Typicality with Feedback

Thomas Sturma, Michèle Wigger

LTCI, Télécom Paris, Palaiseau, France

{thomas.sturma, michele.wigger}@telecom-paris.fr

Abstract—The main objective of this paper is to analyze a closed-loop feedback system where a transmitter probes a discrete memoryless channel and can adapt its inputs based on the previous channel outputs. We prove that, regardless of the transmitter's strategy, the conditional type of the outputs given the inputs remains close to the DMC transition law $P_{Y|X}$. This general result enables the study of fundamental limits in certain adaptive systems.

As an application, we establish a converse result for an integrated sensing and communication (ISAC) model. In this setting, the transmitter also functions as a radar receiver, aiming to simultaneously transmit a message over the channel and estimate the channel state from the backscattered feedback signals. We show that the fundamental limits of the closed loop system are the same as of the open-loop system where the transmitter can use the feedback signal to estimate the state but not to produce adaptive channel inputs. This result holds as long as the sum of the admissible-average-decoding-error-probability, denoted ϵ, and the admissible-excess-distortion-probability, denoted δ, is below 1, i.e., $\delta + \epsilon < 1$.

Index Terms—channel coding, feedback, typicality, strong converse, change of measure, integrated sensing and communication

I. Introduction

In this paper, we investigate a closed-loop system where a transmitter probes a discrete memoryless channel (DMC) and the transmitter can adapt its inputs based on past channel outputs, see Figure 1. We show that, irrespective of the transmitter's strategy in producing the inputs, the empirical conditional distribution—or conditional type—of the channel outputs given the inputs remains close to the underlying channel law $P_{Y|X}$. For open-loop systems, where channel inputs cannot depend on past outputs, the desired statement follows from Chebychef's inequality in a straightforward way and in the asymptotic limit is nothing but the weak law of large numbers. With closed-loop strategies, the situation is more complicated because the transmitter can adapt its inputs so as to maximize the deviation between the empirical distribution and the true channel law $P_{Y|X}$. Proving the desired statement on the conditional type requires a different proof.

Specifically, we identify for each input-output pair (a, b) the transmitter strategy that maximizes the probability that the conditional type of output b given input a deviates from $P_{Y|X}(b|a)$ by more than a fixed threshold μ. The optimal strategy has a simple form: the transmitter feeds input a to the channel until the number of observed b symbols deviates from b by more than n times the threshold μ, at which point it starts feeding non-"a" symbols to the channel until the number of channel uses n is reached. Optimality of this strategy is

proved by first noting the bijection between input strategies and labels on a full $|\mathcal{Y}|$-ary tree of depth n, and then deriving the labeling on the tree that maximizes the desired probability. We finally analyze this probability for the described strategy using a simple inequality on sums of i.i.d. random variables.

We subsequently use the established result on the conditional types to prove a converse result for an Integrated Sensing and Communication (ISAC) system where the transmitter sends a message over a state-dependent discrete memoryless channel (SDMC) with feedback and from the feedback signals also reconstructs the state sequence up to desired distortion. Our converse proof is based on change of measure arguments [1]–[4], [6], [10] and establishes that for any permissible decoding error probability ϵ and excess distortion probability δ whose sum is below 1, i.e., $\epsilon + \delta < 1$, the set of data rates and reconstruction distortions that are simultaneously achievable, coincides with this set in an open-loop system where the transmitter is not allowed to adapt its inputs in function of the feedback signals. Our results assume perfect feedback, and extend to noisy or degraded versions thereof. Extension to arbitrary generalized feedback signals as considered in [8]–[10] is left for future work.

Notation. Sets are denoted using calligraphic fronts such as $\mathcal{X}$ and $\mathcal{Y}$. All random variables are assumed finite. For $i \leq j$ two positive integers, sequences of random variables $(X_i, \ldots, X_j)$ and realizations $(x_i, \ldots, x_j)$ are abbreviated by X_i^j and x_i^j respectively. If $i = 1$, we also use X^j and x^j instead of X_i^j and x_i^j. The probability of an event A and the expectation of a random variable X are denoted by $\mathbf{P}(A)$ and $\mathbf{E}[X]$ respectively. Moreover, throughout this paper, 2^{nR} will denote the integer $\lfloor 2^{nR} \rfloor$, and $[\![a, b]\!]$, for integer numbers $a < b$, denotes the set $\{a, \ldots, b\}$. We denote by $\pi_{x^n y^n}$ the joint type of (x^n, y^n) defined in [11]. The marginal type of x^n is denoted π_{x^n}. Finally, we will use the Landau notation $o(1)$ to indicate any function that tends to 0 when $n \to +\infty$.

