where \( 0 < \beta < 1/2 \),

\[
\mathbb{V}(P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}, P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}}) \leq 2 \sqrt{\log 2} \sqrt{n \delta_n}.
\]

Proof. By the chain rule, we have:

\[
D \left( P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}} \mid \mid P_{S^{1:n}X^{1:n}Y^{1:n}S^{1:n}} \right) = D \left(P_{X^{1:n}S^{1:n}} \mid \mid \mathbb{P}_{S^{1:n}} S_{1:n} \right) + D \left( P_{U^{1:n}X^{1:n}S^{1:n}} \mid \mid P_{X^{1:n}S^{1:n}} \right)
\]

We call \( D_1 \) and \( D_2 \) the first and the second term, then:

\[
D_1 \leq D \left( P_{Z_i^{1:n}X^{1:n}Y^{1:n}} \right) = D \left( P_{Z_i^{1:n}X^{1:n}Y^{1:n}} \right)
\]

\[
\leq \sum_{j=1}^{n} D \left( P_{Z_i^{1:n}X^{1:n}S^{1:n}} \mid \mid P_{Z_i^{1:n}X^{1:n}S^{1:n}} \right),
\]

\[
\leq \sum_{j=1}^{n} \left( 1 - H \left( Z_i^{1:n} \mid Z_i^{1:j-1} \right) \right) \leq n \delta_n,
\]

where (a) comes from the fact that \( Z \) 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 \( P_{Z_i^{1:j-1}} \mid Z_i^{1:j-1} \) 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. Trestis, “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. Arinkin, “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.