Empirical coordination, state masking and state amplification: Core of the decoder's knowledge

Mael Le Treust, Matthieu Bloch

Empirical Coordination, State Masking and State Amplification: Core of the Decoder's Knowledge

Maël Le Treust* and Matthieu Bloch†
* ETIS, UMR 8051 / ENSEA, Université Cergy-Pontoise, CNRS,
6, avenue du Ponceau, 95014 Cergy-Pontoise CEDEX, FRANCE
Email: mael.le-treust@ensea.fr
† School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta, Georgia 30332
Email: matthieu.bloch@ece.gatech.edu

***Abstract*—We revisit the problem of state masking and state amplification for state-dependent channel with causal state information at the encoder from the point of view of empirical coordination. Empirical coordination, which requires all sequences of symbols to be jointly typical for a target joint probability distribution, provides a unified perspective to simultaneously study state masking, state amplification, and capacity-distortion trade-off. Our main result is a characterization of the set of achievable rates, information leakages and joint distributions. We also discuss several specializations and extensions of the result, including the cases of zero message rate, without empirical coordination, strictly causal encoding, two-sided state information and noisy channel feedback. We introduce the notion of "core of the decoder's knowledge," to capture what the decoder can infer about all the signals involved in the model.**

***Index Terms*—Shannon Theory, State-Dependent Channel, Information Leakage, Empirical Coordination, State Masking, State Amplification, Causal Encoding, Noisy Channel Feedback.**

I. Introduction

State-dependent channels with state information at the encoder have attracted significant interest and spawned a vast literature since the works by Shannon [1] and Gel'fand Pinsker [2]. The main idea behind coding schemes for state-dependent channels is to have the encoder match the statistics of the input symbols to those of the channel using his knowledge of the state. This problem turns out to have deep connections with digital watermarking, memory with defect, cognitive radio and secret-key agreement [3].

More recently, the problem of communication over state-dependent channels was modified with the additional requirement of estimating the channel state parameter at the decoder. The authors of [4] have consequently examined the problem of *state masking* by characterizing the information leakage about the state, whereas the authors of [5] have studied the problem of *state amplification*. State masking and state amplification have been considered simultaneously in [6] for secure communication, in [7] for correlated information sources and in [8] for applications to energy harvesting. Another approach consists in determining the minimal distortion between the channel state and the decoder output, for a given amount of reliable information. Optimal capacity-distortion trade-offs have been characterized in [9] for causal encoder, in [10] for non-causal encoder with Gaussian channel, and in [11], for non-causal encoder with common reconstruction.

In this paper, we simultaneously investigate the problems of state masking, state amplification and capacity-distortion trade-off, through the framework of empirical coordination. Empirical coordination, which is connected to the coordination of autonomous agents in the literature of game theory [12], refers to the set of target joint probability distributions that are achievable by empirical frequencies of symbols [13], [14]. Optimal solutions has been provided for strictly causal and causal encoding [16], for perfect channel [15], for strictly causal and causal decoding [17], with source feedforward [18], for lossless decoding [19], with secrecy constraint [20], with two-sided state information [21] and channel feedback [22].

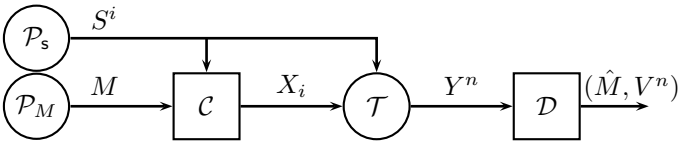

Fig. 1. Causal encoding function $f_i : \mathcal{M} \times \mathcal{S}^i \to \mathcal{X}$, for all $i \in \{1, \dots, n\}$ and non-causal decoding function $g : \mathcal{Y}^n \to \mathcal{M} \times \mathcal{V}^n$.

In this paper, we characterize the set of information rates, information leakages about channel state and joint probability distributions that are achievable. Then, we consider the expectation of the cost or the distortion with respect to the set of achievable joint distributions and this provides the region of achievable rate, information leakage, distortion and cost. We introduce the notion of "core of the decoder's knowledge" corresponding to what the decoder can exactly infer about the other random variables. We determine the optimal solutions for particular cases such as, zero message rate, without empirical coordination, strictly causal encoding, two-sided state information and noisy channel feedback.

System model, definitions and the main result are stated in Sec. II. Particular cases are studied in Sec. III. Conclusions and sketch of proofs are stated in Sec. IV, App. A and B.

* Support of INS2I CNRS through project PEPS JCJC CoReDe 2015.
† Support of National Science Foundation under award CCF 1320304.

II. System model and main result

Figure 1 represents the problem under investigation. S^n, X^n, Y^n, V^n stands for sequences of random variables of channel states $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$, inputs of the channel $x^n \in \mathcal{X}^n$, outputs of the channel $y^n \in \mathcal{Y}^n$ and decoder's output $v^n \in \mathcal{V}^n$, respectively. The sets $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{V}$ have finite cardinality. $M \in \mathcal{M}$ denotes the uniform random message and $\hat{M}$ its decoded version. The set of probability distributions $\mathcal{P}(X)$ over $\mathcal{X}$ is denoted by $\Delta(\mathcal{X})$. The notation $||\mathcal{Q} - \mathcal{P}||_{\mathsf{tv}} = 1/2 \cdot \sum_{x \in \mathcal{X}} |\mathcal{Q}(x) - \mathcal{P}(x)|$ stands for the total variation distance between probability distributions $\mathcal{Q}$ and $\mathcal{P}$. The notation $Y \multimap X \multimap U$ denotes the Markov chain property corresponding to $\mathcal{P}(y|x,u) = \mathcal{P}(y|x)$ for all (u,x,y). Channel state S is i.i.d. distributed with $\mathcal{P}_{\mathsf{s}}$, the channel is memoryless with transition probability $\mathcal{T}_{\mathsf{y|xs}}$ and these statistics are known by encoder $\mathcal{C}$ and decoder $\mathcal{D}$.

