

Rate and Detection-Error Exponent Tradeoff for Joint Communication and Sensing of Fixed Channel States

Meng-Che Chang, Shi-Yuan Wang, Tuna Erdoan, and Matthieu R. Bloch

Abstract

We study the information-theoretic limits of joint communication and sensing when the sensing task is modeled as the estimation of a discrete channel state fixed during the transmission of an entire codeword. This setting captures scenarios in which the time scale over which sensing happens is significantly slower than the time scale over which symbol transmission occurs. The tradeoff between communication and sensing then takes the form of a tradeoff region between the rate of reliable communication and the state detection-error exponent. We investigate such tradeoffs for both mono-static and bi-static scenarios, in which the sensing task is performed at the transmitter or receiver, respectively. In the mono-static case, we develop an exact characterization of the tradeoff in open-loop, when the sensing is not used to assist the communication. We also show the strict improvement brought by a closed-loop operation, in which the sensing informs the communication. In the bi-static case, we develop an achievable tradeoff region that highlights the fundamentally different nature of the bi-static scenario. Specifically, the rate of communication plays a key role in the characterization of the tradeoff and we show how joint strategies, which simultaneously estimate message and state, outperform successive strategies, which only estimate the state after decoding the transmitted message.

I. INTRODUCTION

A core feature envisioned for the next generation of mobile networks is the convergence of communication and sensing [2], [3] (also known as integrated communication and sensing),

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA (email: {mchang301, shi-yuan.wang, etuna3}@gatech.edu, matthieu.bloch@ece.gatech.edu). This work has been supported by the National Science Foundation (NSF) under grants 1955401 and 2148400 as part of the Resilient & Intelligent NextG Systems (RINGS) program. Parts of the results have been presented at the 2022 IEEE International Symposium on Joint Communications & Sensing [1].

motivated in part by the need to offer detection and localization capabilities to interconnected devices interacting with the real world (robots, UAVs, etc.). This convergence is also enabled by the shift of communication frequencies towards the mmWave part of the spectrum, which allows a single Radio-Frequency (RF) hardware to perform both communication and sensing. While joint communication and sensing has already attracted interest in the context of joint communication and radar [4]–[6], theoretical, algorithmic, and hardware-related challenges remain and must be addressed to assess the true potential of joint communication and sensing [3], [7].

The objective of the present work is to further the theoretical understanding of the information-theoretic limits of joint communication and sensing, and specifically to better understand the tradeoffs incurred by a joint operation. While information-theoretic models often abstract fine channel modeling aspects, they provide valuable insights to identify the regimes in which tradeoffs exist and to quantify their severity. In particular, information-theoretic limits of joint communication and sensing are naturally approached from the perspective of joint channel transmission and channel state estimation, where the state simultaneously represents the quantity of interest for sensing and affects the communication. Early works [8]–[10] have considered state-dependent channel models with independent and identically distributed (i.i.d.) channel states, in which the encoder attempts to simultaneously communicate and facilitate the estimation of the state at the receiver, and have leveraged rate-distortion theory to characterize the optimal tradeoff between communication rate and state reconstruction accuracy. Recently, [11], [12] have revisited the model of [10] by shifting the task of estimating the state from the receiver to the transmitter using generalized feedback, a situation more in line with the scenarios envisioned in the context of joint communication and sensing, and characterized again the optimal tradeoff between rate and average state distortion. Extensions to multiple-access channels [13] and broadcast channels [12], [14] have been investigated, as well, although exact characterizations of the tradeoff remain elusive. Recent works have also considered secure joint communication and sensing to quantify and investigate the intrinsic information leakage associated to a joint operation [15]–[17].

A common feature of [10]–[14], [17] is that the i.i.d. nature of the channel state precludes any prediction. Consequently, state detection and estimation strategies are open-loop and the tradeoff between communication and sensing reduces to a resource allocation problem, in which the choice of a channel input distribution dictates the tradeoff.¹ Furthermore, the i.i.d. model is

¹The situation is more nuanced when introducing secrecy constraints as in [17] as feedback is known to improve secrecy.

not well-suited to applications in which the physical phenomena sensed, such as the presence or absence of an obstacle that would disrupt line-of-sight communication, evolve on a time scale that is much slower than the time scale at which communication symbols are transmitted. To address such applications, one may instead adopt a model in which the channel state remains *constant* over the block-length used for communication. Except in rare cases, estimating the channel state is then always possible so that the tradeoff between communication and sensing appears between the rate of communication and the accuracy of the state estimation. For the case of uncountably infinite channel states parameterized by continuous variables, accuracy can be captured by the Cramer-Rao bound [18]. For the case of finitely many channel states, accuracy may be captured by the asymptotic state detection error exponent, as studied thereafter and already reported in our preliminary results [1] concurrently with [19], [20].

Two kinds of joint communication and sensing models are considered in this paper, namely, mono-static and bi-static models, where we borrow these terminologies from mono-static and bi-static radars. In both models, there is a transmitter-receiver pair attempting to convey messages without exact knowledge of the channel state. The difference between these two models is that sensing is performed at the transmitter in the mono-static model and at the receiver in the bi-static model. Since the transmitter always knows the codeword, sensing in the mono-static model can be done coherently with the knowledge of the transmitted waveform. In contrast, the receiver in the bi-static model needs to simultaneously sense the channel and decode the message (or waveform).

In what follows, we study the rate/detection-error exponent tradeoff for both mono-static and bi-static models [2] with fixed channel states. Accordingly, our approach draws on the extensive literature on controlled sensing [21]–[23] and channel estimation with pilot sequences [24], [25]. Our specific contributions are: 1) we characterize the exact rate-detection exponent tradeoff for mono-static open-loop joint communication and sensing;² 2) we show a strict improvement of the rate-detection exponent tradeoff for mono-static closed-loop joint communication and sensing through learning the channel state and adapting the channel code; 3) we provide a partial characterization of the rate-detection exponent tradeoff for bi-static open-loop joint communication and sensing, and show that jointly detecting the state and decoding the code strictly outperforms

²The concurrent works [19], [20] developed characterizations similar to ours in more restricted settings; the results therein are subsumed by Theorem 3 of the present document and [1, Theorem 1].

the naive successive method, in which the detection of the state follows the decoding of the channel code. We further illustrate the results with simple examples that capture the essence of realistic joint communication and sensing scenarios.

The remaining of the paper is organized as follows. We introduce the mono-static and bi-static models for joint communication and sensing in Section III. We present our main results in Section IV, along with numerical examples illustrating the joint communication and sensing tradeoffs. We relegate all proofs to Section V to streamline the presentation.

II. NOTATION

For any set Ω , the indicator function is defined as $\mathbf{1}(\omega \in \Omega) = 1$ if $\omega \in \Omega$ and 0 otherwise. For any discrete set \mathcal{X} , $\mathcal{P}_{\mathcal{X}}$ is the set of all probability distributions on \mathcal{X} . For $n \in \mathbb{N}^*$, a sequence of length n is implicitly denoted $\mathbf{x} \triangleq (x_1, \dots, x_n) \in \mathcal{X}^n$, while $x^i \triangleq (x_1, \dots, x_i) \in \mathcal{X}^i$ denotes a sequence of length i , and $x_j^i = (x_j, \dots, x_i)$ is a sub-sequence of \mathbf{x} . For $\mathbf{x} \in \mathcal{X}^n$, $\hat{p}_{\mathbf{x}}$ denotes the type of \mathbf{x} , i.e., $\hat{p}_{\mathbf{x}}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i = x\}$. For any type P , \mathcal{T}_P^n is the corresponding type class, i.e., the set of all sequences $\mathbf{x} \in \mathcal{X}^n$ such that $\hat{p}_{\mathbf{x}} = P$. $\mathcal{P}_{\mathcal{X}}^n$ is the set of all possible types for length n sequences in \mathcal{X}^n . Let $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ be the set of all conditional probabilities of $Y \in \mathcal{Y}$ given $X \in \mathcal{X}$. Given a sequence $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{y} \in \mathcal{Y}^n$, we define $\hat{p}_{\mathbf{y}|\mathbf{x}}$ as the empirical conditional type, i.e., $\hat{p}_{\mathbf{y}|\mathbf{x}}(b|a) = \sum_{i=1}^n \mathbf{1}\{x_i = a, y_i = b\} / \sum_{i=1}^n \mathbf{1}\{x_i = a\}$ for all $a \in \mathcal{X}$ such that $\hat{p}_{\mathbf{x}}(a) > 0$ and $b \in \mathcal{Y}$. Let $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n$ be the set of all conditional types for length n sequences $\mathbf{x} \in \mathcal{X}^n$ and $\mathbf{y} \in \mathcal{Y}^n$. For any conditional type $P_{Y|X} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n$, we also define $\mathcal{T}_{|\mathbf{x}}(P_{Y|X})$ as the conditional type class of $P_{Y|X}$, i.e., the set of sequences $\mathbf{y} \in \mathcal{Y}^n$ such that $\hat{p}_{\mathbf{y}|\mathbf{x}} = P_{Y|X}$, and define $\mathcal{T}_{|\cdot}(P_{Y|X})$ as the set of sequences $\mathbf{x} \in \mathcal{X}^n$ such that $\hat{p}_{\mathbf{y}|\mathbf{x}} = P_{Y|X}$. Given two conditional distributions $W_{Y|X}$ and $P_{Y|X}$, we set $|W_{Y|X} - P_{Y|X}|_{\infty} \triangleq \max_{a \in \mathcal{X}, b \in \mathcal{Y}} |W_{Y|X}(b|a) - P_{Y|X}(b|a)|$. We let $\mathbb{H}(P_X) \triangleq -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x)$ be the entropy of $X \sim P_X$. If $W_{Y|X}$ is a conditional distribution on $Y \in \mathcal{Y}$ given $X \in \mathcal{X}$, $\mathbb{H}(W_{Y|X}|P_X) \triangleq \mathbb{E}_{P_X} [\mathbb{H}(W_{Y|X}(\cdot|X))]$ is the conditional entropy of $W_{Y|X}$ given an input distribution P_X and $\mathbb{I}(P_X, W_{Y|X}) \triangleq \mathbb{H}(W_{Y|X} \circ P_X) - \mathbb{H}(W_{Y|X}|P_X)$ is the mutual information between X and Y , where $X \sim P_X$ and $Y \sim P_X \circ W_{Y|X} \triangleq \sum_x P_X(x) W_{Y|X}(\cdot|x)$. The relative entropy between $P_{Y|X}$ and $W_{Y|X}$ given an input distribution P_X is $\mathbb{D}(P_{Y|X} \| W_{Y|X}|P_X) \triangleq \mathbb{E}_{P_X} [\mathbb{D}(P_{Y|X}(\cdot|X) \| W_{Y|X}(\cdot|X))]$. Throughout the paper, \log is with respect to (w.r.t.) base e , and therefore all the information quantities should be understood in *nats*. Moreover, for $a, b \in \mathbb{R}$ such that $\lfloor a \rfloor \leq \lceil b \rceil$, we define $[a; b] \triangleq \{\lfloor a \rfloor, \lfloor a \rfloor + 1, \dots, \lceil b \rceil - 1, \lceil b \rceil\}$; otherwise $[a; b] \triangleq \emptyset$. In addition, for any $x \in \mathbb{R}$, we let $|x|^+ \triangleq \max(x, 0)$.

III. JOINT COMMUNICATION AND SENSING MODELS

A. Mono-Static Model

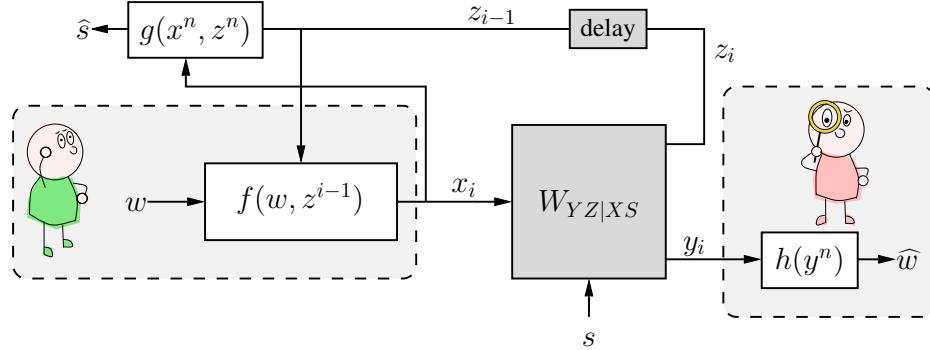


Fig. 1: Mono-static joint communication and sensing model.

The mono-static joint communication and sensing model is illustrated in Fig. 1, in which a transmitter attempts to communicate with a receiver over a state-dependent Discrete Memoryless Channel (DMC), also known as a compound channel, while simultaneously probing the channel state in a strictly causal manner through a sensing channel. Specifically, the transmitter encodes a uniformly distributed message $W \in [1; M]$ into a length n codeword X^n , of which symbols are transmitted over a DMC with transition probability $W_{YZ|XS}$. The state S , a priori unknown to both the transmitter and the receiver, is assumed to be *fixed* during the whole duration of the transmission and takes value in a finite set \mathcal{S} . The transmitter has the ability to estimate the channel state by using past observations obtained from the output Z of the DMC, allowing it to adapt its transmission in an online fashion. We assume in this paper that Chernoff information between channels $W_{Z|X,s}$ and $W_{Z|X,s'}$ is non-zero for all $s \neq s'$, i.e.,

$$\max_{P_X \in \mathcal{P}_{\mathcal{X}}} \max_{\ell \in [0,1]} - \sum_x P_X(x) \log \left(\sum_z W_{Z|X,s}(z|x)^\ell W_{Z|X,s'}(z|x)^{1-\ell} \right) > 0.$$

Formally, the encoder consists of a set of functions

$$f_i^{(m)} : [1; M] \times \mathcal{Z}^{i-1} \rightarrow \mathcal{X} : (w, z^{i-1}) \mapsto x_i \triangleq f_i^{(m)}(w, z^{i-1})$$

defined for every $i \in [1; n]$, while the state estimator is a function

$$g^{(m)} : \mathcal{X}^n \times \mathcal{Z}^n \rightarrow \mathcal{S} : (x^n, z^n) \mapsto \hat{s}.$$

The message decoder is a function

$$h^{(m)} : \mathcal{Y}^n \rightarrow [1; M] : y^n \mapsto \hat{w}.$$

A code $\mathcal{C}^{(m)}$ then consists of the tuple $(\{f_i^{(m)}\}_{i \in [1; n]}, g^{(m)}, h^{(m)})$, as well as the implicitly defined associated message set $[1; M]$.

