Joint Quantum Communication and Sensing

Shi-Yuan Wang*, Tuna Erdoğan*, Uzi Pereg†‡, and Matthieu R. Bloch*
*School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
†Institute for Communications Engineering, Technical University of Munich, Germany
‡Munich Center for Quantum Science and Technology (MCQST)
Email: {shi-yuan.wang,etuna3,matthieu}@gatech.edu, uzi.pereg@tum.de

***Abstract*—To capture the problem of joint communication and sensing in the quantum regime, we consider the problem of reliably communicating over a Classical-Quantum (c-q) channel that depends on a random parameter while simultaneously estimating the random parameter at the transmitter through a noisy feedback channel. Specifically, for non-adaptive estimation strategies, we obtain an exact characterization of the optimal tradeoffs between the rate of communication and the error exponent of parameter estimation. As in the classical setting, the tradeoff is governed by the empirical distribution of the codewords, which simultaneously controls the rate of reliable communication and the error exponent.**

I. Introduction

The use of dual-purpose signal waveforms for communication and sensing is attracting growing interest, motivated in large part by the convergence of frequencies for radar and communication in the mmWave range [1], [2]. Accordingly, there have been attempts at characterizing the fundamental information-theoretic limits of joint communication and sensing, with the hope of quantifying and developing insights into the optimal tradeoffs incurred by the joint objectives.

From an information-theoretic perspective, joint communication and sensing can be modeled as communication over a channel that depends on an unknown random parameter, in which the objective is to not only ensure reliable communication but also accurately estimate the parameter. In this context, the channel parameter abstracts properties of the environment, e.g., the presence of an obstacle, that affect the communication and this modeling allows one to leverage results on joint communication and state estimation [3].[1] Broadly speaking, results obtained thus far fall into two categories. If the parameter to estimate has independent and identically distributed (i.i.d.) statistics and is known strictly causally at the transmitter, the information-theoretic limits of joint communication and sensing take the form of a rate/distortion region capturing the tradeoff between the rate of reliable communication and the minimum distortion incurred when estimating the channel parameter. This region has been characterized for several models, including multi-user networks [4]–[7]. One should note that no prediction is possible when the channel parameter has i.i.d. statistics, so that coding strategies are inherently open-loop and do not exploit the estimation to adapt the communication. On the other hand, if the parameter to estimate is a fixed parameter taking finitely many values, the information-theoretic limits take the form of a rate/error exponent region capturing the tradeoff between the rate of reliable communication and the exponential speed at which the detection error probability decays. The open-loop region has been characterized for discrete memoryless channels [8] (see also [9], [10] for special cases, including Gaussian channels) and leverages insights from controlled sensing and active hypothesis testing [11], [12]. The previous work [8] also highlights the significant performance gains brought by closed-loop strategies, in which the estimation informs the communication.

The objective of our present work is to extend the study initiated in [8] to the quantum regime. While mostly theoretical in nature, our study is motivated by a joint communication sensing operation at very low power, for which quantum effects need to be accounted for. Specifically, we consider the problem of joint quantum communication and sensing over a c-q channel that depends on an a priori unknown random parameter, which brings together two active lines of work in quantum information theory. On one hand, there has been recent interest in studying communication over c-q channel that depends on parameters [13]. On the other hand, there has been a regain of interest for the study of quantum state discrimination [14]–[18]. The remainder of the paper is organized as follows. After a brief review of notation in Section II, we formally introduce the c-q channel model of interest in Section III. The characterization of the rate-exponent region for open-loop strategies constitutes our main result and is given in Section IV along with numerical examples. We develop proof sketches in Section V.

II. Notation

We use logarithms with base e. 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 by $\mathbf{x} \triangleq (x_1, \cdots, x_n) \in \mathcal{X}^n$, while $x^i \triangleq (x_1, \cdots, x_i) \in \mathcal{X}^i$ denotes a sequence of length i. 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 sequence $\mathbf{x} \in \mathcal{X}^n$ such $\hat{p}_{\mathbf{x}} = P$. Finally, $\mathcal{P}_{\mathcal{X},n}$ is the set

S.-Y. Wang, T. Erdoğan, and M. R. Bloch was suppoted by NSF award 1910859. U. Pereg was supported by DFG No. EXC-2111 – 390814868, BMBF No. 16KISQ028, and Israel CHE Fellowship for Quantum Science and Technology.