Definition II.1 *A code with causal encoding $c \in \mathcal{C}(n, \mathcal{M})$ is a tuple of functions $c = (\{f_i\}_{i \in \{1,\ldots,n\}}, g)$ defined by* (1), (2).

$$f_i \quad : \mathcal{M} \times \mathcal{S}^i \longrightarrow \Delta(\mathcal{X}), \qquad \forall i \in \{1, \ldots, n\}, \quad (1)$$

$$g \quad : \mathcal{Y}^n \longrightarrow \mathcal{M} \times \Delta(\mathcal{V}^n). \quad (2)$$

$\mathsf{N}(s|s^n)$ denotes the occurrence number of symbol $s \in \mathcal{S}$ in sequence s^n. The empirical distribution $Q^n \in \Delta(\mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ of sequences (s^n, x^n, y^n, v^n) is defined by (3).

$$\begin{aligned} Q^n(s,x,y,v) &= \frac{\mathsf{N}(s,x,y,v|s^n,x^n,y^n,v^n)}{n}, \\ \forall (s,x,y,v) &\in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}. \end{aligned} \quad (3)$$

Definition II.2 *Fix a target rate R, a target information leakage E and a target probability distribution $\mathcal{Q} \in \Delta(\mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$. The triple $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$ is achievable if for all $\varepsilon > 0$, there exists a $\bar{n} \in \mathbb{N}$ such that for all $n \geq \bar{n}$, there exists a code with causal encoding $c \in \mathcal{C}(n, \mathcal{M})$ that satisfies:*

$$\begin{aligned} \frac{\log_2 |\mathcal{M}|}{n} &\geq \mathsf{R} - \varepsilon, \\ \left| \mathcal{L}_{\mathsf{e}}(c) - \mathsf{E} \right| &\leq \varepsilon, \qquad \text{with} \qquad \mathcal{L}_{\mathsf{e}}(c) = \frac{1}{n} \cdot I(S^n; Y^n), \\ \mathcal{P}_{\mathsf{e}}(c) &= \mathcal{P}_c\Big(M \neq \hat{M}\Big) + \mathcal{P}_c\Big(\Big|\Big|Q^n - \mathcal{Q}\Big|\Big|_{\mathsf{tv}} \geq \varepsilon\Big) \leq \varepsilon. \end{aligned}$$

We denote by $\mathcal{A}$ the set of achievable triples $(\mathsf{R}, \mathsf{E}, \mathcal{Q}) \in \mathcal{A}$.

$Q^n \in \Delta(\mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ is the random variable of the empirical distribution of the sequences of symbols (S^n, X^n, Y^n, V^n) induced by the code $c \in \mathcal{C}(n, \mathcal{M})$ and the probability distributions of the source $\mathcal{P}_{\mathsf{s}}$ and channel $\mathcal{T}_{\mathsf{y|xs}}$.

Theorem II.3 (Causal encoding) *Consider a target joint probability distribution $\mathcal{Q} \in \Delta(\mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$, with marginal distributions $\mathcal{P}_{\mathsf{s}}(s)$, $\mathcal{T}(y|x,s)$, that decomposes as follows:*

$$\mathcal{Q}(s,x,y,v) = \mathcal{P}_{\mathsf{s}}(s) \times \mathcal{Q}(x|s) \times \mathcal{T}(y|x,s) \times \mathcal{Q}(v|s,x,y). \quad (4)$$

• *The triple $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$ is achievable* if and only if *there exists auxiliary random variables (W_1, W_2) with probability distribution $\mathcal{Q}(s, w_1, w_2, x, y, v) \in \mathbb{Q}_{\mathsf{e}}$ that satisfies:*

$$I(S; W_1, W_2, Y) \leq \begin{cases} \mathsf{R} & \leq I(W_1, W_2; Y) - I(W_2; S|W_1), \quad (5) \\ \mathsf{E} & \leq H(S), \quad (6) \\ \mathsf{R} + \mathsf{E} & \leq I(W_1, S; Y). \quad (7) \end{cases}$$

$\mathbb{Q}_{\mathsf{e}}$ is the set of joint distributions $\mathcal{Q}(s, w_1, w_2, x, y, v)$ with marginal $\mathcal{Q}(s, x, y, v)$, that decompose as follows:

$$\mathcal{P}_{\mathsf{s}}(s) \times \mathcal{Q}(w_1) \times \mathcal{Q}(w_2|s, w_1) \times \mathcal{Q}(x|s, w_1) \times \mathcal{T}(y|x, s) \times \mathcal{Q}(v|y, w_1, w_2).$$

Supports satisfy $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq |\mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| + 4$.

Theorem II.3 characterizes the optimal trade-offs between reliable transmission, information leakage and empirical coordination. The proof is stated in [23] and in App. A and B.

Remark II.4 Equation (5) is redundant with (6) and (7) since both Markov chains $X \multimap (S, W_1) \multimap W_2$ and $Y \multimap (X, S) \multimap (W_1, W_2)$, imply the Markov chain $Y \multimap (W_1, S) \multimap W_2$ and equation (7) writes: $I(W_1, W_2, S; Y) = I(W_1, S; Y)$.

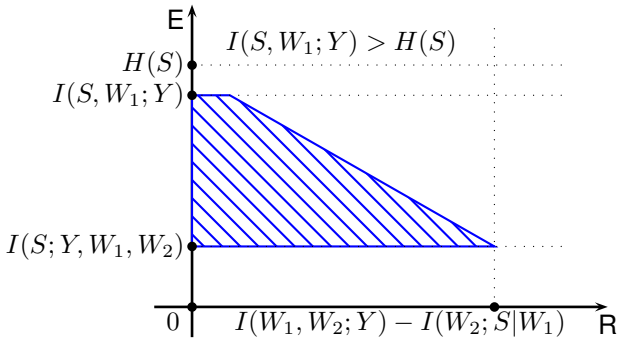

Fig. 2. Region of achievable $(\mathsf{R}, \mathsf{E}) \in \mathcal{A}$ for fixed joint probability distribution $\mathcal{Q}(s, w_1, w_2, x, y, v)$, when $I(S, W_1; Y) > H(S)$ is satisfied.