B. Bi-Static Model

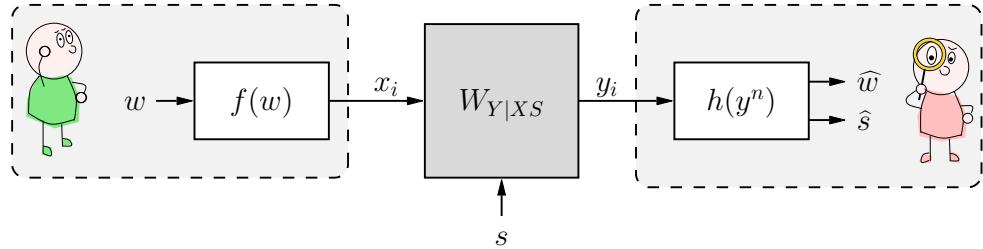


Fig. 2: Bi-static joint communication and sensing model.

As illustrated in Fig. 2, the bi-static joint communication and sensing model differs from the mono-static one in that the receiver should simultaneously sense the state and decode the message. Specifically, the transmitter encodes a uniformly distributed message $W \in [1; M]$ into a length n codeword X^n , of which symbols are transmitted over a DMC with transition probability $W_{Y|XS}$. We again assume that the a priori unknown state S is *fixed* during the whole duration of the transmission and takes value in a finite set \mathcal{S} . The encoder is then defined as

$$f^{(b)} : [1; M] \rightarrow \mathcal{X}^n : w \mapsto \mathbf{x}_w,$$

while the decoder and the state estimator are the functions

$$g^{(b)} : \mathcal{Y}^n \rightarrow \mathcal{S} : y^n \mapsto \hat{s},$$

and

$$h^{(b)} : \mathcal{Y}^n \rightarrow [1; M] : y^n \mapsto \hat{w}.$$

The code $\mathcal{C}^{(b)}$ in the bi-static model then consists of the tuple $(f^{(b)}, g^{(b)}, h^{(b)})$, as well as the message set $[1; M]$.

Remark 1. A key difference between the mono-static model and the bi-static model is that the former reduces the coupling of performance between communication and sensing to the type of

the transmitted codewords. Specifically, the mono-static model endows the state estimator with full knowledge of the transmitted codeword, so that correct decoding is irrelevant to the state detection performance. In contrast, the bi-static model requires the receiver to perform a joint estimation of the codeword and of the state.

Remark 2. The model of Fig. 1 differs from the ones in [10], [12], in which the state is i.i.d. and changing from symbol to symbol. Our model captures a scenario in which the coherence time of the state is much longer than the duration of a transmission. Since the state does not change during the transmission, the transmitter can gradually obtain an accurate estimation about the state and is able to adapt the transmission scheme according to the estimated channel state. As a result our model also captures the ability to adapt to the channel state in an online fashion, while the models in [10], [12] only allow for an offline adaptation based on a target rate/distortion pair. Neither model supersedes the other and both capture scenarios that could be relevant to next generation communication networks.

C. Performance Evaluation Metrics

The performance of the system is measured in terms of the asymptotic rate of reliable communication and asymptotic detection-error exponent. Formally, we define the communication-error probability and the detection-error probability for both mono-static and bi-static cases as follows

- Mono-static case:

$$P_c^{(n)} \triangleq \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(h^{(m)}(Y^n) \neq w | W = w, S = s), \quad (1)$$

$$P_d^{(n)} \triangleq \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(m)}(X^n, Z^n) \neq s | W = w, S = s). \quad (2)$$

- Bi-static case:

$$P_c^{(n)} \triangleq \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(h^{(b)}(Y^n) \neq w | W = w, S = s), \quad (3)$$

$$P_d^{(n)} \triangleq \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(b)}(Y^n) \neq s | W = w, S = s). \quad (4)$$

The rate and the detection-error exponent for both mono-static and bi-static models are

$$R^{(n)} \triangleq \frac{1}{n} \log M \text{ and } E_d^{(n)} \triangleq -\frac{1}{n} \log P_d^{(n)},$$

respectively.

Definition 1 (Achievability). A rate/detection-error exponent (R, E) is achievable in the mono-static (bi-static) joint communication model if for any $s \in \mathcal{S}$ and any $\epsilon > 0$, there exist a sufficiently large n and a code $\mathcal{C}^{(m)}$ ($\mathcal{C}^{(b)}$) of length n such that

$$P_c^{(n)} \leq \epsilon, \quad (5)$$

$$E_d^{(n)} \geq E - \epsilon, \quad (6)$$

$$R^{(n)} \geq R - \epsilon. \quad (7)$$

When the encoder does not perform any online adaptation so that $f_i^{(m)} : [1; M] \mapsto \mathcal{X}$ is independent of the observation Z^{i-1} , the scheme is called *open-loop*. On the other hand, if the encoder utilizes feedback information, the scheme is called *closed-loop*. Our objective is to characterize the set of all achievable rate/detection-error exponent pairs by open-loop strategies in both mono-static and bi-static models as well as the set of all achievable rate/detection-error exponent pairs by closed-loop strategies in the mono-static model.

Definition 2. We define $\mathbf{C}_{\text{open}}^{(m)}$ and $\mathbf{C}^{(b)}$ as the closure of all achievable rate/detection-error exponent pairs by open strategies in the mono-static and bi-static model, respectively. Similarly, we define $\mathbf{C}_{\text{close}}^{(m)}$ as the closure of all achievable rate/detection-error exponent pairs by closed-loop strategies in the mono-static model.

IV. MAIN RESULTS

A. Mono-Static Model

We first restrict ourselves to open-loop schemes, which provide a baseline for assessing the usefulness of adaptation. For simplicity, we denote in this case the encoder that maps a message w to a codeword of n symbols by $f^{(m)} : [1; M] \mapsto \mathcal{X}^n$. The following theorem provides an exact characterization of $\mathbf{C}_{\text{open}}^{(m)}$.

Theorem 3. The closure of all achievable joint communication and sensing rate/detection error exponent pairs for mono-static open-loop schemes is

$$\mathbf{C}_{\text{open}}^{(m)} = \bigcup_{P_X \in \mathcal{P}_{\mathcal{X}}} \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s}) \\ E \leq \phi(P_X) \end{array} \right\} \quad (8)$$

where

$$\phi(P_X) = \min_{s \in \mathcal{S}} \min_{s' \neq s} \max_{\ell \in [0,1]} - \sum_x P_X(x) \log \left(\sum_z W_{Z|X,s}(z|x)^\ell W_{Z|X,s'}(z|x)^{1-\ell} \right). \quad (9)$$

Proof. See Section V-A and Section V-B. \square

A couple of comments are in order. First, since open-loop schemes do not exploit the information about the state contained in past noisy observations of the state, achievable rates are necessarily upper bounded by the compound channel capacity $\max_{P_X} \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s})$. This is a weakness of all open-loop schemes. Second, because of the open-loop nature of the coding schemes, the interplay between communication and sensing is captured by the choice of a distribution P_X that governs the empirical statistics of the codewords and is set *offline*. This is similar to what is obtained in other information-theoretic approaches based on rate-distortion [10], [12].

The results in [19] are also special cases of Theorem 3. In [19], the authors consider a mono-static joint communication and sensing model in which $W_{YZ|XS} = W_{Y|X}W_{Z|X,S}$, i.e., the communication channel is irrelevant to the state. One of the channel models in [19] is a binary setting in which $\mathcal{X} = \mathcal{Z} = \mathcal{S} = \{0, 1\}$ and $W_{Z|X,S} = W_{Z|X,S}$, i.e., at each time t the state-estimator obtains $Z_t = X_t \cdot S \oplus N_t$, where \oplus denotes the modulo-2 sum and $N_t \sim \text{Ber}(q)$ for some $0 < q < 1$. By specializing Theorem 3, one recovers [19, Theorem 1] as follows.

Corollary 4. *Let $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathcal{S} = \{0, 1\}$. When the mono-static joint communication model satisfies $W_{YZ|XS} = W_{Y|X}W_{Z|X,S}$, where $W_{Y|X}$ and $W_{Z|X,S}$ are binary symmetric channels with cross over probability p and q , respectively, then*

$$\mathbf{C}_{\text{open}}^{(m)} = \bigcup_{\alpha \in [0.5, 1]} \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \mathbb{H}(\text{Ber}(\alpha * p)) - \mathbb{H}(\text{Ber}(p)) \\ E \leq \alpha \mathbb{D}(\text{Ber}(0.5) || \text{Ber}(q)), \end{array} \right\} \quad (10)$$

where $\alpha * p = \alpha(1 - p) + (1 - \alpha) * p$.

Proof. Since, $W_{YZ|XS} = W_{Y|X}W_{Z|X,S}$, the achievable rate is irrelevant to the state, and for each $P_X \sim \text{Ber}(\alpha)$ we have

$$\begin{aligned} \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s}) &= \mathbb{I}(P_X, W_{Y|X}) \\ &= \mathbb{H}(P_X \circ W_{Y|X}) - \mathbb{H}(W_{Y|X} | P_X) \end{aligned}$$

$$= \mathbb{H}(\text{Ber}(\alpha * p)) - \mathbb{H}(\text{Ber}(p)).$$

Moreover, for such P_X , $\phi(P_X)$ can be calculated as follows.

$$\begin{aligned} \phi(P_X) &= \min_{s \in \mathcal{S}} \min_{s' \neq s} \max_{\ell \in [0,1]} - \sum_x P_X(x) \times \log \left(\sum_z W_{Z|X,s}(z|x)^\ell W_{Z|X,s'}(z|x)^{1-\ell} \right) \\ &= \max_{\ell \in [0,1]} -P_X(1) \log (q^\ell (1-q)^{1-\ell} + (1-q)^\ell q^{1-\ell}) \end{aligned} \quad (11)$$

$$= -P_X(1) \log (q^{1/2} (1-q)^{1/2} + (1-q)^{1/2} q^{1/2}) \quad (12)$$

$$= \alpha \mathbb{D}(\text{Ber}(0.5) \parallel \text{Ber}(q)),$$

where (12) follows from the fact that $\ell = 1/2$ maximizes (11). The corollary follows by taking the union over all P_X . \square

Remark 3. Note that in the setting considered in Corollary 4, transmitting $X = 0$ does not help the performance of the state-estimation, and the detection-error exponent is a monotonously increasing function of the weight of the codeword.

We also observe that in some cases, there is no tradeoff between maximizing the communication capacity and the detection-error exponent.

Corollary 5. If there exists $x_0 \in \mathcal{X}$ such that for all $x \in \mathcal{X}$ there exists a permutation π_x on \mathcal{Z} such that for every $s \in \mathcal{S}$

$$W_{Z|X,s}(z|x) = W_{Z|X,s}(\pi_x(z)|x_0), \quad (13)$$

then

$$\mathbf{C}_{\text{open}}^{(m)} = \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \max_{P_X} \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s}) \\ E \leq \max_{P_X} \phi(P_X) \end{array} \right\} \quad (14)$$

where $\phi(\cdot)$ is defined in (9).

Proof. For every $x \in \mathcal{X} \setminus \{x_0\}$,

$$\sum_{z \in \mathcal{Z}} W_{Z|X,s}(z|x)^\ell W_{Z|X,s'}(z|x)^{1-\ell} = \sum_{z \in \mathcal{Z}} W_{Z|X,s}(\pi_x(z)|x_0)^\ell W_{Z|X,s'}(\pi_x(z)|x_0)^{1-\ell} \quad (15)$$

$$= \sum_{\pi_x^{-1}(z) \in \mathcal{Z}} W_{Z|X,s}(z'|x_0)^\ell W_{Z|X,s'}(z'|x_0)^{1-\ell} \quad (16)$$

$$= \sum_{z \in \mathcal{Z}} W_{Z|X,s}(z|x_0)^\ell W_{Z|X,s'}(z|x_0)^{1-\ell}. \quad (17)$$

Thus, we know that the detection-error exponent is invariant to the input type under this scenario. \square

In other words, when the channel satisfies certain symmetry conditions, there is no tradeoff between rate and detection-error exponent and one simultaneously achieves the optimal communication rate and the optimal detection performance. One of the compound channel families that falls into such a category is the set of Binary Symmetric Channels (BSCs). The maximal detection-error exponent and the compound capacity are then simultaneously achieved with a uniform input distribution.

We now turn our attention back to the characterization of *closed-loop schemes*, which exploit the feedback to adapt to the state. The next theorem characterizes an inner bound of the set $\mathbf{C}_{\text{closed}}^{(m)}$.

Theorem 6. *The closure of all achievable joint communication and sensing rate/detection error exponent pairs for mono-static closed-loop schemes satisfies*

$$\mathbf{C}_{\text{closed}}^{(m)} \supseteq \bigcup_{\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in (\mathcal{P}_{\mathcal{X}})^{|\mathcal{S}|}} \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \min_{s \in \mathcal{S}} \mathbb{I}(P_{X,s}, W_{Y|X,s}) \\ E \leq \min_{s \in \mathcal{S}} \phi(P_{X,s}) \end{array} \right\} \quad (18)$$

where the notation $\bigcup_{\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in (\mathcal{P}_{\mathcal{X}})^{|\mathcal{S}|}}$ means that we are taking the union over all possible $|\mathcal{S}|$ -tuples of probability distributions in $\mathcal{P}_{\mathcal{X}}$ and $(\mathcal{P}_{\mathcal{X}})^{|\mathcal{S}|}$ is the set of tuples of $|\mathcal{S}|$ elements in $\mathcal{P}_{\mathcal{X}}$.