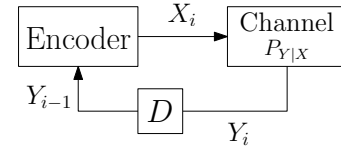## II. Typicality Through a Channel with Feedback

Fig. 1. Typicality through a Channel with Feedback

Consider the setup in Figure 1. An encoder called 'Alex' produces inputs $X_1, \ldots, X_n$ to a channel $P_{Y|X}$ with feedback.

That is, it can observe the past outputs of the channel and choose the inputs in function thereof:

$$X_i = h_i(Y_1, \ldots, Y_{i-1}), \quad i \in \{1, \ldots, n\}, \tag{1}$$

for some functions $h_i(\cdot)$ on appropriate domains. The following lemma asserts that whatever strategy, i.e., sequence of functions $\{h_1, \ldots, h_n\}$, Alex chooses, the conditional type of the outputs Y^n given the inputs X^n is close to the channel law $P_{Y|X}$ with high probability.

Lemma 1: Irrespective of the choice of the functions $\{h_1, \ldots, h_n\}$, for any $(a, b) \in \mathcal{X} \times \mathcal{Y}$ and small value $\mu > 0$:

$$\mathbf{P}\big(\big|\pi_{X^n Y^n}(a, b) - \pi_{X^n}(a) P_{Y|X}(b)\big| > \mu\big) \le (4n\mu^2)^{-1}. \tag{2}$$

While without feedback this statement holds by Chebychef's inequality, the proof is slightly trickier with feedback. Another result examines an upper bound on the probability associated with a specified type of inputs and a specified conditional type of outputs [14].

A. Proof of Lemma 1

Define for any strategy $\mathbf{h} = (h_1(\cdot), \ldots, h_n(\cdot))$ and any $(a, b) \in \mathcal{X} \times \mathcal{Y}$, the random score

$$S_n(\mathbf{h}, a, b) = \sum_{k=1}^{n} \mathbb{1}\{X_k = a\}(\mathbb{1}\{Y_k = b\} - P_{Y|X}(b|a)), \tag{3}$$

for $\{(X_k, Y_k)\}_{k=1}^n$ generated according to the strategy $\mathbf{h}$ and the channel $P_{Y|X}$.

Notice the equivalence:

$$\mathbf{P}\big(\big|\pi_{X^n Y^n}(a, b) - \pi_{X^n}(a) P_{Y|X}(b|a)\big| > \mu\big)$$

$$= \mathbf{P}\left(\left|\sum_{i=1}^{n} \mathbb{1}\{X_i = a\}\big(\mathbb{1}\{Y_i = b\} - P_{Y|X}(b|a)\big)\right| > n\mu\right)$$

$$= \mathbf{P}\left(|S_n(\mathbf{h}, a, b)| > n\mu\right). \tag{4}$$

We shall prove that for any $(a, b) \in \mathcal{X} \times \mathcal{Y}$:

$$\max_{\mathbf{h}} \mathbf{P}\left(|S_n(\mathbf{h}, a, b)| > n\mu\right) \le (4n\mu^2)^{-1}, \tag{5}$$

implying the result in Lemma 1. To this end, we first identify the maximizing strategy $\mathbf{h}$ for any given pair (a, b), and then determine the probability that its score function exceeds $n\mu$.

Remark 1: As suggest by one of the anonymous reviewers, one can prove (5) by first noting that for any given strategy $\mathbf{h}$ the random process $\{S_n(\mathbf{h}, a, b)\}_n$ forms a martingale and then applying Doob's inequality. In our proof we also identify the optimal strategy $\mathbf{h}$, which has a pleasing intuitive form.

1) A maximizing strategy $\mathbf{h}$: Fix $(a, b) \in \mathcal{X} \times \mathcal{Y}$. For a given strategy $\mathbf{h}$, define for each $i \in \{1, \ldots, n\}$ and each tuple y^i:

$$s_i(y^i, \mathbf{h}) = \sum_{\ell=1}^{i} \mathbb{1}\{x_\ell = a\}(\mathbb{1}\{y_\ell = b\} - P_{Y|X}(b|a)), \tag{6}$$

where $x_\ell = h_\ell(y^{\ell-1})$.