Remark II.5 (Strictly causal encoding) The set of achievable triples $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$ for *strictly* causal encoder $f_i : \mathcal{M} \times \mathcal{S}^{i-1} \to \mathcal{X}$, $\forall i \in \{1, \ldots, n\}$ instead of causal encoder, is characterized by replacing the auxiliary random variable W_1 by the channel input X in Theorem II.3, as stated in [23].

A. Achievable region of rate, info. leakage, distortion and cost

We consider a cost function $c : \mathcal{X} \to \mathbb{R}$ for channel input X and a distortion function $d : \mathcal{S} \times \mathcal{V} \to \mathbb{R}$ between the channel state S and the decoder output V. As mentioned in [17] and [19], the empirical coordination approach provides directly the optimal $(\mathsf{R}, \mathsf{E}, \mathsf{C}, \mathsf{D})$ rate-leakage-cost-distortion trade-offs.

Corollary II.6 *A tuple of rate, information leakage, distortion and cost $(\mathsf{R}, \mathsf{E}, \mathsf{D}, \mathsf{C})$ is achievable if and only if there exists a joint probability distribution $\mathcal{Q}(s, x, y, v)$ such that the triple $(\mathsf{R}, \mathsf{E}, \mathcal{Q}) \in \mathcal{A}$ is achievable and that satisfies:*

$$\mathbb{E}_{\mathcal{Q}}\Big[c(X)\Big] = \mathsf{C}, \qquad \mathbb{E}_{\mathcal{Q}}\Big[d(S, V)\Big] = \mathsf{D}. \quad (8)$$

The proof is a direct consequence of Theorem II.3, since $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$ is achievable and equations (8) are satisfied. Otherwise, $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$ is not achievable. Corollary II.6 extends to any general objective function $\Phi : \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V} \mapsto \mathbb{R}$ by considering the expectation $\mathbb{E}_{\mathcal{Q}}\Big[\Phi(U, X, Y, V)\Big]$ with respect to the set of achievable joint distributions $\mathcal{Q}$, as in [17].

III. Optimal solutions for particular cases

A detailed version of these results is provided in [23].

A. Zero-rate $\mathsf{R}=0$ and minimal information leakage $\mathsf{E}^\star(\mathcal{Q})$

We consider the pairs of information leakage and joint probability distribution $(\mathsf{E},\mathcal{Q})\in\mathcal{A}$ that are achievable with zero-rate $\mathsf{R}=0$. By Theorem II.3, there exists a probability distribution $\mathcal{Q}(s,w_1,w_2,x,y,v)\in\mathbb{Q}_{\mathsf{e}}$, that satisfies:

$$0 < I(W_1,W_2;Y) - I(W_2;S|W_1), \tag{9}$$

$$I(S;W_1,W_2,Y) < \mathsf{E}. \tag{10}$$

We define the minimal achievable information leakage $\mathsf{E}^\star(\mathcal{Q}) = \min_{(\mathsf{E},\mathcal{Q})\in\mathcal{A}} \mathsf{E}$, for a fixed $\mathcal{Q}$ with zero-rate $\mathsf{R}=0$.

Corollary III.1 *The minimal information leakage E corresponding to the probability distribution $\mathcal{Q}\in\Delta(\mathcal{S}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V})$ is given by:*

$$\mathsf{E}^\star(\mathcal{Q}) = \min_{\substack{\mathcal{Q}(s,w_1,w_2,x,y,v)\in\mathbb{Q}_{\mathsf{e}},\\ s.t.\ I(W_1,W_2;Y)-I(W_2;S|W_1)>0}} I(S;W_1,W_2,Y). \tag{11}$$

The proof is a direct consequence of Theorem II.3, since $\mathsf{E}^\star(\mathcal{Q})$ is achievable and every $\mathsf{E}<\mathsf{E}^\star(\mathcal{Q})$ is not achievable.

Corollary III.1 states that the minimal information leakage $I(S^n;Y^n) = I(S^n;W_1^n,W_2^n,Y^n)$ is close to $n\cdot I(S;W_1,W_2,Y)$, as if the sequences (S^n,Y^n,W_1^n,W_2^n) were generated with an i.i.d. probability distribution $\mathcal{Q}_{\mathsf{syw_1w_2}}^{\times n}$. In fact, the empirical coordination of the sequences $(S^n,W_1^n,W_2^n,Y^n)\in A_\varepsilon^{\star n}(\mathcal{Q})$ implies that the posterior probability distribution $\mathcal{P}(S^n|W_1^n,W_2^n,Y^n)$ is closely related to the single-letter conditional probability distribution $\mathcal{Q}_{\mathsf{s|yw_1w_2}}$. Based on the triple of symbols (Y,W_1,W_2), the decoder generates symbol V using the conditional probability distribution $\mathcal{Q}_{\mathsf{v|yw_1w_2}}$ and infers the channel state S with the conditional probability distribution $\mathcal{Q}_{\mathsf{s|yw_1w_2}}$. We claim that the random variables (Y,W_1,W_2) determine the "*core of the decoder's knowledge*," regarding other random variables, as S and V.

B. Removing the empirical coordination requirement

In this section, the decoder decodes the message M but does not return a symbol V coordinated with other random variables (S,X,Y). The decoding function writes $g:\mathcal{Y}^n\to\mathcal{M}$ and the error probability should satisfy $\mathcal{P}_{\mathsf{e}}(c)=\mathcal{P}_c\big(M\neq\hat{M}\big)\leq\varepsilon$.

Theorem III.2 (Removing empirical coordination) *A pair of rate and information leakage (R,E) is achievable* if and only if *there exists an auxiliary random variable W_1 with probability distribution $\mathcal{Q}(s,w_1,x,y)\in\mathbb{Q}_{\mathsf{c}}$, such that:*

$$I(S;Y|W_1)\leq \begin{array}{ll} \mathsf{R} & \leq I(W_1;Y), \quad (12)\\ \mathsf{E} & \leq H(S), \quad (13)\\ \mathsf{R}+\mathsf{E} & \leq I(W_1,S;Y). \quad (14)\end{array}$$

$\mathbb{Q}_{\mathsf{c}}$ is the set of distributions $\mathcal{Q}(s,w_1,x,y)$ that decomposes:

$$\mathcal{P}_{\mathsf{s}}(s)\times\mathcal{Q}(w_1)\times\mathcal{Q}(x|s,w_1)\times\mathcal{T}(y|x,s).$$

The support of W_1 is bounded by $|\mathcal{W}_1|\leq|\mathcal{S}\times\mathcal{X}\times\mathcal{Y}|+1$.