Theorem 6 is obtained by considering a simple strategy in which the transmitter learns the state, informs the receiver, and uses a code adapted to the learned channel state. The exact characterization of the optimal tradeoffs for closed-loop schemes remains elusive and presents non-trivial challenges, chief among them the absence of a known optimal detection error-exponent for multi-hypothesis controlled sensing [21]. One can conclude that the maximal achievable detection-error exponent characterized by Theorem 6 is identical to that of the open-loop strategy by observing the following equality

$$\max_{\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in (\mathcal{P}_{\mathcal{X}})^{|\mathcal{S}|}} \min_{s \in \mathcal{S}} \phi(P_{X,s}) = \max_{P_X \in \mathcal{P}_{\mathcal{X}}} \phi(P_X).$$

Therefore, the region characterized in Theorem 6 is sub-optimal because it is already shown in [21] that there exists a closed-loop method that achieves a better detection-error exponent than an open-loop scheme. However, the benefit of Theorem 6 is in improving the communication rate from the compound channel capacity, i.e.,

$$\max_{P_X \in \mathcal{P}_{\mathcal{X}}} \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s})$$

to the worst-case capacity, i.e.,

$$\min_{s \in \mathcal{S}} \max_{P_X \in \mathcal{P}_{\mathcal{X}}} \mathbb{I}(P_X, W_{Y|X,s}).$$

Note that the worst-case capacity is exactly the maximal achievable rate described in Theorem 6 because

$$\max_{\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in (\mathcal{P}_{\mathcal{X}})^{|\mathcal{S}|}} \min_{s \in \mathcal{S}} \mathbb{I}(P_{X,s}, W_{Y|X,s}) = \min_{s \in \mathcal{S}} \max_{P_X \in \mathcal{P}_{\mathcal{X}}} \mathbb{I}(P_X, W_{Y|X,s}).$$

A similar technique exploiting the feedback to improve the capacity in a compound channel can be found in [26].

Remark 4. *The regions characterized in Theorem 6 and Theorem 3 are identical when the condition in Corollary 5 is satisfied and the worst-case capacity is equal to the compound channel capacity.*

B. Bi-Static Model

In joint communication and sensing, codewords convey information and induce a codeword-dependent distribution from which observations are generated and are used to estimate the state. In the mono-static model, state estimation is performed at the transmitter, which has the access to the transmitted codeword, so that a likelihood-based state estimation can be performed with exact knowledge of the transmitted codeword. In the bi-static model, however, the transmitted codeword is unknown to the state estimator $g^{(b)}$.

A natural method to estimate the state is to use a successive approach, by which one first estimates the codeword and then performs a maximum likelihood estimation given the decoded codeword. Denoting by $\mathbf{C}^{(b)}$ the closure of all achievable regions in the bi-static model, our next theorem provides an inner bound of $\mathbf{C}^{(b)}$ based on successive schemes.

Theorem 7. *The closure region of all achievable joint communication and sensing rate/detection error exponent pairs for the bi-static joint communication and sensing model satisfies*

$$\mathbf{C}^{(b)} \supseteq \mathcal{D}_{\text{succ}} \triangleq \bigcup_{P_X \in \mathcal{P}_{\mathcal{X}}} \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s}) \\ E \leq \min(\rho_{\text{succ}}(P_X, R), \phi(P_X)) \end{array} \right\}, \quad (19)$$

where

$$\rho_{\text{succ}}(P_X, R) \triangleq \min_{s \in \mathcal{S}} \min_{\hat{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}} \left(\mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \left| \mathbb{I}(P_X; \hat{P}) - R \right|^+ \right). \quad (20)$$

Proof. Theorem 7 is a direct result of the following inequality

$$\begin{aligned} P_d^{(n)} &= \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(b)}(Y^n) \neq s | W = w, S = s) \\ &= \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(b)}(Y^n) \neq s | h^{(b)} = w, W = w, S = s) \mathbb{P}(h^{(b)} = w | W = w, S = s) \\ &\quad + \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(b)}(Y^n) \neq s | h^{(b)} \neq w, W = w, S = s) \mathbb{P}(h^{(b)} \neq w | W = w, S = s) \\ &\leq \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(b)}(Y^n) \neq s | h^{(b)} = w, W = w, S = s) \\ &\quad + \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(h^{(b)}(Y^n) \neq w | W = w, S = s), \end{aligned}$$

where we have bounded the probability of making a wrong state estimation by 1 when the codeword is decoded incorrectly. Note that the term

$$\max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(h^{(b)}(Y^n) \neq w | W = w, S = s)$$

is the communication-error probability; it has been shown in [27, Theorem 10.2] that, for all P_X , there exists a constant composition code with type P_X such that the communication-error exponent is lower bounded by $\rho_{\text{succ}}(P_X)$. Moreover, when the codeword is correctly decoded, the exponent of

$$\max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(b)}(Y^n) \neq s | h^{(b)} = w, W = w, S = s)$$

can be lower bounded by the Chernoff information $\phi(P_X)$ as shown in Lemma 10 in Section V and proofs therein. \square

Unfortunately, the loss induced by upper bounding the term $\mathbb{P}(g^{(b)}(Y^n) \neq s | h^{(b)} \neq w, W = w, S = s)$ by 1 might result in a loose bound on the state exponent. The difficulty of analyzing the term $\mathbb{P}(g^{(b)}(Y^n) \neq s | h^{(b)} \neq w, W = w, S = s)$ comes from the fact that the receiver does

not know the conditional type $\hat{p}_{\mathbf{y}|\mathbf{x}_w}$. By leveraging random constant composition codes and joint decoding/detection, the following result offers improvements.

Theorem 8. *The closure region of all achievable joint communication and sensing rate/detection error exponent pairs for the bi-static joint communication and sensing model satisfies*

$$\mathbf{C}^{(b)} \supseteq \mathcal{D}_{\text{joint}} \triangleq \bigcup_{P_X \in \mathcal{P}_{\mathcal{X}}} \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s}) \\ E \leq \min(\rho_{\text{joint}}(P_X, R), \phi(P_X)) \end{array} \right\}, \quad (21)$$

where

$$\rho_{\text{joint}}(P_X, R) \triangleq \min_{s \in \mathcal{S}} \min_{\hat{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}} \left(\mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\hat{P}, P_X, R)} \mathbb{I}(P_X, P') - R \right|^+ \right), \quad (22)$$

and

$$\begin{aligned} \mathcal{P}_{s,s'}(\hat{P}, P_X, R) &\triangleq \left\{ P' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} : \mathbb{D}(P' \| W_{Y|X,s'} | P_X) + \mathbb{H}(P' | P_X) \right. \\ &\leq \min \left(\beta(\hat{P}, P_X, R, s), \mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\hat{P} | P_X) \right), P_X \circ P' = P_X \circ \hat{P} \left. \right\}, \end{aligned}$$

where

$$\beta(\hat{P}, P_X, R, s) \triangleq \min_{\substack{P'' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} : \mathbb{I}(P_X, P'') < R, \\ P_X \circ P'' = P_X \circ \hat{P}}} \mathbb{D}(P'' \| W_{Y|X,s} | P_X) + \mathbb{H}(P'' | P_X). \quad (23)$$

The calculation of $\rho_{\text{joint}}(P_X, R)$ in (22) involves a minimization over $\hat{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ and $P' \in \mathcal{P}_{s,s'}(\hat{P}, P_X, R)$. In Corollary 9, we provide a lower bound on $\rho_{\text{joint}}(P_X, R)$ to simplify the expression and show that it is greater than or equal to the compound channel communication-error exponent $\rho_{\text{succ}}(P_X, R)$.

Corollary 9. *For all P_X , it holds that $\rho_{\text{joint}}(P_X, R) \geq \rho_{\text{succ}}(P_X, R)$, and hence, $\mathcal{D}_{\text{joint}} \supseteq \mathcal{D}_{\text{succ}}$. Moreover, $\rho_{\text{joint}}(P_X, R)$ is lower bounded by*

$$\rho_{\text{joint}}(P_X, R) \geq \min_s \min \left(\min_{\hat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \gamma_1(\hat{P}, R, s), \min_{\hat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \gamma_2(\hat{P}, R, s) \right), \quad (24)$$

where for all \hat{P} and s ,

$$\gamma_1(\hat{P}, R, s) \triangleq \mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \left| \mathbb{I}(P_X, \hat{P}) - R \right|^+$$

$$\gamma_2(\widehat{P}, R, s) \triangleq \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(P_X \circ \widehat{P}) - \beta(\widehat{P}, P_X, R, s) - R \right|^+,$$

and

$$\begin{aligned} \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s) &\triangleq \{\widehat{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} : \max_{s'' \in \mathcal{S}} \beta(\widehat{P}, P_X, R, s'') \geq \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\widehat{P} | P_X)\} \\ \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s) &\triangleq \{\widehat{P} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} : \max_{s'' \in \mathcal{S}} \beta(\widehat{P}, P_X, R, s'') < \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\widehat{P} | P_X)\}. \end{aligned}$$

Note that $\rho_{\text{succ}}(P_X, R)$ can be expressed as minimizing $\gamma_1(\widehat{P}, R, s)$ over all possible \widehat{P} and all s . Therefore, we can show that ρ_{joint} is larger than ρ_{succ} if $\gamma_2(\widehat{P}, R, s) \geq \gamma_1(\widehat{P}, R, s)$ for all $\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)$. When $\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)$, one can observe that

$$\mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(P_X \circ \widehat{P}) - \mathbb{I}(P_X, \widehat{P}) = \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\widehat{P} | P_X)$$

is greater than $\beta(P_X, R, s)$ and hence, $\gamma_2(\widehat{P}, R, s) \geq \gamma_1(\widehat{P}, R, s)$. Therefore,

$$\rho_{\text{joint}}(P_X, R) \geq \min_s \min \left(\min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, s)} \gamma_1(\widehat{P}, R, s), \min_{\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, s)} \gamma_1(\widehat{P}, R, s) \right) \quad (25)$$

$$= \rho_{\text{succ}}(P_X, R). \quad (26)$$

Remark 5. It is known that the communication-error exponent is zero when one transmits at the rate of compound channel capacity. Recall that ρ_{succ} is calculated by upper bounding the detection-error probability by 1 when the message decoding error happens. Therefore, ρ_{succ} is also zero when $R = \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s})$. In contrast, ρ_{joint} can be positive even when $R = \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s})$. This can be seen from the expression of $\gamma_2(\widehat{P}, R, s)$ in Corollary 9, and we will illustrate this by an example given in the next sub-section.

C. Numerical Examples

a) Mono-Static Model

We first consider the channel $W_{YZ|X,S}$ defined in Table I. In this example, $\mathcal{Y} = \mathcal{Z} = \mathcal{X} = \{0, 1\}$ and $W_{Z|X,S} = W_{Y|X,S}$. Note that transmitting the symbol $X = 1$ is most useful to identify the channel state because this is the situation in which the output distributions corresponding to different states are most distinguishable. However, if the transmitter only transmits $X = 1$, the communication rate would become zero as seen in Fig. 3a. Therefore, the tradeoff between maximizing the detection-error exponent and maximizing the communication rate can be clearly

seen in Fig. 3a. On the other hand, the difference between $\mathbf{C}_{\text{open}}^{(m)}$ and the inner bound of $\mathbf{C}_{\text{closed}}^{(m)}$ given in Theorem 6 is shown in Fig. 3b, illustrating how the communication rate is increased. Indeed, the inner bound region for $\mathbf{C}_{\text{closed}}^{(m)}$ is larger than $\mathbf{C}_{\text{open}}^{(m)}$ because the compound capacity is here strictly less than the worst-case capacity.

In contrast to the channel given in Table I, for which the tradeoff between sensing and communication exists, the channel given in Table II is a BSC for which, according to Corollary 5, the best error exponent of detection is always achieved regardless of the type of codewords. The result of Theorem 3 for this channel is illustrated in Fig. 4.

TABLE I: Table for $W_{Z|X,S}(0) = W_{Y|X,S}(0)$ for all $X \in \{0, 1\}$ and $S \in \{0, 1, 2\}$.