Lemma 2: The following strategy is a maximizer of (5):

$$h_k^*(y_1, \ldots, y_{k-1}) = \begin{cases} a, & \text{if } |s_{k-1}(y^{k-1}, \mathbf{h})| < n\mu \\ \neq a & \text{otherwise,} \end{cases} \tag{7}$$
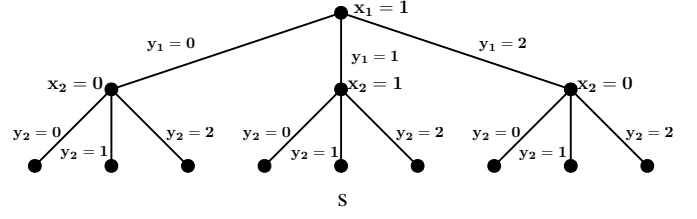

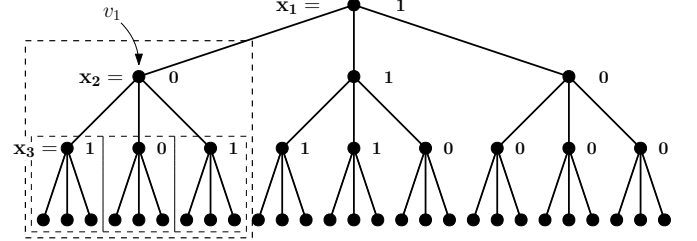Fig. 2. Example of a tree for $n = 2$

Fig. 3. Example of a labeled tree for $n = 3$, $|\mathcal{X}| = 2$ and $|\mathcal{Y}| = 3$.

where $\neq a$ indicates any symbol other than a.

Proof of Lemma 2: We start by noticing that each strategy $\mathbf{h}$ can be represented as labels on all nodes (including the vertex node but excluding leaves) of a full $|\mathcal{Y}|$-ary rooted tree of depth n where the $|\mathcal{Y}|$ edges below a node are labeled by the $|\mathcal{Y}|$ distinct elements of $\mathcal{Y}$. We simply use label $h_k(y_1, \ldots, y_{k-1})$ for the level-k node v that is reached from the root through the edges labeled $y_1, \ldots, y_{k-1}$. Figure 2 illustrates for example the tree (for $n = 2$ and $\mathcal{Y} = \{0, 1, 2\}$) and the node labels corresponding to strategy

$$h_1(\varnothing) = 1 \tag{8}$$

$$h_2(y_1) = \begin{cases} 0 & \text{if } y_1 \in \{0, 2\} \\ 1 & \text{if } y_1 = 1. \end{cases} \tag{9}$$

A second example (for $n = 3$, $|\mathcal{X}| = 2$ and $|\mathcal{Y}| = 3$) is provided in Figure 3, where for readability, we do not show the labels on the edges; they always correspond to the elements of $\mathcal{Y}$ in the same order.

We shall refer by $\mathcal{T}_{\mathbf{h}}$ to the tree corresponding to a strategy $\mathbf{h}$. Notice the following facts for this tree $\mathcal{T}_{\mathbf{h}}$:

- There is a one-to-one correspondance between the leaves of the tree and the output sequences y^n. We will thus simply refer to the sequences y^n as the leaves of the tree.
- The score $s_n(y^n, \mathbf{h})$ of a given leave node is obtained by running from the root node of the tree $\mathcal{T}_{\mathbf{h}}$ to leave y^n, and by summing up for each edge the term $(\mathbb{1}\{y_k = b\} - P_{Y|X}(b|a))$ if its parent node is a and 0 otherwise. We will thus also write $s_i(y^i, \mathcal{T}_{\mathbf{h}})$ instead of $s_i(y^i, \mathbf{h})$. For a general labeled tree $\mathcal{T}$, we define $s_i(y^i, \mathcal{T})$ as

$$\sum_{\ell=1}^{i} \mathbb{1}\{x_\ell(y^{\ell-1}) = a\}(\mathbb{1}\{y_\ell = b\} - P_{Y|X}(b|a)), \tag{10}$$

where $x_\ell(y^{\ell-1})$ is the label in $\mathcal{T}$ at the node on the path from the root indicated by $y_1, \ldots, y_{\ell-1}$.

- The success probability $\mathbf{P}(|S_n(\mathbf{h}, a, b)| > n\mu)$ of strategy $\mathbf{h}$ is directly obtained from the tree $\mathcal{T}_{\mathbf{h}}$ by summing over the probabilities of the leaves with sufficiently high scores.

We generally define for any labeled tree $\mathcal{T}$:

$$\mathcal{S}(\mathcal{T}) \triangleq \{y^n : |s_n(y^n, \mathcal{T})| > n\mu\}, \qquad (11)$$

and the success probability of the tree

$$P(\mathcal{T}) \triangleq \sum_{y^n \in \mathcal{S}(\mathcal{T})} \prod_{i=1}^{n} P_{Y|X}(y_i | x_i(y^{i-1})), \qquad (12)$$

where $x_i(y^{i-1})$ is the label on $\mathcal{T}$ of the node on the path from the root indicated by $y_1, \ldots, y_{i-1}$. Then,

$$P(\mathcal{T}_{\mathbf{h}}) = \mathbf{P}(|S_n(\mathbf{h}, a, b)| > n\mu). \qquad (13)$$

Given these observations, we will denote by $\mathcal{F}$ the set of all labeled trees, and the bijection between the elements of $\mathcal{F}$ and the set of strategies.