Achievability proof comes from Theorem II.3, by removing auxiliary random variable $W_2=\emptyset$ and considering single block coding instead of block-Markov coding.

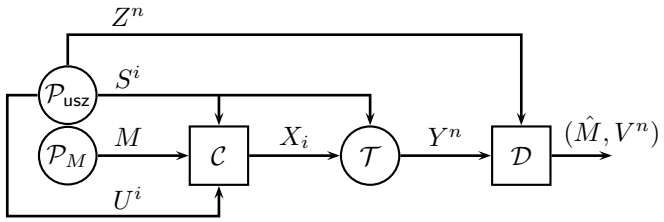

Fig. 3. Causal encoding function $f_i:\mathcal{M}\times\mathcal{U}^i\times\mathcal{S}^i\to\mathcal{X}$, for all $i\in\{1,\ldots,n\}$ and non-causal decoding function $g:\mathcal{Y}^n\times\mathcal{Z}^n\to\mathcal{M}\times\mathcal{V}^n$.

Remark III.3 (Strictly causal encoder - no coordination) The set of achievable pairs (R,E) for *strictly* causal encoder $f_i:\mathcal{M}\times\mathcal{S}^{i-1}\to\mathcal{X}$, $\forall i\in\{1,\ldots,n\}$ instead of causal encoder, is characterized by replacing the auxiliary random variable W_1 by the channel input X in Theorem III.2, [23].

C. Two-sided state information

The case of two-sided state information is represented by Fig. 3. The distribution $\mathcal{P}_{\mathsf{usz}}\in\Delta(\mathcal{U}\times\mathcal{S}\times\mathcal{Z})$ generates i.i.d. correlated channel state S, information source U and state information Z at the decoder. The encoding is causal $f_i:\mathcal{M}\times\mathcal{U}^i\times\mathcal{S}^i\to\mathcal{X}$, for all $i\in\{1,\ldots,n\}$ and the decoding is non-causal $g:\mathcal{Y}^n\times\mathcal{Z}^n\to\mathcal{M}\times\mathcal{V}^n$. The information leakage $\mathcal{L}_{\mathsf{e}}(c)$ is defined with respect to the pair of sequences (U^n,S^n) and the observation (Y^n,Z^n) of the decoder.

$$\mathcal{L}_{\mathsf{e}}(c) \;=\; \frac{1}{n}\cdot I(U^n,S^n;Y^n,Z^n). \tag{15}$$

Theorem III.4 (Two-sided state information) *Consider a target joint distribution $\mathcal{Q}\in\Delta(\mathcal{U}\times\mathcal{S}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V})$ with marginal $\mathcal{P}_{\mathsf{usz}}(u,s,z)$, $\mathcal{T}(y|x,s)$, that decomposes as:*

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

• *The triple $(\mathsf{R},\mathsf{E},\mathcal{Q})$ is achievable* if and only if *there exists auxiliary random variables (W_1,W_2) with probability distribution $\mathcal{Q}(u,s,z,w_1,w_2,x,y,v)\in\mathbb{Q}_2$, such that:*

$$I(U,S;W_1,W_2,Y,Z)\leq \begin{array}{ll} \mathsf{R} & \leq I(W_1,W_2;Y,Z)-I(W_2;U,S|W_1),\\ \mathsf{E} & \leq H(U,S),\\ \mathsf{R}+\mathsf{E} & \leq I(W_1,U,S;Y,Z).\end{array}$$

$\mathbb{Q}_2$ is the set of distributions with marginal $\mathcal{Q}$ that decompose:

$$\mathcal{P}_{\mathsf{usz}}\times\mathcal{Q}_{\mathsf{w_1}}\times\mathcal{Q}_{\mathsf{w_2|usw_1}}\times\mathcal{Q}_{\mathsf{x|usw_1}}\times\mathcal{T}_{\mathsf{y|xs}}\times\mathcal{Q}_{\mathsf{v|yzw_1w_2}}.$$

Supports of (W_1,W_2) are bounded by $\max(|\mathcal{W}_1|,|\mathcal{W}_2|)\leq d+4$ with $d=|\mathcal{U}\times\mathcal{S}\times\mathcal{Z}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{V}|$.

The achievability proof of Theorem III.4 follows directly from the proof of Theorem II.3, by replacing the random variable of the channel state S by the pair (U,S) and the random variable of the channel output Y by the pair (Y,Z), as in [21].

D. Noisy channel feedback observed by the encoder

We characterize the set of achievable triples $(\mathsf{R},\mathsf{E},\mathcal{Q})$ when the encoder has noisy feedback Y_2 from the state-dependent broadcast channel $\mathcal{T}(y_1,y_2|x,s)$, as in Fig. 4. Encoding function writes $f_i:\mathcal{M}\times\mathcal{S}^i\times\mathcal{Y}_2^{i-1}\to\mathcal{X}$, $\forall i\in\{1,\ldots,n\}$ and both decoding function and information leakage remain unchanged.

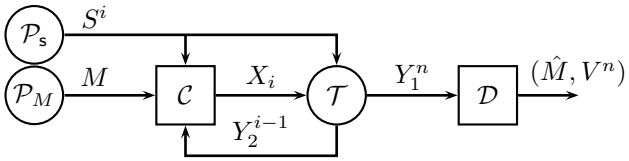

Fig. 4. Noisy feedback Y_2^{i-1} from state-dependent broadcast $\mathcal{T}(y_1, y_2|x, s)$ channel. Encoding writes $f_i : \mathcal{M} \times \mathcal{S}^i \times \mathcal{Y}_2^{i-1} \to \mathcal{X}$, $\forall i \in \{1, \dots, n\}$.