		X	
		0	1
S	0	0.95	0.45
	1	0.9	0.2
	2	0.5	0.03

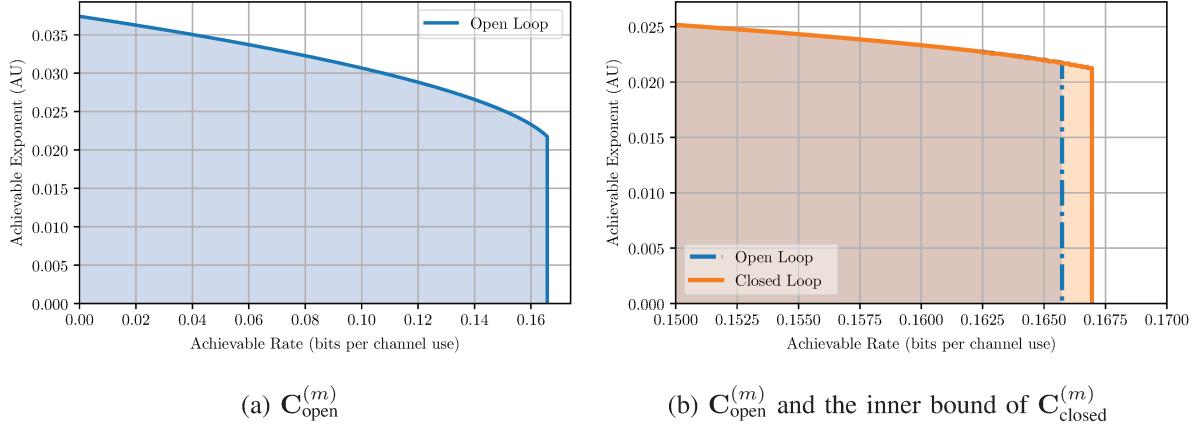


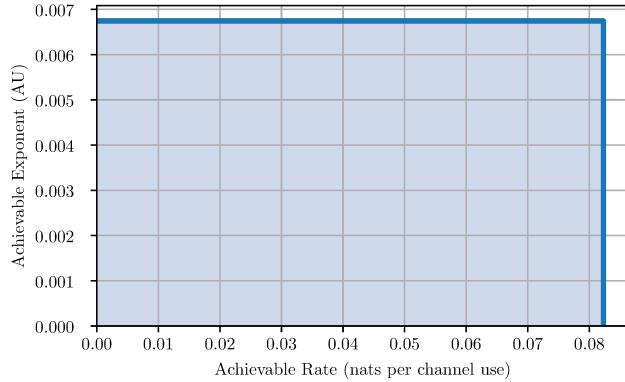
Fig. 3: Closure region $\mathbf{C}_{\text{open}}^{(m)}$ and inner bounds for $\mathbf{C}_{\text{closed}}^{(m)}$ corresponding to the channel of Table I.

b) Bi-Static Model

For the bi-static model, we present a numerical example as defined in Table III. Here, $\mathcal{Y} = \mathcal{Z} = \mathcal{X} = \{0, 1\}$ and $\mathcal{S} = \{0, 1\}$. When $\mathcal{S} = 0$, the channel $W_{Y|X,0}$ is a BSC with cross-over probability 0.3, and $W_{Y|X,1}$ is a BSC with cross-over probability 0.6. From Fig. 5, one observes that $\mathcal{D}_{\text{joint}}$ is strictly larger than $\mathcal{D}_{\text{succ}}$ especially when the code rate is high. When the rate is low, the exponent ρ_{joint} is dominated by the Chernoff information $\phi(P_X)$, which is approximately

TABLE II: Table for $W_{Z|X,S}(0) = W_{Y|X,S}(0)$ for all $X \in \{0, 1\}$ and $S \in \{0, 1, 2\}$.

$S \backslash X$	0	1
0	0.9	0.1
1	0.8	0.2
2	0.7	0.3

Fig. 4: Closure region $C_{\text{open}}^{(m)}$ corresponding to the channel of Table II.

0.048 in this example. When the rate is high, the exponent is dominated by a rate-dependent term, capturing the fact the receiver's inability to easily decode the transmitted sequence is what drives performance. By the definition of ρ_{succ} , the value of ρ_{succ} is zero when the code rate is at the compound channel capacity. However, the joint detection scheme reaches a positive exponent even when transmitting at the compound capacity, highlighting the benefits of joint detection over successive decoding.

TABLE III: Table for $W_{Y|X,S}(0)$ for all $X \in \{0, 1\}$ and $S \in \{0, 1\}$.

$S \backslash X$	0	1
0	0.7	0.3
1	0.4	0.6

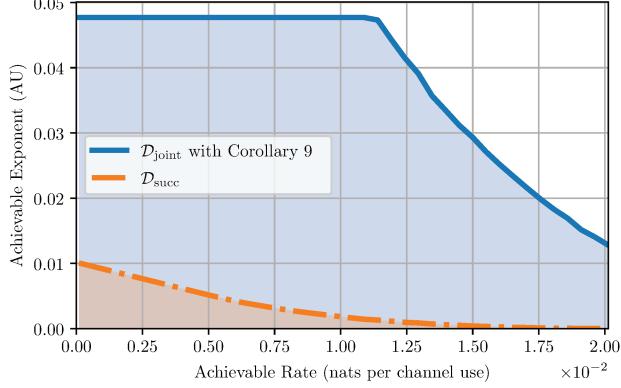


Fig. 5: Region $\mathcal{D}_{\text{succ}}$ and $\mathcal{D}_{\text{joint}}$ corresponding to Table III.

V. DETAILED PROOFS

A. Achievability Proof of Theorem 3

We show that all (R, E) pairs within the region $\mathbf{C}_{\text{open}}^{(m)}$ are achievable. Since we restrict ourselves to open-loop schemes, we may fix P_X as the type of all codewords. Fix any $\epsilon > 0$. By [27, Theorem 10.2], there exists a code with encoder $f^{(m)}$ and decoder $h^{(m)}$ such that $f^{(m)}(w) \in \mathcal{T}_{P_X}^n$ for all w , the rate is at least $\min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s}) - \epsilon$, and $\max_w \mathbb{P}(h(Y^n) \neq w | S = s) < \epsilon$ for all $s \in \mathcal{S}$. Then, the detection-error exponent $\phi(P_X)$ is given by Lemma 10 adapted from [21, Theorem 1].

Lemma 10. *Suppose that the codeword corresponding to the message $w \in [1; M]$ has type $P_X \in \mathcal{P}_X$, the conditional detection-error exponent $E_{d,w} \triangleq -\frac{1}{n} \log \max_{s \in \mathcal{S}} \mathbb{P}(g(Z^n) \neq s | S = s, W = w)$ in an open-loop scheme is asymptotically upper bounded by*

$$\phi(P_X) \triangleq \min_s \min_{s' \neq s} \max_{\ell \in [0,1]} - \sum_x P_X(x) \log \left(\sum_z W_{Z|X,s}(z|x)^\ell W_{Z|X,s'}(z|x)^{1-\ell} \right).$$

Moreover, it is also asymptotically achievable by a maximum likelihood estimator g_{ML} . Specifically,

$$\max_{s \in \mathcal{S}} \mathbb{P}(g_{\text{ML}}(Z^n) \neq s | S = s, W = w) \leq \Theta_n(1) e^{-\phi(P_X)} \quad (27)$$

for all $w \in [1; M]$, and for all $s \in \mathcal{S}$

$$\mathbb{P}(g_{\text{ML}}(Z^n) \neq s | S = s, W = w) \leq \Theta_n(1) e^{-\psi_s(P_X)}, \quad (28)$$

for all $w \in [1; M]$, where

$$\psi_s(P_X) \triangleq \min_{s' \neq s} \max_{\ell \in [0,1]} - \sum_x P_X(x) \log \left(\sum_z W_{Z|X,s}(z|x)^\ell W_{Z|X,s'}(z|x)^{1-\ell} \right). \quad (29)$$

Taking the union over all possible P_X leads to the region in (8).

B. Converse Proof of Theorem 3

Assume that the rate/detection-error exponent pair (R, E) is achievable. Then, for all $\epsilon > 0$, there exists n sufficiently large and a codebook $\mathcal{C}^{(m)}$ such that

$$\begin{aligned} \frac{\log |\mathcal{C}^{(m)}|}{n} &\geq R - \epsilon \\ \max_{s \in \mathcal{S}} \max_{w \in [1; M]} \mathbb{P}(h(Y^n) \neq w | W = w, S = s) &\leq \epsilon \\ -\frac{1}{n} \log \max_{s \in \mathcal{S}} \max_{w \in [1; M]} \mathbb{P}(g(Z^n) \neq s | W = w, S = s) &\geq E - \epsilon. \end{aligned}$$

Since there is at most a polynomial number of types, there exists a set of types \mathcal{T} such that, for all $P_X \in \mathcal{T}$, the subcode $\mathcal{C}_{P_X} \triangleq \{f(w) : \hat{p}_{f(w)} = P_X\} \subset \mathcal{C}^{(m)}$ satisfies

$$\max_{s \in \mathcal{S}} \max_{w \in f^{-1}(\mathcal{C}_{P_X})} \mathbb{P}(h(Y^n) \neq w | W = w, S = s) \leq \epsilon \quad (30)$$

and

$$\frac{\log |\mathcal{C}_{P_X}|}{n} > \frac{\log |\mathcal{C}^{(m)}|}{n} - \delta \geq R - \epsilon - \delta \quad (31)$$

for some δ vanishing with ϵ . Fix any $P_X \in \mathcal{T}$. Let

$$\tilde{P}_X^n(\mathbf{x}) \triangleq \frac{1}{|\mathcal{C}_{P_X}|} \sum_{\tilde{\mathbf{x}} \in \mathcal{C}_{P_X}} \mathbf{1}\{\mathbf{x} = \tilde{\mathbf{x}}\}.$$

Observe that

$$\overline{P}_X(x) \triangleq \frac{1}{n} \sum_{i=1}^n \tilde{P}_{X_i}(x) = P_X(x), \quad (32)$$

where \tilde{P}_{X_i} is the i -th marginal distribution of \tilde{P}_X^n . Then, for any state s ,

$$\begin{aligned} (1 - \epsilon) \log |\mathcal{C}_{P_X}| - h_b(\epsilon) &\stackrel{(a)}{\leq} \mathbb{I}\left(\tilde{P}_X^n, W_{Y|X,s}^{\otimes n}\right) \\ &\stackrel{(b)}{=} \mathbb{H}(Y^n) - \sum_{\mathbf{x} \in \mathcal{X}^n} \tilde{P}_X^n(\mathbf{x}) \mathbb{H}(Y^n | X^n = x^n, s) \\ &= \mathbb{H}(Y^n) - \sum_{\mathbf{x} \in \mathcal{X}^n} \tilde{P}_X^n(x^n) \sum_{i=1}^n \mathbb{H}(Y_i | X^n = x^n, Y^{i-1}, s) \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \mathbb{H}(Y_i | Y^{i-1}) - \sum_{i=1}^n \sum_{x_i \in \mathcal{X}} \tilde{P}_{X_i}(x_i) \mathbb{H}(Y_i | X_i = x_i, s) \\
&\leq \sum_{i=1}^n \mathbb{I}(\tilde{P}_{X_i}, W_{Y|X,s}) \\
&\stackrel{(c)}{\leq} n \mathbb{I}(\bar{P}_X, W_{Y|X,s})
\end{aligned} \tag{33}$$

where $h_b(\cdot)$ is the binary entropy function, (a) follows from Fano's inequality and standard techniques, (b) follows by identifying $Y^n \sim \tilde{P}_X^n \circ W_{Y|X,s}^{\otimes n}$ and the definition of mutual information, and (c) follows from the concavity of mutual information in the input distribution. Then, we obtain

$$\frac{\log |\mathcal{C}_{P_X}|}{n} \leq \frac{\mathbb{I}(P_X, W_{Y|X,s}) + \frac{1}{n}}{1 - \epsilon}, \tag{34}$$

where we have used (32).

Since (34) is valid for any ϵ and state s , the size of sub-codebook \mathcal{C}_{P_X} is upper bounded by the mutual information in a compound channel, i.e.,

$$\frac{\log |\mathcal{C}_{P_X}|}{n} < \min_{s \in \mathcal{S}} \mathbb{I}(P_X, W_{Y|X,s}) + \tau \tag{35}$$

for some $\tau > 0$ vanishing with ϵ . On the other hand, for this $P_X \in \mathcal{T}$,

$$\begin{aligned}
&E - \epsilon \\
&\leq -\frac{1}{n} \log \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g(Z^n) \neq s | W = w, S = s) \\
&\stackrel{(a)}{\leq} -\frac{1}{n} \log \max_{s \in \mathcal{S}} \max_{w \in f^{-1}(\mathcal{C}_{P_X})} \mathbb{P}(g(Z^n) \neq s | W = w, S = s) \\
&\stackrel{(b)}{\leq} \phi(P_X) + \delta,
\end{aligned} \tag{36}$$

where (a) follows by restricting the set to only the terms corresponding to messages in \mathcal{C}_{P_X} ; (b) follows since the detection error is upper bounded by $\phi(P_X)$ for any message with type P_X by Lemma 10. Combining (35) and (36) and choosing $P_X^* \triangleq \operatorname{argmin}_{P_X \in \mathcal{T}} \phi(P_X)$, we conclude that for all $\epsilon > 0$, there exist $\tau, \delta > 0$ vanishing with ϵ such that

$$R \leq \min_s \mathbb{I}(P_X^*, W_{Y|X,s}) + \tau + \epsilon + \delta \tag{37}$$

$$E \leq \phi(P_X^*) + \delta + \epsilon. \tag{38}$$

Since ϵ, τ, δ can be chosen arbitrarily small as the block length n goes to infinity, E is upper bounded by $\phi(P_X)$ for some $P_X \in \mathcal{P}_{\mathcal{X}}$ and the rate R is achieved by this P_X . Taking the union over all possible P_X completes the result of converse of Theorem 3.

C. Proof of Theorem 6

1) Definition of the Code

Before the protocol starts, we fix $n \in \mathbb{N}^*$, $\Delta_1, \Delta_2 \in (0, 1)$ such that $\Delta_1 + \Delta_2 < 1$, and we also fix some $|\mathcal{S}|$ -tuple $\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in \left(\mathcal{P}_X^{(1-\Delta_1-\Delta_2)n}\right)^{|\mathcal{S}|}$, where $\left(\mathcal{P}_X^{(1-\Delta_1-\Delta_2)n}\right)^{|\mathcal{S}|}$ is the set of tuples of types in $\mathcal{P}_X^{(1-\Delta_1-\Delta_2)n}$, as well as the number of messages M as

$$M = n(1 - \Delta_1 - \Delta_2) \times \left(\min_{s'' \in \mathcal{S}} \mathbb{I}(P_{X,s''}, W_{Y|X,s''}) - \delta \right) \quad (39)$$

for some $\delta > 0$.