After these preliminary remarks, we tackle the proof of Lemma 2. Notice that it is equivalent to proving that there exists an optimal labeled tree $\mathcal{T}^* \in \mathcal{F}$, i.e., $P(\mathcal{T}^*) = \max_{\mathcal{T} \in \mathcal{F}} P(\mathcal{T})$, with the following properties:

1) The tree $\mathcal{T}^*$ is *well-ordered*, meaning only nodes labeled a can have descendants labeled a.
2) A level-k node of $\mathcal{T}^*$ is labeled a if, and only if, $|s_k(y^k, \mathcal{T}^*)| \le n\mu$, where y^k indicates the labels of the path from the root to this node.

To show that 1.) holds, we start with any optimal tree $\mathcal{T}_0 \in \mathcal{F}$ and iteratively construct new trees $\mathcal{T}_1, \mathcal{T}_2, \ldots$ so that $P(\mathcal{T}_k) \ge P(\mathcal{T}_{k-1})$ for any $k \ge 1$. Moreover, after a finite number of steps κ the tree $\mathcal{T}_\kappa$ is well-ordered and optimal.

Tree $\mathcal{T}_k$ is obtained from $\mathcal{T}_{k-1}$ as follows:

- Pick a vertex v_k from tree $\mathcal{T}_{k-1}$ that is labeled $\tilde{a} \ne a$ but for which all parents and one of these children are labeled a. (For example, consider vertex v_1 in Figure 3.)
- For any $y \in \mathcal{Y}$, denote by $\mathcal{B}_y$ the subtree starting at the child of v_k along the edge y. Label all leave nodes of $\mathcal{B}_y$ by $\tilde{a}$ and augment all of these leaves with $|\mathcal{Y}|$ new children, where the edges are labeled by all elements of $\mathcal{Y}$. (Figure 3 indicates the trees $\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2$ for node v_1 with dashed boxes and Figure 4 depicts the augmented tree of $\mathcal{B}_0$.)
- Starting from $\mathcal{T}_{k-1}$, create a new *random* tree $\tilde{\mathcal{A}}$ as follows. For each $y \in \mathcal{Y}$, with probability $P_{Y|X}(y|\tilde{a})$ replace node v_k and its underlying subtree by the augmented tree of $\mathcal{B}_y$. (Figure 5 depicts the realization of $\tilde{\mathcal{A}}$ constructed with the augmented of $\mathcal{B}_0$.)
- Among all $|\mathcal{Y}|$ possible realizations of the tree $\tilde{\mathcal{A}}$, let $\mathcal{T}_k$ be the one that has largest probability of success $P(\cdot)$.

It can be verified that for each iteration k:

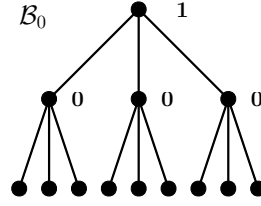$$P(\mathcal{T}_{k-1}) = P(\tilde{\mathcal{A}}) \le P(\mathcal{T}_k). \qquad (14)$$

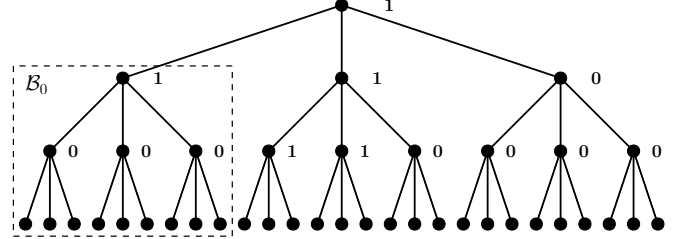

Fig. 4. The augmented tree of $\mathcal{B}_0$

Fig. 5. The realization of $\tilde{\mathcal{A}}$ constructed with the augmented tree of $\mathcal{B}_0$

Above inequality holds simply because the best is no worse than the average. The equality is proved in Appendix A in [13].

To prove 2), notice that once 1) is established, we can focus on well-ordered trees only. By the property of these trees, the path from the root to any leaf $\bar{y}^n$ starts with a certain number of a-labels on the nodes, and then once it reaches a specific node $v^*(\bar{y}^n)$ it switches to non-a labels for $v^*(\bar{y}^n)$ and all its descendants. Notice that in an ordered tree $\mathcal{T}_{\text{ord}}$, all leaves y^n that are descendants of a switching node $v^*(\bar{y}^n)$ have same score as $\bar{y}^n$: $s_n(y^n, \mathcal{T}_{\text{ord}}) = s_n(\bar{y}^n, \mathcal{T}_{\text{ord}})$.