Theorem III.5 (Noisy channel feedback)
Consider target joint distribution $\mathcal{Q} \in \Delta(\mathcal{S} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{V})$ with marginal $\mathcal{P}_{\mathsf{s}}(s)$, $\mathcal{T}(y_1, y_2|x, s)$, that decomposes as:

$$\mathcal{P}_{\mathsf{s}}(s) \times \mathcal{Q}(x|s) \times \mathcal{T}(y_1, y_2|x, s) \times \mathcal{Q}(v|s, x, y_1, y_2).$$

• *The triple $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$ is achievable* if and only if *there exists auxiliary random variables (W_1, W_2) with probability distribution $\mathcal{Q}(s, w_1, w_2, x, y_1, y_2, v) \in \mathbb{Q}_{\mathsf{f}}$, such that:*

$$I(S; W_1, W_2, Y_1) \leq \begin{array}{ll} \mathsf{R} & \leq I(W_1, W_2; Y_1) - I(W_2; S, Y_2|W_1), \\ \mathsf{E} & \leq H(S), \\ \mathsf{R}+\mathsf{E} & \leq I(W_1, S; Y_1). \end{array}$$

$\mathbb{Q}_{\mathsf{f}}$ is the set of distributions with marginal $\mathcal{Q}$ that decompose:

$$\mathcal{P}_{\mathsf{s}} \times \mathcal{Q}_{\mathsf{w}_1} \times \mathcal{Q}_{\mathsf{x}|\mathsf{s}\mathsf{w}_1} \times \mathcal{T}_{\mathsf{y}_1\mathsf{y}_2|\mathsf{x}\mathsf{s}} \times \mathcal{Q}_{\mathsf{w}_2|\mathsf{s}\mathsf{w}_1\mathsf{y}_2} \times \mathcal{Q}_{\mathsf{v}|\mathsf{y}_1\mathsf{w}_1\mathsf{w}_2}.$$

Supports of (W_1, W_2) are bounded by $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq d + 4$ with $d = |\mathcal{S} \times \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{V}|$.

The achievability proof of Theorem III.5 follows directly from the proof of Theorem II.3, by replacing the pair (S^n, W_1^n) by the triple (S^n, W_1^n, Y_2^n) when the encoder finds W_2^n. The decoding functions and the leakage analysis remain unchanged.

Remark III.6 (Noisy feedback improve coordination)
Channel feedback increases the set of achievable triples $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$, since the new conditional distribution $\mathcal{Q}_{\mathsf{w}_2|\mathsf{s}\mathsf{w}_1\mathsf{y}_2}$ depends on channel outputs Y_2 whereas the previous one $\mathcal{Q}_{\mathsf{w}_2|\mathsf{s}\mathsf{w}_1}$ does not. The information constraints of Theorem III.5 are reduced to that of Theorem II.3 as soon as $\mathcal{Q}_{\mathsf{w}_2|\mathsf{s}\mathsf{w}_1\mathsf{y}_2} = \mathcal{Q}_{\mathsf{w}_2|\mathsf{s}\mathsf{w}_1} \iff W_2 \multimap (S, W_1) \multimap Y_2 \iff I(W_2; Y_2|S, W_1) = 0$. More details on channel feedback for empirical coordination are provided in [22].

IV. Conclusion

We have simultaneously investigated the problems of state masking, state amplification and empirical coordination for state-dependent channel with causal state information at the encoder. We have characterized the achievable triples of rate, information leakage and joint probability distribution and we provide optimal distortion and cost levels. The information of the decoder regarding other random variables is characterized precisely and called the "core of the decoder's knowledge." We provide optimal solutions for zero message rate, without empirical coordination, strictly causal encoding, two-sided state information and noisy channel feedback.

Appendix

The full versions of the proofs are available in [23].

A. Sketch of achievability proof of Theorem II.3

Achievability proof is based on rate splitting and on the proof of Theorem V.1 in [21]. Consider a triple $(\mathsf{R}, \mathsf{E}, \mathcal{Q})$ and a probability distribution $\mathcal{Q}_{\mathsf{s}\mathsf{w}_1\mathsf{w}_2\mathsf{x}\mathsf{y}\mathsf{v}} \in \mathbb{Q}_{\mathsf{e}}$ satisfying equations (5), (6), (7) of Theorem II.3. We introduce parameters R_{L}, R_{K} and block-Markov code $c \in \mathcal{C}$ defined over $B \in \mathbb{N}$ blocks of length $n \in \mathbb{N}$.

Random Codebook. We generate $2^{n(H(S)+\varepsilon)}$ sequences of states $S^n(l, j) \sim \mathcal{P}_{\mathsf{s}}^{\times n}$ indexed by $(l, j) \in \mathcal{M}_{\mathsf{L}} \times \mathcal{M}_{\mathsf{J}}$ with cardinalities $|\mathcal{M}_{\mathsf{L}}| = 2^{n\mathsf{R}_{\mathsf{L}}}$ and $|\mathcal{M}_{\mathsf{J}}| = 2^{n\mathsf{R}_{\mathsf{J}}}$. We generate $2^{n(\mathsf{R}+\mathsf{R}_{\mathsf{L}}+\mathsf{R}_{\mathsf{K}})}$ sequences $W_1^n(m, l, k)$, drawn from $\mathcal{Q}_{\mathsf{w}_1}^{\times n}$ with index $(m, l, k) \in \mathcal{M} \times \mathcal{M}_{\mathsf{L}} \times \mathcal{M}_{\mathsf{K}}$ with $|\mathcal{M}_{\mathsf{K}}| = 2^{n\mathsf{R}_{\mathsf{K}}}$. For each triple (m, l, k), we generate the same number $2^{n(\mathsf{R}+\mathsf{R}_{\mathsf{L}}+\mathsf{R}_{\mathsf{K}})}$ of sequences $W_2^n(m, l, k, \hat{m}, \hat{l}, \hat{k})$ with indexes $(\hat{m}, \hat{l}, \hat{k})$, drawn from $\mathcal{Q}_{\mathsf{w}_2|\mathsf{w}_1}^{\times n}$ depending on $W_1^n(m, l, k)$.