We first define $g_{\text{ML},i}$ as the maximum likelihood estimator at each time $i \in \mathbb{N}^*$, i.e.,

$$g_{\text{ML},i}(x^i, z^i) \triangleq \underset{s \in \mathcal{S}}{\text{argmax}} \prod_{\ell=1}^i W_{Z|X,s}(z_\ell | x_\ell).$$

Then, the code $\mathcal{C}^{(m)} = (\{f_i\}_{i \in [1;n]}, g^{(m)}, h^{(m)})$ is defined through the following steps.

a) Initial Estimation

We define $P_X^\# = \underset{P_X \in \mathcal{P}_X^{\Delta_1 n}}{\text{argmax}} \phi(P_X)$ and pick any length $n\Delta_1$ sequence $\mathbf{v} = (v_1, \dots, v_{\Delta_1 n})$ from the type class $\mathcal{T}_{P_X^\#}$. Then, for $1 \leq i \leq \Delta_1 n$, the encoder is defined as

$$f_i(w, z^{i-1}) = v_i \quad (40)$$

for all $w \in [1; M]$ and $z^{i-1} \in \mathcal{Z}^{i-1}$. At time $\Delta_1 n + 1$, the transmitter estimates the state by using the maximum likelihood estimator $\tilde{s} = g_{\text{ML},\Delta_1 n}(x^{\Delta_1 n}, z^{\Delta_1 n})$.

b) State Information Transmission

The transmitter then conveys the information of the estimated state to the receiver by encoding the estimated state \tilde{s} into a codeword. Since $|\mathcal{S}|$ does not grow with n , there exists a length $\Delta_2 n$ channel code (\hat{f}, \hat{g}) with arbitrarily small error probability, where $\hat{f} : \mathcal{S} \mapsto \mathcal{X}^{\Delta_2 n}$ is the encoder and $\hat{g} : \mathcal{Y}^{\Delta_2 n} \mapsto \mathcal{S}$ is the decoder. Denoting $\hat{\mathbf{x}}(\tilde{s}) = (\hat{x}_1(\tilde{s}), \dots, \hat{x}_{\Delta_2 n}(\tilde{s})) = \hat{f}(\tilde{s})$ the codeword corresponding to \tilde{s} . Then, for $\Delta_1 n < i \leq (\Delta_1 + \Delta_2)n$, the encoder is defined as

$$f_i(w, z^{i-1}) = \hat{x}_{i-\Delta_1 n}(g_{\text{ML},\Delta_1 n}(x^{\Delta_1 n}, z^{\Delta_1 n})) \quad (41)$$

for all $w \in [1; M]$ and $z^{i-1} \in \mathcal{Z}^{i-1}$.

c) Message Transmission

It is known that for every $\bar{P}_X \in \mathcal{P}_X$, there exists a channel code with arbitrarily small error probability $\epsilon > 0$ such that the rate is at least $\mathbb{I}(\bar{P}_X, W_{Y|X,s}) - 2\tau$ for any $\tau > 0$ when the channel state is $s \in \mathcal{S}$ and the number of channel uses is large enough. Therefore, for any $s \in \mathcal{S}$, there

exist an $((1 - \Delta_1 - \Delta_2)n, \epsilon)$ constant composition code with the type $P_{X,s}$ for the state s channel with rate $\min_{s'' \in \mathcal{S}} \mathbb{I}(P_{X,s''}, W_{Y|X,s''}) - \delta$ for any $\epsilon > 0$, where the set of types $\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in \left(\mathcal{P}_X^{(1-\Delta_1-\Delta_2)n}\right)^{|\mathcal{S}|}$ is fixed at the beginning of the protocol. Let the channel code for the channel with state s be characterized by $(\tilde{f}_s, \tilde{h}_s)$, where $\tilde{f}_s : [1; M] \mapsto \mathcal{X}^{(1-\Delta_1-\Delta_2)n}$ is the encoder and $\tilde{h}_s : \mathcal{Y}^{(1-\Delta_1-\Delta_2)n} \mapsto [1 : M]$ is the decoder. Denoting $\tilde{\mathbf{x}}(w, s) = (\tilde{x}_1(w, s), \dots, \tilde{x}_{(1-\Delta_1-\Delta_2)n}(w, s)) = \tilde{f}_s(w)$ the codeword corresponding to the message $w \in [1; M]$. Then, for $(\Delta_1 + \Delta_2)n < i \leq n$, we define the encoder as

$$f_i(w, z^{i-1}) = \tilde{x}_{i-(\Delta_1+\Delta_2)n}(w, g_{\text{ML}, \Delta_1 n}(x^{\Delta_1 n}, z^{\Delta_1 n})) \quad (42)$$

for all $w \in [1; M]$ and $z^{i-1} \in \mathcal{Z}^{i-1}$.

d) Message Decoding and State Estimation

Finally, the message decoder is

$$h^{(m)}(y^n) = \tilde{h}_{\hat{g}(y_{\Delta_1 n+1}^{\Delta_2 n})}(y_{(\Delta_1+\Delta_2)n+1}^n) \quad (43)$$

for all $y^n \in \mathcal{Y}^n$, and the state estimator is

$$g^{(m)}(x^n, z^n) = \operatorname{argmax}_{s \in \mathcal{S}} \prod_{\ell=(\Delta_1+\Delta_2)n+1}^n W_{Z|X,s}(z_\ell | x_\ell) \quad (44)$$

for all $x^n \in \mathcal{X}^n$ and $z^n \in \mathcal{Z}^n$. Note that we only use the subsequences $x_{(\Delta_1+\Delta_2)n+1}^n$ and $z_{(\Delta_1+\Delta_2)n+1}^n$ for the state estimation in (44).

2) Analysis of the Communication Rate and Detection-Error Exponent

a) Rate Analysis

For any $s \in \mathcal{S}$ any $w \in [1; M]$, the error probability of communication is

$$\begin{aligned} & \mathbb{P}(h(Y^n) \neq W | W = w, S = s) \\ & \leq \mathbb{P}(g_{\text{ML}, \Delta_1 n}(X^{\Delta_1 n}, Z^{\Delta_1 n}) \neq s | W = s, S = s) + \mathbb{P}(\hat{h}(Y_{\Delta_1 n+1}^{\Delta_2 n}) \neq s, \tilde{s} = s | W = s, S = s) \\ & \quad + \mathbb{P}\left(\tilde{h}_s(Y_{(\Delta_1+\Delta_2)n+1}^n) \neq w, \tilde{s} = s, \hat{h}(Y_{\Delta_1 n+1}^{\Delta_2 n}) = s \mid W = w, S = s\right), \end{aligned} \quad (45)$$

where we have applied the union bound. The first term of (45) comes from the event in which the initial estimation of the state \tilde{s} is incorrect; the second term of (45) is the event in which the decoding of the initial estimated state \tilde{s} is incorrect; the last term of (45) is the decoding error probability. For any $s \in \mathcal{S}$ and $w \in [1; M]$, all terms on the right-hand side of (45) are arbitrarily small when n is sufficiently large by our construction of the code. The rate of communication is

$$R = \frac{1}{n} \log e^{n(1-\Delta_1-\Delta_2)(\min_{s'' \in \mathcal{S}} \mathbb{I}(P_{X,s''}, W_{Y|X,s''}) - \delta)} = (1 - \Delta_1 - \Delta_2) \left(\min_{s'' \in \mathcal{S}} \mathbb{I}(P_{X,s''}, W_{Y|X,s''}) - \delta \right).$$

By making Δ_1, Δ_2 and δ arbitrarily small, we conclude that $\min_{s'' \in \mathcal{S}} \mathbb{I}(P_{X,s''}, W_{Y|X,s''})$ is achievable.

b) Detection-Error Exponent Analysis

The error probability of detection is

$$\begin{aligned} P_d^{(n)} &= \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P}(g^{(m)}(X^n, Z^n) \neq s | W = w, S = s) \\ &= \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \mathbb{P} \left(\operatorname{argmax}_{s \in \mathcal{S}} \prod_{\ell=(\Delta_1+\Delta_2)n+1}^n W_{Z|X,s}(z_\ell | x_\ell) \neq s \middle| W = w, S = s \right). \end{aligned} \quad (46)$$

Note that $\operatorname{argmax}_{s \in \mathcal{S}} \prod_{\ell=(\Delta_1+\Delta_2)n+1}^n W_{Z|X,s}(z_\ell | x_\ell) \neq s$ is the error event of applying the ML estimator when the type of the input sequence is $\hat{p}_{X_{(1-\Delta_1-\Delta_2)n+1}^n} = P_{X,\tilde{s}}$. For all $t \in [n(\Delta_1 + \Delta_2) + 1; n]$, the channel input X_t is chosen without using the feedback $Z_{n(\Delta_1+\Delta_2)+1}^n$. Moreover, for any $w \in [1; M]$, the sequence $X_{n(\Delta_1+\Delta_2)+1}^n$ has a constant type, which depends on \tilde{s} . Therefore, we can apply Lemma 10 and the law of total probability to upper bound the error probability of (46) as

$$P_d^{(n)} \leq \max_{s \in \mathcal{S}} \max_{w \in [1;M]} \sum_{s'' \in \mathcal{S}} \mathbb{P}(\tilde{s} = s'' | W = w, S = s) \times \Theta_n(1) \times e^{-n(1-\Delta_1-\Delta_2)\psi_s(P_{X,s''})} \quad (47)$$

$$\leq \max_{s \in \mathcal{S}} \Theta_n(1) \times e^{-n(1-\Delta_1-\Delta_2) \min_{s'' \in \mathcal{S}} \psi_s(P_{X,s''})} \quad (48)$$

$$= \Theta_n(1) \times e^{-n(1-\Delta_1-\Delta_2) \min_{s \in \mathcal{S}} \min_{s'' \in \mathcal{S}} \psi_s(P_{X,s''})} \quad (49)$$

$$= \Theta_n(1) \times e^{-n(1-\Delta_1-\Delta_2) \min_{s'' \in \mathcal{S}} \phi(P_{X,s''})}, \quad (50)$$

where in (48) we lower bound the exponent $\psi_s(P_{X,s''})$ by the minimum one and $\Theta_n(1)$ is some constant, and in (50) we swap the order of $\min_{s \in \mathcal{S}}$ and $\min_{s'' \in \mathcal{S}}$ and apply the definition of ϕ . Then, we have shown that

$$E_d^{(n)} \geq (1 - \Delta_1 - \Delta_2) \min_{s \in \mathcal{S}} \phi(P_{X,s}). \quad (51)$$

By taking the union over all possible choices of $\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in \left(\mathcal{P}_X^{(1-\Delta_1-\Delta_2)n}\right)^{|\mathcal{S}|}$ and making Δ_1, Δ_2 arbitrarily small, we conclude that the following region is achievable.

$$\bigcup_{\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in \left(\mathcal{P}_X^{(1-\Delta_1-\Delta_2)n}\right)^{|\mathcal{S}|}} \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \min_{s \in \mathcal{S}} \mathbb{I}(P_{X,s}, W_{Y|X,s}) \\ E \leq \min_{s \in \mathcal{S}} \phi(P_{X,s}) \end{array} \right\} \quad (52)$$

To recover the result of Theorem 6, note that for any $P_X \in \mathcal{P}_{\mathcal{X}}$, there exists some type $\widehat{P}_X \in \mathcal{P}_{\mathcal{X}}^{(1-\Delta_1-\Delta_2)n}$ such that $|P_X(x) - \widehat{P}_X(x)| \leq \eta$ for any $x \in \mathcal{X}$ and $\eta > 0$ for some n large enough. A similar continuity argument can be used to guarantee that the rate and exponent derived by \widehat{P}_X would be $\xi(\eta)$ -close to the result obtained by P_X , where the difference $\xi(\eta)$ vanishes with n . By taking the union of all $\{P_{X,s}\}_{s \in \mathcal{S}}$, we conclude that the closure of all achievable regions is at least

$$\bigcup_{\{P_{X,s''}\}_{s'' \in \mathcal{S}} \in \mathcal{P}_{\mathcal{X}}^{|\mathcal{S}|}} \left\{ \begin{array}{l} (R, E) \in \mathbb{R}_+^2 : \\ R \leq \min_{s \in \mathcal{S}} \mathbb{I}(P_{X,s}, W_{Y|X,s}) \\ E \leq \min_{s \in \mathcal{S}} \phi(P_{X,s}) \end{array} \right\}, \quad (53)$$

which completes the proof.

D. Proof of Theorem 8

We first construct the code by the following steps. Fix a specific type P_X . The length n codeword corresponds to the message $w \in [1; M]$ is \mathbf{x}_w and is uniformly drawn from the type class $\mathcal{T}_{P_X}^n$. The message decoder and the state estimator are jointly defined as

$$(\hat{w}, \hat{s}) = \operatorname{argmax}_{w \in [1; M], s \in \mathcal{S}} \mathbb{P}(\mathbf{y}|S = s, X^n = \mathbf{x}_w). \quad (54)$$

The message decoder $h^{(b)}$ and the state estimator $g^{(b)}$ are then well-defined by (54). Note that the codewords $\{\mathbf{x}_w\}_{w \in [1; M]}$ are random and so is the code $\mathcal{C}^{(b)}$. Since we have fixed the definition of $h^{(b)}$ and $g^{(b)}$, with a slight abuse of notation, we denote $\mathcal{C}^{(b)} = \{\mathbf{x}_\ell\}_{\ell \in [1; M]}$ as the set of codewords in the derivation below. We next derive the detection-error exponent when averaging over $\mathcal{C}^{(b)}$, which codewords are drawn uniformly from $\mathcal{T}_{P_X}^n$.

a) Detection-error Analysis

Without loss of generality we assume that the message $w = 1$ is transmitted and the state $S = s$ for some $s \in \mathcal{S}$. The error event is the set of all received \mathbf{y} that would result in detection errors and is defined as

$$\mathcal{E} \triangleq \{\mathbf{y} \in \mathcal{Y}^n : \max_{i \in [1; M]} \mathbb{P}(\mathbf{y}|S = s, X^n = \mathbf{x}_i) < \max_{j \in [1; M]} \mathbb{P}(\mathbf{y}|S = s', X^n = \mathbf{x}_j) \text{ for some } s' \neq s\}.$$

For any $\widehat{P}_{Y|X} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n$, it is known from [28] that the probability of receiving $\mathbf{y} \in \mathcal{T}_{\cdot|\mathbf{x}_1}(\widehat{P}_{Y|X})$ is upper bounded by

$$\mathbb{P}(\mathbf{y} \in \mathcal{T}_{\cdot|\mathbf{x}_1}(\widehat{P}_{Y|X}) \mid X^n = \mathbf{x}_1, S = s) \leq e^{-n\mathbb{D}(\widehat{P}_{Y|X} \parallel W_{Y|X,s}|P_X)}. \quad (55)$$