Fix an ordered tree $\mathcal{T}$ and a leaf $\bar{y}^n$, and let k be the depth of "switching" node $v^*(\bar{y}^n)$. We argue that if $|s_i(\bar{y}^i)| > n\mu$ for some $i \in \{1, \ldots, k-1\}$ or $|s_k(\bar{y}^k)| \le n\mu$, then we can find a new ordered tree $\mathcal{T}'$ with a larger set of leaves exceeding the desired score than the original tree. I.e., $\mathcal{S}(\mathcal{T}') \supseteq \mathcal{S}(\mathcal{T})$ which by (12) implies $P(\mathcal{T}') \ge P(\mathcal{T})$. It can also be verified that the procedure will produce $\mathcal{T}^*$ after a finite number of steps. The new ordered tree $\mathcal{T}'$ is obtained as follows:

- Assume $|s_i(\bar{y}^i)| > n\mu$ for some $i \in \{1, \ldots, k-1\}$. In this case, denote the level-i node on the path to $\bar{y}^n$ by $\tilde{v}$ and set in $\mathcal{T}'$ the new node $\tilde{v}$ as the "switching" node of y^n (and naturally of all descendant leaves of $\tilde{v}$). Notice that all descendant leaves y^n of $\tilde{v}$ have score $|s_n(y^n, \mathcal{T}')| > n\mu$ implying in particular that this holds for all leaves that are descendants of the original "switching" node $v^*(\bar{y}^n)$. Since the score of all other leaves did not change from $\mathcal{T}$ to $\mathcal{T}'$, we can conclude the inclusion $\mathcal{S}(\mathcal{T}') \supseteq \mathcal{S}(\mathcal{T})$.
- If $|s_k(\bar{y}^k)| \le n\mu$, then all leaves y^n that are descendants of $v^*(\bar{y}^n)$ has scores below $n\mu$ and do not belong to $\mathcal{S}(\mathcal{T})$. We construct $\mathcal{T}'$ from $\mathcal{T}$ by relabeling $v^*(\bar{y}^n)$ by a, and trivially conclude $\mathcal{S}(\mathcal{T}') \supseteq \mathcal{S}(\mathcal{T})$. $\blacksquare$

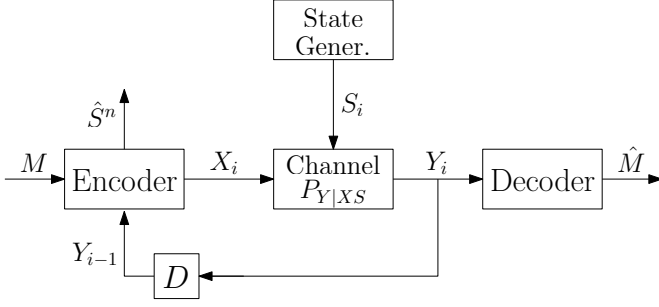*2) Success Probability of Strategy $\mathbf{h}^*$ in (7) :*

Fig. 6. ISAC Model

Proposition 1: We have:

$$\mathbf{P}\left(|S_n(\mathbf{h}^*, a, b)| > n\mu\right) \le (4n\mu^2)^{-1} \qquad (15)$$

Proof: Define the i.i.d. variables $Z_k = \mathbb{1}\{\tilde{Y}_i = b\} - P_{Y|X}(b|a)$ for $k = 1, 2, \ldots$ where $\{\tilde{Y}_i\}_{i=1}^{+\infty}$ is i.i.d. $\sim P_{Y|X}(\cdot|a)$. By Kolmogorov's maximal inequality [12, Theorem 2.5.5]:

$$\mathbf{P}\left(\max_{1 \le k \le n} \left|\sum_{i=1}^{k} Z_i\right| > n\mu\right) \le \frac{\mathbf{E}[(\sum_{k=1}^{n} Z_k)^2]}{n^2\mu^2} \le \frac{1}{4n\mu^2}. \quad (16)$$

We next relate the probability on the left-hand side of (16) with the success probability of our optimal strategy $\mathbf{h}^*$. Notice that for all y^n sequences where $\max_{1 \le k \le n} |\sum_{i \le k} Z_i| \le n\mu$ when $\tilde{Y}^n = y^n$, we also have $|s_k(y^k, \mathbf{h}^*)| \le n\mu$ for all $k \in \{1, \ldots, n\}$. This is, *both are* smaller *than the threshold.* Moreover, these sequences have same probability of arising for the random sequences $\tilde{Y}^n$ and Y^n as the optimal strategy $\mathbf{h}^*$ always produces a, i.e., $x_1 = \ldots = x_n = a$. We obtain:

$$\mathbf{P}\left(\max_{1 \le k \le n} \left|\sum_{i=1}^{k} Z_i\right| > n\mu\right) \le \mathbf{P}\left(|S_n(Y^n, \mathbf{h}^*)| \le n\mu\right). \quad (17)$$

The proof is then concluded by considering the complement probabilities as well as (16). ∎

III. CONVERSE THEOREM FOR AN ISAC MODEL

We use our Lemma 1 to prove a converse for the integrated sensing and communication (ISAC) model in Figure 6.