Encoding function. At block b, the encoder observes the previous channel states S_{b-1}^n and finds the indexes (l_{b-1}, j_{b-1}) such that $\big(S^n(l_{b-1}, j_{b-1}), S_{b-1}^n\big) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical. Encoder observes the message m_b and the index l_{b-1} and recalls the sequence $W_1^n(m_{b-1}, l_{b-2}, k_{b-1})$ of block $b-1$. It finds k_b such that $\big(S_{b-1}^n, W_1^n(m_{b-1}, l_{b-2}, k_{b-1}), W_2^n(m_{b-1}, l_{b-2}, k_{b-1}, m_b, l_{b-1}, k_b)\big) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical. Encoder sends X_b^n drawn from $\mathcal{Q}_{\mathsf{x}|\mathsf{s}\mathsf{w}_1}^{\times n}$ depending $W_1^n(m_b, l_{b-1}, k_b)$ and S_b^n.

Decoding function. At block b, decoder recalls Y_{b-1}^n and the indexes $(m_{b-1}, l_{b-2}, k_{b-1})$ of $W_1^n(m_{b-1}, l_{b-2}, k_{b-1})$ correctly decoded. Decoder observes Y_b^n and finds (m_b, l_{b-1}, k_b) such that $\big(Y_b^n, W_1^n(m_b, l_{b-1}, k_b)\big) \in A_\varepsilon^{\star n}(\mathcal{Q})$ and $\big(Y_{b-1}^n, W_1^n(m_{b-1}, l_{b-2}, k_{b-1}), W_2^n(m_{b-1}, l_{b-2}, k_{b-1}, m_b, l_{b-1}, k_b)\big) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical. Decoder returns the message m_b and V_{b-1}^n drawn from $\mathcal{Q}_{\mathsf{v}|\mathsf{y}\mathsf{w}_1\mathsf{w}_2}^{\times n}$ depending on $\big(Y_{b-1}^n, W_1^n(m_{b-1}, l_{b-2}, k_{b-1}), W_2^n(m_{b-1}, l_{b-2}, k_{b-1}, m_b, l_{b-1}, k_b)\big)$. Decoder knows that the sequences $\big(S_{b-1}^n, W_1^n(m_{b-1}, l_{b-2}, k_{b-1}), W_2^n(m_{b-1}, l_{b-2}, k_{b-1}, m_b, l_{b-1}, k_b), X_{b-1}^n, Y_{b-1}^n, V_{b-1}^n\big) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical and S_{b-1}^n belongs to the bin with index $l_{b-1} \in \mathcal{M}_{\mathsf{L}}$.

$$\begin{aligned} \mathsf{R}_{\mathsf{L}} &= \mathsf{E} - I(S; W_1, W_2, Y) - 2\varepsilon \geq 0, && (16) \\ \mathsf{R}_{\mathsf{L}} + \mathsf{R}_{\mathsf{J}} &= H(S) + \varepsilon, && (17) \\ \mathsf{R}_{\mathsf{K}} &= I(W_2; S|W_1) + \varepsilon, && (18) \\ \mathsf{R} + \mathsf{R}_{\mathsf{L}} + \mathsf{R}_{\mathsf{K}} &< I(W_1; Y) + I(W_2; Y|W_1) - \varepsilon. && (19) \end{aligned}$$

Equations (17), (18), (19) imply that for all $n \geq \bar{n}$:

$$\begin{aligned} &\mathbb{E}_c\Big[\mathcal{P}\Big(S_{b-1}^n \notin A_\varepsilon^{\star n}(\mathcal{Q})\Big)\Big] \leq \varepsilon, \\ &\mathbb{E}_c\Big[\mathcal{P}\Big(\forall (l_{b-1}, j_{b-1}) \in \mathcal{M}_{\mathsf{L}} \times \mathcal{M}_{\mathsf{J}},\ \ (S^n(l_{b-1}, j_{b-1}), S_{b-1}^n) \notin A_\varepsilon^{\star n}\Big)\Big] \leq \varepsilon, \\ &\mathbb{E}_c\Big[\mathcal{P}\Big(\forall k_b \in \mathcal{M}_{\mathsf{K}},\ (S_{b-1}^n, W_1^n(m_{b-1}, l_{b-2}, k_{b-1}), \\ &\quad W_2^n(m_{b-1}, l_{b-2}, k_{b-1}, m_b, l_{b-1}, k_b)) \notin A_\varepsilon^{\star n}(\mathcal{Q})\Big)\Big] \leq \varepsilon, \\ &\mathbb{E}_c\Big[\mathcal{P}\Big(\exists (m_b, l_{b-1}, k_b) \neq (m_b', l_{b-1}', k_b'),\ \text{s.t.}\ \big\{(Y_b^n, W_1^n(m_b', l_{b-1}', k_b')) \in \\ &\quad A_\varepsilon^{\star n}(\mathcal{Q})\big\} \cap \big\{(Y_{b-1}^n, W_1^n(m_{b-1}, l_{b-2}, k_{b-1}), W_2^n(m_{b-1}, l_{b-2}, k_{b-1}, \\ &\quad m_b', l_{b-1}', k_b')) \in A_\varepsilon^{\star n}(\mathcal{Q})\big\}\Big)\Big] \leq \varepsilon. \end{aligned}$$

For a large number of blocks $B \in \mathbb{N}$, the sequences are jointly typical for distribution $\mathcal{Q}$ and the rate R is correctly decoded,.

Upper bound on the information leakage

$$\begin{aligned} & n \cdot \mathcal{L}_{\mathsf{e}}(c) = I(S^n; Y^n|C) \\ \leq\ & I(S^n; W_1^n, W_2^n, L, M|C) + I(S^n; Y^n|W_1^n, W_2^n, L, M, C) && (20) \\ \leq\ & n \cdot (\mathsf{E} - I(S; Y|W_1, W_2) - \varepsilon) + n \cdot (I(S; Y|W_1, W_2) + \varepsilon) && (21) \\ \leq\ & n \cdot (\mathsf{E} + 2\varepsilon). && (22) \end{aligned}$$

Eq. (21) comes from equations (25) and (29) and concludes.