For all $\mathbf{y} \in \mathcal{Y}^n$ and for any $k \neq 1$, we define the random variable

$$\begin{aligned} I_k(\mathbf{y}) &\triangleq \mathbf{1} \left(\max_{\ell \neq k} \mathbb{P}(\mathbf{y}|S=s, X^n = \mathbf{x}_\ell) \leq \mathbb{P}(\mathbf{y}|S=s', X^n = \mathbf{x}_k) \text{ for some } s' \neq s \right) \\ &= \mathbf{1} \left(\exists s' \neq s \text{ s.t. } e^{n \sum_{x,y} P_X(x) \hat{p}_{\mathbf{y}|\mathbf{x}_k}(y|x) \log W_{Y|X,s'}(y|x)} \geq \max_{\ell \neq k} e^{n \sum_{x,y} P_X(x) \hat{p}_{\mathbf{y}|\mathbf{x}_\ell}(y|x) \log W_{Y|X,s}(y|x)} \right), \end{aligned}$$

as well as the variable

$$J_1(\mathbf{y}) \triangleq \mathbf{1} \left(\exists s' \neq s \text{ s.t. } e^{n \sum_{x,y} P_X(x) \hat{p}_{\mathbf{y}|\mathbf{x}_1}(y|x) \log W_{Y|X,s'}(y|x)} \geq e^{n \sum_{x,y} P_X(x) \hat{p}_{\mathbf{y}|\mathbf{x}_1}(y|x) \log W_{Y|X,s}(y|x)} \right). \quad (56)$$

The randomness of the variables $I_k(\mathbf{y})$ and $J_1(\mathbf{y})$ comes from the random coding, i.e., the type $\hat{p}_{\mathbf{y}|\mathbf{x}_k}$ is random. $I_k(\mathbf{y})$ is equal to one when there exists some state $s' \neq s$ such that the likelihood calculated according to the state s' and the codeword k is greater than the maximal possible likelihood according to the state s . Then, the average detection error probability can be upper bounded by the following

$$\begin{aligned} &\mathbb{E}_{\mathcal{C}^{(b)}} [\mathbb{P}(\hat{s} \neq s | X^n = \mathbf{x}_1, S = s)] \\ &\leq \frac{1}{|\mathcal{T}_{P_X}^n|} \sum_{\mathbf{x}_1 \in \mathcal{T}_{P_X}^n} \sum_{\hat{P}_{Y|X}^{(1)}} \sum_{\mathbf{y} \in \mathcal{T}_{|\mathbf{x}_1|}(\hat{P}_{Y|X}^{(1)})} \mathbb{P}(\mathbf{y}|X^n = \mathbf{x}_1, S = s) \mathbb{P}(\text{error}|X^n = \mathbf{x}_1, S = s, Y^n = \mathbf{y}) \quad (57) \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{|\mathcal{T}_{P_X}^n|} \sum_{\mathbf{x}_1 \in \mathcal{T}_{P_X}^n} \sum_{\hat{P}_{Y|X}^{(1)}} \sum_{\mathbf{y} \in \mathcal{T}_{|\mathbf{x}_1|}(\hat{P}_{Y|X}^{(1)})} \mathbb{P}(\mathbf{y}|X^n = \mathbf{x}_1, S = s) \sum_{k \neq 1} \mathbb{E}_{\mathcal{C}^{(b)} \setminus \mathbf{x}_1} [I_k(\mathbf{y}) | X^n = \mathbf{x}_1, S = s] \\ &+ \frac{1}{|\mathcal{T}_{P_X}^n|} \sum_{\mathbf{x}_1 \in \mathcal{T}_{P_X}^n} \sum_{\hat{P}_{Y|X}^{(1)}} \sum_{\mathbf{y} \in \mathcal{T}_{|\mathbf{x}_1|}(\hat{P}_{Y|X}^{(1)})} \mathbb{P}(\mathbf{y}|X^n = \mathbf{x}_1, S = s) J_1(\mathbf{y}), \quad (58) \end{aligned}$$

where the expectation is taken over all possibilities of codewords $\{\mathbf{x}_\ell\}_{\ell \neq 1}$. Note that $I_k(\mathbf{y})$ is random because the codewords $\{\mathbf{x}_\ell\}_{\ell \neq 1}$ are drawn from $\mathcal{T}_{P_X}^n$ uniformly, and hence, $\{\hat{p}_{\mathbf{y}|\mathbf{x}_\ell}\}_{\ell \neq 1}$ are random, as well.

The difficulty of analyzing $I_k(\mathbf{y})$ comes from the term $\max_{\ell \neq k} e^{n \sum_{x,y} P_X(x) \hat{p}_{\mathbf{y}|\mathbf{x}_\ell}(y|x) \log W_{Y|X,s}(y|x)}$, which involves a maximization over exponentially many indices ℓ . Observe that

$$\max_{\ell \neq k} e^{n \sum_{x,y} P_X(x) \hat{p}_{\mathbf{y}|\mathbf{x}_\ell}(y|x) \log W_{Y|X,s}(y|x)} = \max_{\ell \neq k} e^{-n(\mathbb{D}(\hat{p}_{\mathbf{y}|\mathbf{x}_\ell} \| W_{Y|X,s} | P_X) + \mathbb{H}(\hat{p}_{\mathbf{y}|\mathbf{x}_\ell} | P_X))}. \quad (59)$$

Note that the type $\hat{p}_{\mathbf{y}|\mathbf{x}_1}$ is fixed when $\mathbf{y} \in \mathcal{T}_{|\mathbf{x}_1|}(\hat{P}_{Y|X}^{(1)})$ for some $\hat{P}_{Y|X}^{(1)} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n$. Therefore, in the following we define a set \mathcal{A}_k of conditional types $\left\{ \{\hat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1, k\}} \right\} \in \left(\mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n \right)^{2^{nR}-2}$ such that when $\mathbf{x}_\ell \in \mathcal{T}_{\mathbf{y}|\cdot}(\hat{P}_{Y|X}^{(\ell)})$ for all $\ell \neq \{1, k\}$, we can control the right-hand side of (59). Moreover,

\mathcal{A}_k needs to have the property that $\{\{\widehat{P}_{Y|X_\ell}\}_{\ell \notin \{1,k\}}\} \in \mathcal{A}_k$ with high probability. Specifically, we define

$$\begin{aligned} \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R) &\triangleq \left\{ \{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}} \in (\mathcal{P}_{Y|X}^n)^{2^{nR}-2} : \forall P'' \in \mathcal{P}_{Y|X}^n \text{ s.t. } \mathbb{I}(P_X, P'') < R \right. \\ &\quad \left. \text{and } P_X \circ P'' = P_X \circ \widehat{P}_{Y|X}^{(1)} \exists j \notin \{1, k\} \text{ s.t. } \widehat{P}_{Y|X}^{(j)} = P'', \widehat{P}_{Y|X}^{(j)} \circ P_X = \widehat{P}_{Y|X}^{(1)} \circ P_X \forall j \notin \{1, k\} \right\}, \end{aligned}$$

as the set of conditional types such that for all $P'' \in \mathcal{P}_{Y|X}^n$ satisfying $\mathbb{I}(P_X, P'') < R$ there exists some index $j \notin \{1, k\}$ such that the corresponding conditional type $\widehat{P}_{Y|X}^{(j)} = P''$. The constraint $\widehat{P}_{Y|X}^{(j)} \circ P_X = \widehat{P}_{Y|X}^{(1)} \circ P_X$ for all $j \notin \{1, k\}$ comes from the fact that when $\mathbf{y} \in \mathcal{T}_{|\mathbf{x}_1}(\widehat{P}_{Y|X}^{(1)})$, the type of \mathbf{y} is $\widehat{P}_{Y|X}^{(1)} \circ P_X$, and hence $\mathbb{P}(\mathbf{x}_\ell \in \mathcal{T}_{\mathbf{y} \cdot}(\widehat{P}_{Y|X}^{(\ell)})) = 0$ if $\widehat{P}_{Y|X}^{(\ell)} \circ P_X \neq \widehat{P}_{Y|X}^{(1)} \circ P_X$.

Then, by using the law of total probability, we have

$$\begin{aligned} &\mathbb{E}_{\mathcal{C}^{(b)} \setminus \mathbf{x}_1} [I_k(\mathbf{y}) | X^n = \mathbf{x}_1, S = s] \\ &= \sum_{\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}}} \prod_{\ell \notin \{1,k\}} \mathbb{P}(\mathbf{x}_\ell \in \mathcal{T}_{\mathbf{y} \cdot}(\widehat{P}_{Y|X}^{(\ell)})) \mathbb{E}_{\mathbf{x}_k} \left[I_k(\mathbf{y}) \middle| X^n = \mathbf{x}_1, S = s, \widehat{p}_{\mathbf{y} \mid \mathbf{x}_\ell} = \widehat{P}_{Y|X}^{(\ell)} \forall \ell \notin \{1, k\} \right] \\ &\leq \sum_{\substack{\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}} \\ \in \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R)}} \prod_{\ell \notin \{1,k\}} \mathbb{P}(\mathbf{x}_\ell \in \mathcal{T}_{\mathbf{y} \cdot}(\widehat{P}_{Y|X}^{(\ell)})) \mathbb{E}_{\mathbf{x}_k} \left[I_k(\mathbf{y}) \middle| X^n = \mathbf{x}_1, S = s, \widehat{p}_{\mathbf{y} \mid \mathbf{x}_\ell} = \widehat{P}_{Y|X}^{(\ell)} \forall \ell \notin \{1, k\} \right] \\ &\quad + \sum_{\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}} \notin \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R)} \prod_{\ell \notin \{1,k\}} \mathbb{P}(\mathbf{x}_\ell \in \mathcal{T}_{\mathbf{y} \cdot}(\widehat{P}_{Y|X}^{(\ell)})). \end{aligned}$$

Note that when $\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}} \in \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R)$, it holds that

$$\begin{aligned} &\min_{\ell \notin \{k, 1\}} \mathbb{D}(\widehat{P}_{Y|X}^{(\ell)} \| W_{Y|X,s} | P_X) + \mathbb{H}(\widehat{P}_{Y|X}^{(\ell)} | P_X) \\ &\leq \min_{P'' \in \mathcal{P}_{Y|X}^n : \mathbb{I}(P_X, P'') < R, P_X \circ P'' = P_X \circ \widehat{P}_{Y|X}^{(1)}} \mathbb{D}(P'' \| W_{Y|X,s} | P_X) + \mathbb{H}(P'' | P_X) \end{aligned} \quad (60)$$

$$\triangleq \beta_n(\widehat{P}_{Y|X}^{(1)}, P_X, R, s), \quad (61)$$

where (60) follows since for every $P'' \in \mathcal{P}_{Y|X}^n$, there is some $\widehat{P}_{Y|X}^{(\ell)}$ such that $\widehat{P}_{Y|X}^{(\ell)} = P''$ by definition of \mathcal{A}_k . Then, for $\mathbf{y} \in \mathcal{T}_{|\mathbf{x}_1}(\widehat{P}_{Y|X}^{(1)})$, it holds that

$$\begin{aligned} &\max_{\ell \neq k} e^{n \sum_{x,y} P_X(x) \widehat{p}_{\mathbf{y} \mid \mathbf{x}_\ell}(y|x) \log W_{Y|X,s}(y|x)} \\ &\geq \max \left(e^{-n\beta_n(\widehat{P}_{Y|X}^{(1)}, P_X, R, s)}, e^{-n\mathbb{D}(\widehat{P}_{Y|X}^{(1)} \| W_{Y|X,s} | P_X) + \mathbb{H}(\widehat{P}_{Y|X}^{(1)} | P_X)} \right) \end{aligned}$$

by applying the right-hand side of (59) and (61). Therefore, when $\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}} \in \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R)$ and $\mathbf{y} \in \mathcal{T}_{|\mathbf{x}_1}(\widehat{P}_{Y|X}^{(1)})$, the indicator function I_k can be upper bounded by

$$I_k(\mathbf{y}) \leq \sum_{s' \neq s} \mathbf{1} \left(\mathbf{x}_k \in \mathcal{T}_{\mathbf{y}|.}(P'_{Y|X}) \text{ for some } P'_{Y|X} \in \mathcal{P}_{s,s'}^n(\widehat{P}_{Y|X}^{(1)}, P_X, R) \right), \quad (62)$$

where

$$\begin{aligned} \mathcal{P}_{s,s'}^n(\widehat{P}_{Y|X}^{(1)}, P_X, R) &\triangleq \left\{ P'_{Y|X} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n : e^{n \sum_{x,y} P_X(x) P'_{Y|X}(y|x) \log W_{Y|X,s'}(y|x)} \right. \\ &\geq e^{-n \min(\beta_n(\widehat{P}_{Y|X}^{(1)}, P_X, R, s), \mathbb{D}(\widehat{P}_{Y|X}^{(1)} \| W_{Y|X,s} P_X) + \mathbb{H}(\widehat{P}_{Y|X}^{(1)} | P_X))}, P'_{Y|X} \circ P_X = \widehat{P}_{Y|X}^{(1)} \circ P_X \left. \right\}. \end{aligned}$$