A transmitter seeks to communicate a random message M to a receiver over a state-dependent discrete memoryless channel (SDMC) $P_{Y|XS}$. The message M is assumed to be uniformly distributed over the set $[\![1, 2^{nR}]\!]$, where $R > 0$ denotes the communication rate and $n > 0$ the blocklength. The SDMC is influenced by a state sequence S^n, which is i.i.d. according to a given probability mass function (pmf) P_S. At each time i, the channel output Y_i is generated based on the current input X_i and the state S_i, according to the SDMC transition probability $P_{Y|XS}$. The outputs are observed at the receiver, and through backscattering also at the transmitter.

The transmitter creates its channel inputs based on the message and the past feedback outputs

$$X_i = f_i^{(n)}(M, Y_1, \ldots, Y_{i-1}), \qquad i = 1, \ldots, n, \quad (18)$$

for some sequence of encoding functions $\{f_i^{(n)}\}_{i=1}^n$. It also uses the feedback outputs Y^n and the produced input sequence X^n (or M) to reconstruct the channel's state sequence S:

$$\hat{S}^n = h^{(n)}(X^n, Y^n). \qquad (19)$$

The receiver attempts to guess the original message M based on the full sequence of channel outputs Y^n using a decoding function $g^{(n)}$:

$$\hat{M} = g^{(n)}(Y^n). \qquad (20)$$

Reliability of communication is evaluated via the average probability of decoding error at the receiver

$$P_e^{(n)} = \mathbf{P}\left(\hat{M} \ne M\right). \qquad (21)$$

Sensing performance at the transmitter is measured by the average expected distortion between the true state sequence S^n and its reconstruction $\hat{S}^n$:

$$\text{dist}^n(\hat{S}^n, S^n) = \frac{1}{n}\sum_{i=1}^{n} d(\hat{S}_i, S_i), \qquad (22)$$

where d denotes a given positive and bounded distortion function.

Our goal is to have small decoding error probability and small excess distortion probability.

Definition 1: A rate-distortion pair (R, D) is (ε, δ)-achievable if there exist sequences of encoding, decoding and estimation functions $\{\{f_i^{(n)}\}_{i=1}^n, g^{(n)}, h^{(n)}\}_{n=1}^{+\infty}$ satisfying

$$\varlimsup_{n \to +\infty} P_e^{(n)} \le \varepsilon \qquad (23)$$

$$\varlimsup_{n \to \infty} \mathbf{P}\left(\text{dist}^n\left(\hat{S}^n, S^n\right) > D\right) \le \delta. \qquad (24)$$

Theorem 1: For any $\varepsilon + \delta < 1$, if a rate-distorsion pair (R, D) is (ε, δ)-achievable then there exists P_X satisfying

$$R = I_{P_X P_S P_{Y|XS}}(X; Y) \qquad (25)$$

$$D \ge \mathbf{E}_{P_X P_S P_{Y|XS}}\left[d(\hat{s}(X, Y), S)\right] \qquad (26)$$

where

$$\hat{s}(x, y) = \min_{\hat{s} \in \mathcal{S}} \sum_s P_{S|XY}(s|x, y)d(\hat{s}, s). \qquad (27)$$

Achievability for $\epsilon = \delta = 0$ follows directly from the results in [8], [9], where it was also shown that the transmitter does not need to rely on the feedback outputs $Y_1, \ldots, Y_{i-1}$ to produce the next input X_i. The contribution of our article lies in proving the converse to this statement in the case where the encoder can use the feedback to produce the channel inputs.

Remark 2: Notice that under the average probability of error criterion considered in this article, above theorem does not hold when $\epsilon + \delta \ge 1$. Better performances can be achieved.

Remark 3: Theorem 1 and Remark 2 can easily be extended also to setups where the transmitter observes a (deterministically or stochastically) degraded version of the outputs $\{Y_i\}$.

A. Proof of Theorem 1

Fix $\epsilon, \delta > 0$ with sum $\epsilon + \delta < 1$. Fix also a sequence of encoding and decoding functions $\{\{f_i^{(n)}\}_{i=1}^n, g^{(n)}\}_{n=1}^\infty$ satisfying (23) and (24) and consider the optimal state estimator (see [9, Lemma 1] $h^n(s^n, x^n) = (\hat{s}(x_1, y_1), \ldots, \hat{s}(x_n, y_n))$, where $\hat{s}(x, y)$ is defined in (27).