$$\begin{aligned} & I(S^n; W_1^n, W_2^n, L, M|C) = I(S^n; W_2^n, L|W_1^n, M, C) && (23) \\ \leq\ & H(W_2^n, L|W_1^n, M, C) \leq \log_2 |\mathcal{M}_{\mathsf{L}}| + H(W_2^n|W_1^n, L, M, C) && (24) \\ \leq\ & \log_2 |\mathcal{M}_{\mathsf{L}}| + \log_2 |\mathcal{M}_{\mathsf{K}}| = n \cdot \Big(\mathsf{E} - I(S; Y|W_1, W_2) - \varepsilon\Big). && (25) \end{aligned}$$

Eq. (23) is due to the independence of S^n and (W_1^n, M, C).
Eq. (25) comes from the size $|\mathcal{M}_{\mathsf{K}}|$ of the bin of sequences W_2^n.

$$\begin{aligned} & I(S^n; Y^n | W_1^n, W_2^n, L, M, C) \\ = \; & H(Y^n | W_1^n, W_2^n, L, M, C) - H(Y^n | S^n, W_1^n, W_2^n, L, M, C) && (26) \\ = \; & H(Y^n | W_1^n, W_2^n, L, M, C) - n \cdot H(Y|S, W_1) && (27) \\ \leq \; & n \cdot (H(Y|W_1, W_2) + \varepsilon) - n \cdot H(Y|S, W_1, W_2) && (28) \\ \leq \; & n \cdot (I(S; Y|W_1, W_2) + \varepsilon). && (29) \end{aligned}$$

Eq. (27) is due to the cascade of memoryless channels $\mathcal{Q}(x|w_1, s) \times \mathcal{T}(y|x, s)$. Eq. (28) comes from empirical coordination hence sequences (Y^n, W_1^n, W_2^n) are jointly typical.

Lower bound on the information leakage

We introduce the set $\mathcal{S}^\star(w_1^n, w_2^n, y^n, l)$ of sequences $s^n \in \mathcal{S}^n$ that are jointly typical with (w_1^n, w_2^n, y^n) and that belong to bin $l \in \mathcal{M}_{\mathsf{L}}$.

$$\begin{aligned} & \mathbb{E}_c \left[\left| \mathcal{S}^\star(w_1^n, w_2^n, y^n, l) \right| \right] = \mathbb{E}_c \left[\sum_{\substack{s^n \in A_\varepsilon^{\star n} \\ (w_1^n, w_2^n, y^n)}} \mathbb{1}\left\{ s^n \in \mathcal{B}(l) \right\} \right] && (30) \\ = \; & \sum_{s^n \in A_\varepsilon^{\star n}(w_1^n, w_2^n, y^n)} \mathbb{E}_c \left[\mathbb{1}\left\{ s^n \in \mathcal{B}(l) \right\} \right] && (31) \\ \leq \; & \sum_{s^n \in A_\varepsilon^{\star n}(w_1^n, w_2^n, y^n)} 2^{-n \cdot \mathsf{R}_{\mathsf{L}}} \leq 2^{n \cdot (H(S|W_1, W_2, Y) - \mathsf{R}_{\mathsf{L}} + \varepsilon)}. && (32) \end{aligned}$$

Eq. (32) is due to the random code that induces a uniform probability distribution over the bins $\mathcal{B}(l)$ and to the size of the set of typical sequences. Markov's inequality gives:

$$\begin{aligned} & \mathcal{P}_c \left[\left| \mathcal{S}^\star(w_1^n, w_2^n, y^n, l) \right| \geq 2^{n \cdot (H(S|W_1, W_2, Y) - \mathsf{R}_{\mathsf{L}} + 2\varepsilon)} \right] \\ \leq \; & \frac{\mathbb{E}_c \left[\left| \mathcal{S}^\star(w_1^n, w_2^n, y^n, l) \right| \right]}{2^{n \cdot (H(S|W_1, W_2, Y) - \mathsf{R}_{\mathsf{L}} + 2\varepsilon)}} \leq \frac{2^{n \cdot (H(S|W_1, W_2, Y) - \mathsf{R}_{\mathsf{L}} + \varepsilon)}}{2^{n \cdot (H(S|W_1, W_2, Y) - \mathsf{R}_{\mathsf{L}} + 2\varepsilon)}} && (33) \\ \leq \; & 2^{-n\varepsilon} \leq \varepsilon. && (34) \end{aligned}$$

We define the event $F = 0$ if the size $\left| \mathcal{S}^\star(w_1^n, w_2^n, y^n, l) \right| < 2^{n \cdot (H(S|W_1, W_2, Y) - \mathsf{R}_{\mathsf{L}} + 2\varepsilon)}$ and $F = 1$ otherwise.

$$\begin{aligned} & n \cdot \mathcal{L}_{\mathsf{e}}(c) = I(S^n; Y^n | C) = I(S^n; Y^n, W_1^n, W_2^n, L | C) && (35) \\ = \; & n \cdot H(S) - H(S^n | Y^n, W_1^n, W_2^n, L, C) && (36) \\ \geq \; & n \cdot H(S) - H(S^n | Y^n, W_1^n, W_2^n, L, C, F = 0) \\ - \; & \mathcal{P}(F = 1) \cdot \log_2 |\mathcal{S}| - h_b(\mathcal{P}(F = 1)) && (37) \\ \geq \; & n \cdot (H(S) - H(S|W_1, W_2, Y) - \mathsf{R}_{\mathsf{L}} - 3\varepsilon) \geq n \cdot (\mathsf{E} - \varepsilon). && (38) \end{aligned}$$

Eq. (35) is due to decoding of (W_1^n, W_2^n, L) based on Y^n.
Eq. (36) is due to the i.i.d. property of the channel state S.
Eq. (37) is inspired by Fano's inequality.
Eq. (38) is due to size $\left| \mathcal{S}^\star(w_1^n, w_2^n, y^n, l) \right|$ and parameter R_{L}.