Then,

$$\begin{aligned} &\sum_{\substack{\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}} \\ \in \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R)}} \prod_{\ell \notin \{1,k\}} \mathbb{P}(\mathbf{x}_\ell \in \mathcal{T}_{\mathbf{y}|.}(\widehat{P}_{Y|X}^{(\ell)})) \mathbb{E}_{\mathbf{x}_k} \left[I_k(\mathbf{y}) \middle| X^n = \mathbf{x}_1, S = s, \widehat{p}_{\mathbf{y}|\mathbf{x}_\ell} = \widehat{P}_{Y|X}^{(\ell)} \forall \ell \notin \{1,k\} \right] \\ &\leq \sum_{s' \neq s} \sum_{P'_{Y|X} \in \mathcal{P}_{s,s'}^n(\widehat{P}_{Y|X}^{(1)}, P_X, R)} \mathbb{P}(\mathbf{x}_k \in \mathcal{T}_{\mathbf{y}|.}(P'_{Y|X})) \end{aligned} \quad (63)$$

$$\leq \exp \left(-n \min_{s' \neq s} \min_{P'_{Y|X} \in \mathcal{P}_{s,s'}^n(\widehat{P}_{Y|X}^{(1)}, P_X, R)} (\mathbb{I}(P_X, P'_{Y|X}) - o_n(1)) \right), \quad (64)$$

where we have used the fact [29, (41)] that

$$\mathbb{P}(\mathbf{x}_k \in \mathcal{T}_{\mathbf{y}|.}(P'_{Y|X})) = \frac{|\mathcal{T}_{P_X P'_{Y|X}}^n|}{|\mathcal{T}_{P_X}^n| |\mathcal{T}_{P_X \circ P'_{Y|X}}^n|} \leq \exp(-n \mathbb{I}(P_X, P'_{Y|X})),$$

and $|\mathcal{P}_{s,s'}^n(\widehat{P}_{Y|X}^{(1)}, P_X, R)| \leq \text{poly}(n)$. Moreover, for all $P'' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n$ such that $\mathbb{I}(P_X, P'') < R$ and for all $\ell \notin \{1,k\}$, it holds that

$$\mathbb{P}(\mathbf{x}_\ell \notin \mathcal{T}_{\mathbf{y}|.}(P'')) \leq 1 - e^{-n(\mathbb{I}(P_X, P'') + o_n(1))} \quad (65)$$

$$\leq e^{-e^{-n(\xi + o_n(1))}}, \quad (66)$$

for some $\xi < R$, where in (65) we lower bound $\mathbb{P}(\mathbf{x}_k \in \mathcal{T}_{\mathbf{y}|.}(P''_{Y|X}))$ by $e^{-n(\mathbb{I}(P_X, P'') + o_n(1))}$ [29, (14)] and in (66) we use the fact that $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$. Then,

$$\begin{aligned} \sum_{\substack{\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1,k\}} \\ \notin \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R)}} \prod_{\ell \notin \{1,k\}} \mathbb{P}(\mathbf{x}_\ell \in \mathcal{T}_{\mathbf{y}|.}(\widehat{P}_{Y|X}^{(\ell)})) &\leq \sum_{\substack{P'' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n : \mathbb{I}(P_X, P'') < R, \\ P_X \circ P'' = P_X \circ \widehat{P}_{Y|X}^{(1)}}} \prod_{\ell \notin \{1,k\}} \mathbb{P}(\mathbf{x}_\ell \notin \mathcal{T}_{\mathbf{y}|.}(P'')) \end{aligned} \quad (67)$$

$$\leq |\mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n| \left(e^{-e^{-n(\xi + o_n(1))}} \right)^{e^{nR}} \quad (68)$$

$$\leq e^{n o_n(1)} \times e^{-e^{-n(\xi - R - o_n(1))}}, \quad (69)$$

which has the double exponential form and decays much faster than the right-hand side of (64), where (67) follows since $\{\widehat{P}_{Y|X}^{(\ell)}\}_{\ell \notin \{1, k\}} \notin \mathcal{A}_k(\widehat{P}_{Y|X}^{(1)}, P_X, R)$ implies that there exists some $P'' \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n$ such that $\mathbb{I}(P_X, P'') < R$, $P_X \circ P'' = P_X \circ \widehat{P}_{Y|X}^{(1)}$ but no \mathbf{x}_ℓ lies in $\mathcal{T}_{\mathbf{y}|\cdot}(P'')$. Therefore, for $\mathbf{y} \in \mathcal{T}_{\cdot|\mathbf{x}_1}(\widehat{P}_{Y|X}^{(1)})$, it holds that

$$\mathbb{E}_{\mathcal{C}^{(b)} \setminus \mathbf{x}_1} [I_k(\mathbf{y}) | X^n = \mathbf{x}_1, S = s] \leq \exp \left(-n \min_{s' \neq s} \min_{P'_{Y|X} \in \mathcal{P}_{s,s'}^n(\widehat{P}_{Y|X}^{(1)}, P_X, R)} (\mathbb{I}(P_X, P'_{Y|X}) - o_n(1)) \right). \quad (70)$$

By applying inequality (70) to the first term on the right-hand side of (58) and using (55), we obtain

$$\begin{aligned} & \frac{1}{|\mathcal{T}_{P_X}^n|} \sum_{\mathbf{x}_1 \in \mathcal{T}_{P_X}^n} \sum_{\widehat{P}_{Y|X}^{(1)}} \sum_{\mathbf{y} \in \mathcal{T}_{\cdot|\mathbf{x}_1}(\widehat{P}_{Y|X}^{(1)})} \mathbb{P}(\mathbf{y} | X^n = \mathbf{x}_1, S = s) \sum_{k \neq 1} \mathbb{E}_{\mathcal{C}^{(b)} \setminus \mathbf{x}_1} [I_k(\mathbf{y}) | X^n = \mathbf{x}_1, S = s] \\ & \leq \text{poly}(n) \\ & \times \exp \left(-n \min_{\widehat{P}_{Y|X}^{(1)}} \left(\mathbb{D}(\widehat{P}_{Y|X}^{(1)} \| W_{Y|X,s} | P_X) + \min_{s' \neq s} \min_{P'_{Y|X} \in \mathcal{P}_{s,s'}^n(\widehat{P}_{Y|X}^{(1)}, P_X, R)} (\mathbb{I}(P_X, P'_{Y|X}) - o_n(1)) - R \right) \right). \end{aligned} \quad (71)$$

Besides, the term

$$\frac{1}{|\mathcal{T}_{P_X}^n|} \sum_{\mathbf{x}_1 \in \mathcal{T}_{P_X}^n} \sum_{\widehat{P}_{Y|X}^{(1)}} \sum_{\mathbf{y} \in \mathcal{T}_{\cdot|\mathbf{x}_1}(\widehat{P}_{Y|X}^{(1)})} \mathbb{P}(\mathbf{y} | X^n = \mathbf{x}_1, S = s) J_1(\mathbf{y})$$

is the detection-error probability by using the ML detector when the codeword is known. From Lemma 10, we have

$$\frac{1}{|\mathcal{T}_{P_X}^n|} \sum_{\mathbf{x}_1 \in \mathcal{T}_{P_X}^n} \sum_{\widehat{P}_{Y|X}^{(1)}} \sum_{\mathbf{y} \in \mathcal{T}_{\cdot|\mathbf{x}_1}(\widehat{P}_{Y|X}^{(1)})} \mathbb{P}(\mathbf{y} | X^n = \mathbf{x}_1, S = s) J_1(\mathbf{y}) \leq e^{-n\phi(P_X)} \quad (72)$$

Combining (58), (71), (72), we have

$$\mathbb{E}_{\mathcal{C}^{(b)}} [\mathbb{P}(\hat{s} \neq s | X^n = \mathbf{x}_1, S = s)] \leq e^{-n \min(\phi(P_X), \rho_{\text{joint}}(P_X, R) - \epsilon(n))}, \quad (73)$$

where $\epsilon(n) \rightarrow 0$ when $n \rightarrow \infty$. The definition of $\rho_{\text{joint}}(P_X, R)$ involves the optimization over conditional distributions, i.e., $\min_{\widehat{P} \in \mathcal{P}_{Y|X}}$ and $\min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)}$, but the result we obtained in (71) involves optimization over conditional types of length n sequences. However, when n is sufficiently large, any conditional distribution in $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}$ can be approximated by conditional types in $\mathcal{P}_{\mathcal{Y}|\mathcal{X}}^n$ with some deviation $\delta(n)$ with $\lim_{n \rightarrow \infty} \delta(n) = 0$. Since all the functions in the definition

of $\rho_{\text{joint}}(P_X, R)$ is continuous in the conditional distribution, the difference between the right-hand side of (71) and $\rho_{\text{joint}}(R_X, R)$ can be bounded by some $\epsilon(\delta(n))$ and again vanishes with n . The right-hand side of (73) is irrelevant to the message and the state, and hence,

$$\mathbb{E}_{\mathcal{C}^{(b)}} \left[\frac{1}{M} \sum_w \mathbb{P}(\hat{s} \neq s | X^n = \mathbf{x}_w, S = s) \right] \leq e^{-n \min(\phi(P_X), \rho_{\text{joint}}(P_X, R) - \epsilon(\delta))} \quad (74)$$

for all $s \in \mathcal{S}$ by the linearity of expectation.

b) Communication-error Analysis

By [27, Lemma 10.1 and Theorem 10.2], we have that,

$$\mathbb{E}_{\mathcal{C}^{(b)}} \left[\max_s \max_w \mathbb{P}(\hat{w} \neq w | X^n = \mathbf{x}_w, S = s) \right] \leq e^{-n(\rho_{\text{succ}}(P_X, R) - \kappa)},$$

where κ vanishes with n .

c) Derandomization and Expurgation

By the Markov's inequality, we have for any $\zeta_1 > 0$

$$\begin{aligned} & \mathbb{P}_{\mathcal{C}^{(b)}} \left(\exists s \in \mathcal{S} \text{ s.t. } \frac{1}{M} \sum_w \mathbb{P}(\hat{s} \neq s | X^n = \mathbf{x}_w, S = s) > e^{-n(\min(\phi(P_X), \rho_{\text{joint}}(P_X, R) - \epsilon(\delta)) - \zeta_1)} \right. \\ & \quad \left. \text{or } \max_s \max_w \mathbb{P}(\hat{w} \neq w | X^n = \mathbf{x}_w, S = s) \geq e^{-n(\rho_{\text{succ}}(P_X, R) - \zeta_2)} \right) \\ & \leq |\mathcal{S}| \frac{\mathbb{E}_{\mathcal{C}^{(b)}} \left[\frac{1}{M} \sum_w \mathbb{P}(\hat{s} \neq s | X^n = \mathbf{x}_w, S = s) \right]}{e^{-n(\min(\phi(P_X), \rho_{\text{joint}}(P_X, R) - \epsilon(\delta)) - \zeta_1)}} + \frac{\mathbb{E}_{\mathcal{C}^{(b)}} \left[\max_s \max_w \mathbb{P}(\hat{w} \neq w | X^n = \mathbf{x}_w, S = s) \right]}{e^{-n(\rho_{\text{succ}}(P_X, R) - \zeta_2)}} \end{aligned} \quad (75)$$

$$\leq |\mathcal{S}| e^{-n\zeta_1} + e^{-n\zeta_2}, \quad (76)$$

which goes to zero when $n \rightarrow \infty$. Therefore, there exists some code $\mathcal{C}^{(b)}$ such that

$$\max_s \frac{1}{M} \sum_{w=1}^M \mathbb{P}(\hat{s} \neq s | X^n = \mathbf{x}_w, S = s) < e^{-n(\min(\phi(P_X), \rho_{\text{joint}}(P_X, R) - \epsilon(\delta)) - \zeta_1)}, \quad (77)$$

$$\max_s \max_w \mathbb{P}(\hat{w} \neq w | X^n = \mathbf{x}_w, S = s) < e^{-n(\rho_{\text{succ}}(P_X, R) - \zeta_2)} \quad (78)$$

when n is sufficiently large. By the codebook expurgation argument, there exists a code $\bar{\mathcal{C}}^{(b)} \triangleq \{\bar{\mathbf{x}}_w\} \subset \mathcal{C}^{(b)}$ with size $|\bar{\mathcal{C}}^{(b)}| = |\mathcal{C}^{(b)}|/2$ such that

$$\max_s \max_w \mathbb{P}(\hat{s} \neq s | X^n = \bar{\mathbf{x}}_w, S = s) < 2e^{-n(\min(\phi(P_X), \rho_{\text{joint}}(P_X, R) - \epsilon(\delta)) - \zeta_1)}, \quad (79)$$

$$\max_s \max_w \mathbb{P}(\hat{w} \neq w | X^n = \mathbf{x}_w, S = s) < e^{-n(\rho_{\text{succ}}(P_X, R) - \zeta_2)}. \quad (80)$$

For every $R < \min_s \mathbb{I}(P_X, W_{Y|X,s})$, there exists some ζ_2 small enough such that $\rho_{\text{succ}}(P_X, R) - \zeta_2 > 0$. Therefore, for every $R < \min_s \mathbb{I}(P_X, W_{Y|X,s})$ there exists a code $\bar{C}^{(b)}$ with rate $R - o_n(1)$ such that

$$-\frac{1}{n} \log \max_s \max_w \mathbb{P}(\hat{s} \neq s | X^n = \bar{\mathbf{x}}_w, S = s) \geq \min(\phi(P_X), \rho_{\text{joint}}(P_X, R)) - \epsilon, \quad (81)$$

$$\max_s \max_w \mathbb{P}(\hat{w} \neq w | X^n = \bar{\mathbf{x}}_w, S = s) < \epsilon \quad (82)$$

for any $\epsilon > 0$ whenever n is sufficiently large. The theorem is proved by taking the union over all possible P_X .