Define for any blocklength n the parameter $\mu_n = n^{-1/4}$ and notice that $\mu_n \to 0$ while $n\mu_n^2 \to \infty$ as $n \to +\infty$. Consider now the three conditions:

$$g^{(n)}(y^n) = m \quad (28)$$

$$\text{dist}^n(h^{(n)}(x^n, y^n), s^n) \leq D \quad (29)$$

$$\left| \pi_{x^n, s^n, y^n}(a, b, c) - \pi_{x^n}(a) P_S(b) P_{Y|XS}(c|a, b) \right| \leq \mu_n \quad (30)$$

where in the above, $x_i = f_i(m, y^{i-1})$ for any $i = 1, \ldots, n$.

Fix a constant η such that $0 < \eta < 1 - \varepsilon - \delta$, and define the subset $\tilde{\mathcal{M}}_n \subseteq [\![1, 2^{nR}]\!]$ to consist of all messages m satisfying

$$\mathbf{P}\left(\tilde{M} \neq M \text{ or } \text{dist}^n\left(\hat{S}^n, S^n\right) > D \mid M = m \right) \leq 1 - \eta, \quad (31)$$

and let $\tilde{M}$ be uniformly distributed over $\tilde{\mathcal{M}}_n$. Notice that (23) and (24) imply that (see Appendix B in [13]):

$$\frac{|\tilde{\mathcal{M}}_n|}{2^{nR}} \geq \left(1 - \frac{P_e^{(n)} + P_D^{(n)}}{1 - \eta} \right) =: \gamma_n, \quad (32)$$

where we define $P_D^{(n)} := \mathbf{P}\left(\text{dist}^n\left(\hat{S}^n, S^n\right) > D \right)$.

Define now for any $m \in \mathcal{M}_n$ the set $\mathcal{D}_{n,m}$ consisting of all pairs (s^n, y^n) such that (28), (29), and (30) hold, and introduce random variables $(\tilde{S}^n, \tilde{Y}^n)$ jointly distributed with $\tilde{M}$ through the conditional pmfs:

$$P_{\tilde{S}^n, \tilde{Y}^n | \tilde{M} = m}(s^n, y^n)$$
$$\triangleq \frac{P_S^{\otimes n}(s^n) \prod_{i=1}^n P_{Y|XS}(y_i | x_i(m, y^{i-1}), s_i)}{\Delta_{n,m}}$$
$$\cdot \mathbb{1}\{(s^n, y^n) \in \mathcal{D}_{n,m}\}, \quad (33)$$

where $\Delta_{n,m} \triangleq \mathbf{P}((S^n, Y^n) \in \mathcal{D}_{n,m} \mid M = m)$.

Notice that:

$$\Delta_{n,m} = 1 - \mathbf{P}((S^n, Y^n) \notin \mathcal{D}_{n,m} \mid M = m) \quad (34)$$
$$= 1 - \mathbf{P}(M \neq \hat{M} \text{ or } \text{dist}^n(\hat{S}^n, S^n) > D \text{ or }$$
$$\exists (a, b, c) : |\pi_{X^n, Y^n, S^n}(a, b, c)$$
$$- \pi_{X^n}(a) P_S(c) P_{Y|X}(b|a)| > \mu_n \mid M = m) \quad (35)$$
$$\geq \eta - \sum_{a,b,c} \mathbf{P}(|\pi_{X^n, Y^n, S^n}(a, b, c)$$
$$- \pi_{X^n}(a) P_{Y|X}(b|a) P_S(c)| > \mu_n \mid M = m) \quad (36)$$

By Lemma 1, we obtain for any m and triple (a, b, c):

$$\mathbf{P}\Big(|\pi_{X^n Y^n S^n}(a, b, c)$$
$$- \pi_{X^n}(a) P_{Y|X}(b|a) P_S(c)| > \mu_n \mid M = m \Big) \leq \frac{1}{4n\mu_n^2}, \quad (37)$$

and thus, by (36):

$$\Delta_{n,m} \geq \eta - \frac{|\mathcal{X}||\mathcal{Y}||\mathcal{S}|}{4n\mu_n^2}, \quad \forall m \in \tilde{\mathcal{M}}_n. \quad (38)$$

We now prove the desired bounds on the rate and the distortion. By combining (32) and (28), we establish the following inequality, the proof of which is provided in the extended version of the paper (See Appendix C in [13]):

$$R \leq H(\tilde{Y}_{T_n}) - \frac{1}{n} H(\tilde{Y}^n | \tilde{M}) + o(1), \quad (39)$$

where we defined T_n to be independent of $(\tilde{M}, \tilde{X}^n, \tilde{Y}^n)$ and uniform over $[\![1, n]\!]$.