B. Sketch of converse of Theorem II.3

Consider a code $c \in \mathcal{C}(n, M)$ with small error probability and auxiliary random variables $W_{1,i} = (M, S^{i-1})$ and $W_{2,i} = Y_{i+1}^n$ satisfying Markov chains of the set of probability distributions $\mathbb{Q}_{\mathsf{e}}$.

$$\begin{aligned} & n \cdot H(S) \geq n \cdot \mathsf{E} \geq I(S^n; Y^n) - n \cdot \varepsilon && (39) \\ \geq \; & \sum_{i=1}^n I(S_i; Y^n, M | S^{i-1}) - H(M|Y^n) - n \cdot \varepsilon && (40) \\ \geq \; & \sum_{i=1}^n I(S_i; Y_{i+1}^n, M, S^{i-1}, Y_i) - n \cdot 2\varepsilon && (41) \\ = \; & \sum_{i=1}^n I(S_i; W_{1,i}, W_{2,i}, Y_i) - n \cdot 2\varepsilon. && (42) \end{aligned}$$

Eq. (39) comes from the definition of the information leakage.
Eq. (41) is due to Fano's inequality and the i.i.d. property of S.

$$\begin{aligned} n \cdot \mathsf{R} \; \leq \; & I(M; Y^n) + n \cdot \varepsilon && (43) \\ \leq \; & \sum_{i=1}^n I(M, S^{i-1}, Y_{i+1}^n; Y_i) - \sum_{i=1}^n I(S_i; Y_{i+1}^n | S^{i-1}, M) + n \cdot \varepsilon && (44) \\ = \; & \sum_{i=1}^n I(W_{1,i}, W_{2,i}; Y_i) - \sum_{i=1}^n I(S_i; W_{2,i} | W_{1,i}) + n \cdot \varepsilon. && (45) \end{aligned}$$

Eq. (43) is due to Fano's inequality and properties of mutual info.
Eq. (44) is due to Csiszár Sum Identity and properties of MI.

$$\begin{aligned} & n \cdot \Big(\mathsf{E} + \mathsf{R} \Big) \leq I(S^n; Y^n) + I(M; Y^n) + H(M|Y^n) && (46) \\ \leq \; & I(S^n, M; Y^n) + n \cdot \varepsilon \leq \sum_{i=1}^n I(S_i, M, S^{i-1}; Y_i) && (47) \\ + \; & \sum_{i=1}^n I(S_{i+1}^n, Y_{i+1}^n; Y_i | S_i, M, S^{i-1}) + n \cdot \varepsilon && (48) \\ = \; & \sum_{i=1}^n I(S_i, W_{1,i}; Y_i) + n \cdot \varepsilon. && (49) \end{aligned}$$

Eq. (47) is due to Fano's ineq. and independence M and S^n.
Eq. (48) comes from the causal encoding that implies the Markov chain: $Y_i \multimap (S_i, M, S^{i-1}) \multimap (S_{i+1}^n, Y_{i+1}^n)$.

REFERENCES

[1] C. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289–293, Oct 1958.
[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. of Control and I.T.*, vol. 9, no. 1, pp. 19–31, 1980.
[3] A. Khisti, S. Diggavi, and G. Wornell, "Secret-key agreement with channel state information at the transmitter," *IEEE Transactions on Information Forensics and Security*, vol. 6, pp. 672 – 681, 2011.
[4] N. Merhav and S. Shamai, "Information rates subject to state masking," *IEEE Trans. on Information Theory*, vol. 53, pp. 2254–2261, June 2007.
[5] Y.-H. Kim, A. Sutivong, and T. Cover, "State amplification," *IEEE Trans. on Information Theory*, vol. 54, no. 5, pp. 1850–1859, 2008.
[6] O. Koyluoglu, R. Soundararajan, and S. Vishwanath, "State amplification under masking constraints," in *49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 936–943, Sept 2011.
[7] T. Courtade, "Information masking and amplification: The source coding setting," in *IEEE Internat. Symp. on Info. Th.*, pp. 189–193, July 2012.
[8] K. Tutuncuoglu, O. Ozel, A. Yener, and S. Ulukus, "State amplification and state masking for the binary energy harvesting channel," in *IEEE Information Theory Workshop (ITW)*, pp. 336–340, Nov 2014.
[9] C. Choudhuri, Y.-H. Kim, and U. Mitra, "Causal state communication," *IEEE Trans. on Information Theory*, vol. 59, pp. 3709–3719, June 2013.
[10] T. M. C. A. Sutivong, M. Chiang and Y.-H. Kim, "Channel capacity and state estimation for state-dependent gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1486–1495, Apr. 2005.
[11] Y. Steinberg, "Coding and common reconstruction," *IEEE Transactions on Information Theory*, vol. 55, pp. 4995–5010, Nov 2009.
[12] O. Gossner, P. Hernandez, and A. Neyman, "Optimal use of communication resources," *Econometrica*, vol. 74, pp. 1603–1636, Nov. 2006.
[13] G. Kramer and S. Savari, "Communicating probability distributions," *IEEE Trans. on Information Theo.*, vol. 53, no. 2, pp. 518 – 525, 2007.
[14] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity," *IEEE Trans. on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.
[15] P. Cuff and L. Zhao, "Coordination using implicit communication," *Information Theory Workshop (ITW), IEEE*, pp. 467– 471, 2011.
[16] 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.
[17] M. Le Treust, "Empirical coordination for the joint source-channel coding problem," *submitted to IEEE Trans. on Information Theory, http://arxiv.org/abs/1406.4077*, 2014.
[18] B. Larrousse, S. Lasaulce, and M. Bloch, "Coordination in distributed networks via coded actions with application to power control," *Submitted to IEEE Transactions on Information Theory, http://arxiv.org/abs/1501.03685*, 2014.
[19] M. Le Treust, "Correlation between channel state and information source with empirical coordination constraint," in *IEEE Information Theory Workshop (ITW)*, pp. 272–276, Nov 2014.
[20] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *IEEE Trans. on Information Theory*, vol. 60, pp. 7584–7605, Dec 2014.
[21] M. Le Treust, "Empirical coordination with two-sided state information and correlated source and state," in *IEEE Int. S. Info. Th. (ISIT)*, 2015.
[22] M. Le Treust, "Empirical coordination with channel feedback and strictly causal or causal encoding," in *IEEE Inter. Symp. Info. Th. (ISIT)*, 2015.
[23] M. Le Treust and M. Bloch, "State masking, state amplification and empirical coordination with causal encoding," Technical report, 2016. https://cloud.ensea.fr/index.php/s/xJ1Hf13r3s9xVob