E. Proof of Corollary 9

Recall that the definition of $\rho_{\text{joint}}(P_X, R)$ is

$$\rho_{\text{joint}}(P_X, R) \triangleq \min_{s \in \mathcal{S}} \min_{\hat{P}} \left(\mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\hat{P}, P_X, R)} \mathbb{I}(P_X, P') - R \right|^+ \right). \quad (83)$$

For any \hat{P} , P_X , $s \in \mathcal{S}$ and $s' \neq s$, the definition of $\mathcal{P}_{s,s'}(\hat{P}, P_X, R)$ is

$$\begin{aligned} \mathcal{P}_{s,s'}(\hat{P}, P_X, R) &\triangleq \left\{ P' : \mathbb{D}(P' \| W_{Y|X,s'} | P_X) + \mathbb{H}(P' | P_X) \right. \\ &\leq \min \left(\beta(\hat{P}, P_X, R, s), \mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\hat{P} | P_X) \right), P_X \circ P' = P_X \circ \hat{P} \left. \right\}. \end{aligned}$$

The value of $\min_{s'' \in \mathcal{S}} \left(\beta(\hat{P}, P_X, R, s), \mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\hat{P} | P_X) \right)$ depends on \hat{P} , s and P_X . Therefore, for each $s \in \mathcal{S}$ and P_X , we partition the set $\mathcal{P}_{Y|X}$ into $\mathcal{P}'_{Y|X}(P_X, s)$ and $\mathcal{P}''_{Y|X}(P_X, s)$, where

$$\begin{aligned} \mathcal{P}'_{Y|X}(P_X, R, s) &\triangleq \{ \hat{P} \in \mathcal{P}_{Y|X} : \max_{s'' \in \mathcal{S}} \beta(\hat{P}, P_X, R, s'') \geq \mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\hat{P} | P_X) \} \\ \mathcal{P}''_{Y|X}(P_X, R, s) &\triangleq \{ \hat{P} \in \mathcal{P}_{Y|X} : \max_{s'' \in \mathcal{S}} \beta(\hat{P}, P_X, R, s'') < \mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(\hat{P} | P_X) \}. \end{aligned}$$

For each $s \in \mathcal{S}$, and P_X , we define

$$\hat{\rho}(P_X, R, s, \hat{P}) \triangleq \mathbb{D}(\hat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\hat{P}, P_X, R)} \mathbb{I}(P_X, P') - R \right|^+. \quad (84)$$

The exponent $\rho_{\text{joint}}(P_X, R)$ is then

$$\rho_{\text{joint}}(P_X, R) = \min_{s \in \mathcal{S}} \min_{\hat{P} \in \mathcal{P}'_{Y|X}(P_X, R, s)} \hat{\rho}(P_X, R, s, \hat{P}), \min_{\hat{P} \in \mathcal{P}''_{Y|X}(P_X, R, s)} \hat{\rho}(P_X, R, s, \hat{P}) \quad (85)$$

We first analyze the term $\min_s \min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \widehat{\rho}(P_X, R, s, \widehat{P})$. Let the arguments of the minimization of the term

$$\min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)} \mathbb{I}(P_X, P') - R \right|^+$$

be achieved by the tuple (\widehat{P}, s', P') , where \widehat{P} , s' and P' are the minimizer corresponding to $\min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)}$, $\min_{s' \neq s}$ and $\min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)}$, respectively. Given the tuple (\widehat{P}, s', P') , if $\mathbb{I}(P_X, P') \geq \mathbb{I}(P_X, \widehat{P})$, then

$$\begin{aligned} & \min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)} \mathbb{I}(P_X, P') - R \right|^+ \\ &= \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + |\mathbb{I}(P_X, P') - R|^+ \\ &\geq \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + |\mathbb{I}(P_X, \widehat{P}) - R|^+ \\ &\geq \min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + |\mathbb{I}(P_X, \widehat{P}) - R|^+. \end{aligned}$$

On the other hand, if the minimizer P' satisfies $\mathbb{I}(P_X, P') < \mathbb{I}(P_X, \widehat{P})$, then by definition of $\mathcal{P}_{s,s'}(\widehat{P}, P_X)$ and the fact that $\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)$ it holds that

$$\mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) \geq \mathbb{D}(P' \| W_{Y|X,s'} | P_X)$$

Moreover, since $P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)$, it holds that

$$\mathbb{D}(P' \| W_{Y|X,s'} | P_X) + \mathbb{H}(P' | P_X) \leq \beta(\widehat{P}, P_X, R, s) \quad (86)$$

$$= \beta(P', P_X, R, s) \quad (87)$$

$$\leq \max_{s'' \in \mathcal{S}} \beta(P', P_X, R, s''), \quad (88)$$

where in (87) we use the fact that $P_X \circ P' = P_X \circ \widehat{P}$. Therefore, by definition of $\mathcal{P}'_{\mathcal{Y}|\mathcal{X}}$ and (88), it holds that $P' \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s')$. Then,

$$\min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X)} \mathbb{I}(P_X, P') - R \right|^+ \quad (89)$$

$$= \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + |\mathbb{I}(P_X, P') - R|^+ \quad (90)$$

$$\geq \mathbb{D}(P' \| W_{Y|X,s'} | P_X) + |\mathbb{I}(P_X, P') - R|^+ \quad (91)$$

$$\geq \min_{P' \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s')} \mathbb{D}(P' \| W_{Y|X,s'} | P_X) + |\mathbb{I}(P_X, P') - R|^+. \quad (92)$$

Taking the minimum over both cases of $\mathbb{I}(P_X, P') \geq \mathbb{I}(P_X, \widehat{P})$ and $\mathbb{I}(P_X, P') < \mathbb{I}(P_X, \widehat{P})$ and all possible $s \in \mathcal{S}$, we obtain that

$$\min_s \min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \widehat{\rho}(P_X, R, s, \widehat{P}) \geq \min_s \min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \mathbb{I}(P_X, \widehat{P}) - R \right|^+ . \quad (93)$$

Our next step is to analyze $\min_s \min_{\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \widehat{\rho}(P_X, R, s, \widehat{P})$. We assume again that the minimization of the term

$$\min_{\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)} \mathbb{I}(P_X, P') - R \right|^+$$

is achieved by the tuple (\widehat{P}, s', P') .

Given the tuple (\widehat{P}, s', P') , if $\mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) \geq \mathbb{D}(P' \| W_{Y|X,s'} | P_X)$, then

$$\min_{\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)} \mathbb{I}(P_X, P') - R \right|^+ \quad (94)$$

$$= \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + |\mathbb{I}(P_X, P') - R|^+ \quad (95)$$

$$\geq \mathbb{D}(P' \| W_{Y|X,s'} | P_X) + |\mathbb{I}(P_X, P') - R|^+ \quad (96)$$

$$\geq \min_{P' \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s')} \mathbb{D}(P' \| W_{Y|X,s'} | P_X) + |\mathbb{I}(P_X, P') - R|^+ , \quad (97)$$

where (97) follows from the fact that $P' \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s')$ and has the same form as the right-hand side of (93). On the other hand, if $\mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) < \mathbb{D}(P' \| W_{Y|X,s'} | P_X)$, we have

$$\begin{aligned} & \min_{\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \min_{s' \neq s} \min_{P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X)} \mathbb{I}(P_X, P') - R \right|^+ \\ &= \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + |\mathbb{I}(P_X, P') - R|^+ \\ &\geq \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + |\mathbb{D}(P' \| W_{Y|X,s'} | P_X) + \mathbb{H}(P_X \circ P') - \mathbb{D}(P' \| W_{Y|X,s'} | P_X) - \mathbb{H}(P' | P_X) - R|^+ \\ &\stackrel{(a)}{\geq} \min_{\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(P_X \circ \widehat{P}) - \beta(\widehat{P}, P_X, R, s) - R \right|^+ , \end{aligned} \quad (98)$$

where (a) follows since $P' \in \mathcal{P}_{s,s'}(\widehat{P}, P_X, R)$ implies that $-\mathbb{D}(P' \| W_{Y|X,s'} | P_X) - \mathbb{H}(P' | P_X) \geq -\beta(\widehat{P}, P_X, R, s)$ and we use the fact that $P_X \circ P' = P_X \circ \widehat{P}$.

Finally, by taking the minimum over all s and applying the fact that $\rho_{\text{joint}}(P_X, R)$ is the minimum of the right-hand sides of (93) and (98), we conclude that

$$\rho_{\text{joint}}(P_X, R) \geq \min_s \min_{\widehat{P} \in \mathcal{P}'_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \gamma_1(\widehat{P}, R, s), \quad \min_{\widehat{P} \in \mathcal{P}''_{\mathcal{Y}|\mathcal{X}}(P_X, R, s)} \gamma_2(\widehat{P}, R, s), \quad (99)$$

where for all \widehat{P} , R and s ,

$$\begin{aligned} \gamma_1(\widehat{P}, R, s) &\triangleq \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \mathbb{I}(P_X, \widehat{P}) - R \right|^+ \\ \gamma_2(\widehat{P}, R, s) &\triangleq \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \left| \mathbb{D}(\widehat{P} \| W_{Y|X,s} | P_X) + \mathbb{H}(P_X \circ \widehat{P}) - \beta(P_X, R, s) - R \right|^+. \end{aligned}$$

REFERENCES

- [1] M.-C. Chang, T. Erdogan, S.-Y. Wang, and M. R. Bloch, “Rate and detection error-exponent tradeoffs of joint communication and sensing,” in *Proc. of IEEE International Symposium on Joint Communications & Sensing*, Vienna, Austria, Mar. 2022, pp. 1–6.
- [2] B. Paul, A. R. Chiriyath, and D. W. Bliss, “Survey of RF communications and sensing convergence research,” *IEEE Access*, vol. 5, pp. 252–270, 2017.
- [3] G. Fettweis, M. Schlüter, R. Thomä, H. Boche, H. Schotten, A. Barreto, A. Mueller, A. Bilgic, C. Hildebrand, D. Schupke, E. Grass, F. Hofmann, G. Kahmen, S. Blicker, I. Korthals, H. J. Einsiedler, I. Rechenberger, G. Zimmermann, H. Heiss, U. Klepser, M. Claus, H. Seidel, J. Goerlich, L. Lemke, M. Kottkamp, M. Meyer, P. Sen, P. Scheele, P. Bender, A. Deder, R. Quay, S. Stanczak, S. Jovanoska, S. Köpsell, T. Dallmann, T. Multerer, T. Neugebauer, T. Wild, U. Barth, W. Heinrich, M. Krstic, G. D. Galdo, R. Salman, S. Brüggenwirth, and R. Irmer, “Joint communication & sensing,” VDE, VDE Positionspapier, Jul. 2021.
- [4] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, “Inner bounds on performance of radar and communications co-existence,” *IEEE Transactions on Signal Processing*, vol. 64, no. 2, pp. 464–474, Jan. 2016.
- [5] L. Zheng, M. Lops, Y. C. Eldar, and X. Wang, “Radar and communication coexistence: An overview: A review of recent methods,” *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 85–99, Sep. 2019.
- [6] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo, “Joint radar and communication design: Applications, state-of-the-art, and the road ahead,” *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3834–3862, Jun. 2020.
- [7] T. Wild, V. Braun, and H. Viswanathan, “Joint design of communication and sensing for beyond 5g and 6g systems,” *IEEE Access*, vol. 9, pp. 30 845–30 857, 2021.
- [8] Y. H. Kim, A. Sutivong, and T. M. Cover, “State amplification,” *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1850–1859, May 2008.
- [9] C. Choudhuri, Y.-H. Kim, and U. Mitra, “Causal state amplification,” in *Proc. of IEEE International Symposium on Information Theory*, Saint-Petersburg, Russia, Aug. 2011, pp. 2110–2114.
- [10] W. Zhang, S. Vedantam, and U. Mitra, “Joint transmission and state estimation: A constrained channel coding approach,” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 7084–7095, Oct. 2011.
- [11] M. Kobayashi, G. Caire, and G. Kramer, “Joint state sensing and communication: Optimal tradeoff for a memoryless case,” in *Proc. of IEEE International Symposium on Information Theory*, Vail, CO, Jun. 2018, pp. 111–115.

- [12] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An information-theoretic approach to joint sensing and communication," *IEEE Transactions on Information Theory*, pp. 1–1, 2022.
- [13] M. Kobayashi, H. Hamad, G. Kramer, and G. Caire, "Joint state sensing and communication over memoryless multiple access channels," in *Proc. of IEEE International Symposium on Information Theory*, Paris, France, Jul. 2019, pp. 270–274.
- [14] M. Ahmadipour, M. Wigger, and M. Kobayashi, "Joint sensing and communication over memoryless broadcast channels," in *Proc. of IEEE Information Theory Workshop*, Apr. 2021.
- [15] N. Su, F. Liu, and C. Masouros, "Secure radar-communication systems with malicious targets: Integrating radar, communications and jamming functionalities," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 83–95, jan 2021.
- [16] Z. Ren, L. Qiu, J. Xu, and D. W. K. Ng, "Robust transmit beamforming for secure integrated sensing and communication," *arXiv preprint*, vol. 2207.13863, Jul. 2022.
- [17] O. Günlü, M. R. Bloch, R. Schaefer, and A. Yener, "Secure joint communication and sensing," in *Proc. of IEEE International Symposium on Information Theory*, Helsinki, Finland, Jun. 2022, pp. 844–849.
- [18] Y. Xiong, F. Liu, Y. Cui, W. Yuan, and T. X. Han, "Flowing the information from shannon to fisher: Towards the fundamental tradeoff in ISAC," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, Dec. 2022, pp. 5601–5606.
- [19] H. Joudeh and F. M. J. Willems, "Joint communication and binary state detection," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 1, pp. 113–124, Mar. 2022.
- [20] H. Wu and H. Joudeh, "On joint communication and channel discrimination," in *Proc. of IEEE International Symposium on Information Theory*, Espoo, Finland, Jun. 2022, pp. 3321–3326.
- [21] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled Sensing for Multihypothesis Testing," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2451–2464, Oct. 2013.
- [22] M. Naghshvar and T. Javidi, "Sequentiality and adaptivity gains in active hypothesis testing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 768–782, Oct. 2013.
- [23] —, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, Dec. 2013.
- [24] A. Tchamkerten and I. E. Telatar, "On the use of training sequences for channel estimation," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1171–1176, Mar. 2006.
- [25] A. Mahajan and S. Tatikonda, "Opportunistic capacity and error exponent region for the compound channel with feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4331–4341, Jul. 2012.
- [26] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3629–3644, 2009.
- [27] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge: Cambridge University Press, Jan. 2011.
- [28] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.
- [29] R. Gallager, "Fixed composition arguments and lower bounds to error probability," <http://web.mit.edu/gallager/www/notes/notes5.pdf>, online; accessed 8 October 2022.