Lemma 3: There exists a subsequence $\{n_i\}_{i \geq 1}$ and a pmf P_X on $\mathcal{X}$ such that:

$$\lim_{i \to +\infty} P_{\tilde{X}_T, \tilde{Y}_T, \tilde{S}_T}(x, y, s) = P_X(x) P_S(s) P_{Y|XS}(y|x, s) \quad (40)$$

$$\lim_{i \to +\infty} \frac{1}{n_i} H(\tilde{Y}^{n_i} | \tilde{M}) \geq H(Y|X) \quad (41)$$

where the random variables (X, Y) in the entropy term are distributed according to the joint marginal of pmf $P_X P_S P_{Y|XS}$.

Proof: Based on (30) and (38), see Appendix D in the extended version [13]. ∎

Subsequently, by virtue of this lemma and (39), and as the blocklength n_i tends to $+\infty$, it must hold that:

$$R \leq H(Y) - H(Y|X) = I(X; Y). \quad (42)$$

We prove the distortion bound. By (29), with probability 1:

$$D \geq \frac{1}{n_i} \sum_{j=1}^{n_i} d\left(\hat{s}\left(\tilde{X}_j, \tilde{Y}_j\right), \tilde{S}_j \right) \quad (43)$$

$$= \mathbf{E}\left[d\left(\hat{s}\left(\tilde{X}_{T_n}, \tilde{Y}_{T_n}\right), \tilde{S}_{T_n} \right) \right] \quad (44)$$

and by (40), we can conclude by taking $n_i \to +\infty$ that:

$$D \geq \mathbf{E}[d(\hat{s}(X, Y), S)]. \quad (45)$$

IV. CONCLUSION

In this paper, we analyzed a closed-loop communication system where the encoder adapts its transmissions based on previous outputs of a discrete memoryless channel. We established that, despite the presence of feedback, the conditional type of the channel outputs given the inputs remains close to the true channel law $P_{Y|X}$. The core of the proof involved identifying the worst-case encoder strategy and bounding the probability of significant deviation using a random walk argument.

We then applied this general result to an Integrated Sensing and Communication (ISAC) model, where a transmitter seeks to jointly communicate a message and estimate the state of a channel based on the backscattered feedback signals. By leveraging the conditional type result and a change-of-measure technique, we proved that the fundamental limits are the same as in the corresponding open loop system with non-adaptive channel inputs.

REFERENCES

[1] T. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 2–14, 1989.

[2] M. Hamad, M. Wigger, and M. Sarkiss, "Strong converses using typical changes of measures and asymptotic Markov chains," *IEEE Trans. Inf. Theory,* vol. 70, no. 3, pp. 1720-–1737, Mar. 2024.

[3] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," *2009 IEEE International Symposium on Information Theory*, Seoul, Korea (South), 2009, pp. 2316–2320.

[4] W. Gu and M. Effros, "A strong converse in source coding for super-source networks," *2011 IEEE International Symposium on Information Theory Proceedings*, St. Petersburg, Russia, 2011, pp. 395–399.

[5] H. Tyagi and S. Watanabe, "Strong Converse Using Change of Measure Arguments," in *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 689–703, Feb. 2020.

[6] D. Takeuchi and S. Watanabe, "Tight Exponential Strong Converse for Source Coding Problem With Encoded Side Information," in *IEEE Transactions on Information Theory*, vol. 71, no. 3, pp. 1533–1545, March 2025

[7] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, vol. 1, no. 4, pp. 591–606, 1957.

[8] M. Kobayashi, G. Caire, and G. Kramer, "Joint state sensing and communication: Optimal tradeoff for a memoryless case," *2018 IEEE International Symposium on Information Theory Proceedings*, Vail, CO, USA, 2018, pp. 111–115.

[9] M. Ahmadipour, M. Kobayashi, M. Wigger and G. Caire, "An Information-Theoretic Approach to Joint Sensing and Communication," in *IEEE Transactions on Information Theory*, vol. 70, no. 2, pp. 1124–1146, Feb. 2024.

[10] M. Ahmadipour, M. Wigger and S. Shamai, "Strong Converses for Memoryless Bi-Static ISAC," 2023 *IEEE International Symposium on Information Theory* , Taipei, Taiwan, 2023, pp. 1818–1823.

[11] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[12] R. Durrett, *Probability – theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Vol. 49 (Fifth edition of 1991 original ed.). Cambridge University Press, 2019.

[13] Thomas Sturma and Michèle Wigger, "Typicality with feedback," 2024. [Online]. Available : https://arxiv.org/abs/2507.16384

[14] H. Palaiyanur (2011), "The impact of causality on information-theoretic source and channel coding problems," 2011. UC Berkeley.