[1] In communication and information theory, the channel parameter is often called a "state;" we will not follow this terminology to avoid confusion with quantum states.

of all possible types of length n sequence on $\mathcal{X}^n$. Moreover, for two integers $\lfloor a \rfloor$ and $\lceil b \rceil$ such that $\lfloor a \rfloor \leqslant \lceil b \rceil$, we define $[a;b] \triangleq \{\lfloor a \rfloor, \lfloor a \rfloor + 1, \cdots, \lceil b \rceil - 1, \lceil b \rceil\}$; otherwise $[a;b] \triangleq \emptyset$.

Let $\mathcal{H}$ be a finite-dimensional Hilbert space. Let $\mathcal{P}(\mathcal{H})$ and $\mathcal{D}(\mathcal{H})$ be the sets of all positive operators and all density operators acting on $\mathcal{H}$, respectively. Given an ensemble $\mathcal{E} \triangleq \{P_X, \rho_B^x\}$, we define the Holevo information $\chi(\mathcal{E}) \triangleq \mathbb{H}(\rho_B) - \sum_x P_X(x)\mathbb{H}(\rho_B^x)$. In particular, if $\mathcal{E}$ corresponds to a c-q channel $\mathcal{N}_{X\to B}(x)$ with an input distribution P_X, we also write $\mathbb{I}(P_X, \mathcal{N}_{X\to B}) \triangleq \chi(\mathcal{E})$. For $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ and $s \in (0,1)$, the Petz-Renyi divergence [19] is defined as $\mathbb{D}_s(\rho \,\|\, \sigma) \triangleq \frac{1}{s-1} \log \mathrm{tr}\left(\rho^s \sigma^{1-s}\right)$. For any Hermitian linear operator X, $\{X > 0\}$ denotes the projection onto the eigenspaces of positive eigenvalues.

III. Joint Quantum Communication and Sensing Model

As illustrated in Fig. 1, we consider a c-q channel that depends on a random parameter $\theta \in \Theta$ described by $\mathcal{N}_{X\to AB}^{(\theta)} : \mathcal{X} \to \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, where $|\mathcal{X}| < \infty$ and $|\Theta| < \infty$. Without loss of generality, we assume the set Θ is partially ordered. For every value of $\theta \in \Theta$, the channel maps a classical input symbol x to a density operator $\rho_{AB}^{x,\theta}$. The system A captures observations obtained by the transmitter (Alice) while the system B captures observations obtained by the receiver (Bob). The prior probabilities $\{p_\theta\}_{\theta\in\Theta}$ of the unknown parameter θ are known and θ is assumed fixed during the entire transmission. The objective is for Alice to simultaneously transmit information to Bob while estimating the unknown parameter θ. Specifically, Alice encodes a uniformly-distributed message $W \in [1;M]$ with a known classical encoding function $f : [1;M] \to \mathcal{X}^n$ to create a classical codeword of length n, $X^n \triangleq f(W)$, and transmits X^n over the c-q channel. Alice receives the state

$$\rho_{A^n}^{f(W),\theta} \triangleq \mathrm{tr}_{B^n}\left(\mathcal{N}_{X\to AB}^{(\theta)\otimes n}(f(W))\right) \tag{1}$$

and uses a collection of Positive Operator-Valued Measures (POVMs) $\{\{\Pi_\theta^{(w)}\}_{\theta\in\Theta}\}_{w\in[1;M]}$ (i.e., for every $w \in [1;M]$, $\{\Pi_\theta^{(w)}\}_{\theta\in\Theta}$ is a POVM), to obtain an estimate of the channel parameter $\hat{\theta}$. On the other hand, Bob uses a POVM $\{\Lambda_w\}_{w\in[1;M]}$ on his observations

$$\rho_{B^n}^{f(W),\theta} \triangleq \mathrm{tr}_{A^n}\left(\mathcal{N}_{X\to AB}^{(\theta)\otimes n}(f(W))\right) \tag{2}$$

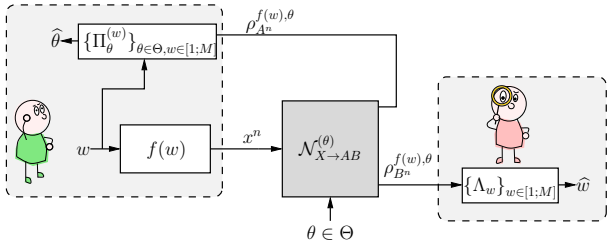


Fig. 1: Joint quantum communication and sensing model.

to obtain an estimate $\widehat{W}$ of the transmitted message. A code $\mathcal{C}$ then consists of the tuple $(n, M, f, \{\{\Pi_\theta^{(w)}\}_{\theta\in\Theta}\}_{w\in[1;M]}, \{\Lambda_w\}_{w\in[1;M]})$.

Remark III.1. *Our code definition does not allow adaptivity, and we refer to the coding strategy as open-loop. The study of closed-loop strategies will be the topic of a forthcoming work.*

The performance of the code is measured in terms of the detection-error and communication-error probabilities defined as

$$P_e^*\left(\{\{\Pi_\theta^{(w)}\}_\theta\}_w\right) \triangleq \max_{\theta\in\Theta}\max_{w\in[1;M]} \mathrm{tr}\left((\mathbf{I} - \Pi_\theta^{(w)})\rho_{A^n}^{f(w),\theta}\right) \tag{3}$$

$$P_c^{(n)}\left(\{\Lambda_w\}_w\right) \triangleq \max_{\theta\in\Theta}\max_{w\in[1;M]} \mathrm{tr}\left((\mathbf{I} - \Lambda_w)\,\rho_{B^n}^{f(w),\theta}\right), \tag{4}$$

respectively. The rate and detection-error exponent are

$$\frac{1}{n}\log M \quad \text{and} \quad E_{\mathrm{d}}^{(n)} \triangleq -\frac{1}{n}\log P_e^*, \tag{5}$$

respectively.

Definition III.2 (Achievable pair)**.** *A rate/detection-error exponent (R, E) is achievable if for every $\epsilon > 0$, there exist a sufficiently large n and a code $\mathcal{C}$ of length n such that*

$$P_{\mathrm{c}}^{(n)} \leqslant \epsilon, E_{\mathrm{d}}^{(n)} \geqslant E - \epsilon, \text{ and } \frac{1}{n}\log M \geqslant R - \epsilon. \tag{6}$$

For the analysis of the detection error, we also introduce the average Bayesian and maximal error probabilities as follows:

$$P_e(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta\}_\theta) \triangleq \sum_{\theta\in\Theta} \mathrm{tr}\left(p_\theta \rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \Pi_\theta)\right), \tag{7}$$

$$P_e^*(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_{\theta\in\Theta}; \{\Pi_\theta\}_\theta) \triangleq \max_{\theta\in\Theta} \mathrm{tr}\left(p_\theta \rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \Pi_\theta)\right). \tag{8}$$

In terms of asymptotics, the error exponents of these two probabilities are the same for n large enough, as (7) would be asymptotically dominated by (8).

IV. Main Result

Our main result is an exact characterization of the region of achievable rate/error-exponent pairs.

Theorem IV.1. *The set of achievable rate/detection-error exponent pairs is*

$$\bigcup_{P_X\in\mathcal{P}_\mathcal{X}} \left\{ \begin{array}{l} (R,E) \in \mathbb{R}_+^2 : \\ R \leqslant \min_{\theta\in\Theta} \mathbb{I}\left(P_X, \mathcal{N}_{X\to B}^{(\theta)}\right) \\ E \leqslant \phi(P_X) \end{array} \right\}$$

with

$$\phi(P_X) = \min_\theta \min_{\theta'\neq\theta} \sup_{s\in[0,1]} \sum_x P_X(x)(1-s)\mathbb{D}_s(\rho_A^{x,\theta} \| \rho_A^{x,\theta'})$$

and $\rho_{AB}^{x,\theta} = \mathcal{N}_{X\to AB}^{(\theta)}(x)$.

As expected, the result is the direct generalization of our result reported for classical channels in [8, Theorem 3]. As detailed in Section V-A, although the proof requires a specific new approach relying in part on [16, Theorem 2]

TABLE I: $p_\theta^x(0)$ for all x and θ.

θ \ x	0	1
0	0.9	0.3
1	0.9	0.2
2	0.7	0.2

TABLE II: $p_\theta^x(0)$ for all x and θ.

θ \ x	0	1
0	0.9	0.1
1	0.8	0.2
2	0.7	0.3

for discriminating multiple hypotheses. Theorem IV.1 shows once more that the tradeoff between rate and error exponent is governed by the type of the codewords.

Remark IV.2. *Our result for the error exponent, which is in fact the quantum Chernoff information, is different from [17, Proposition 32], which is not the tightest characterization even without adaptivity. The main reason is their derivation for the upper bound relies on the divergence sphere optimization [20], and the log-Euclidean Renyi divergence [21] turns out to be the optimization solution. However, our exponent is consistent with the upper bound obtained in [18] for the c-q channel. As detailed later in the proof, regardless of adaptivity, we both employ a classical interpretation (See also [18, Lemma 1] and [14]) to construct a classical hypothesis testing properly lower-bounding the error performance for any quantum test.*

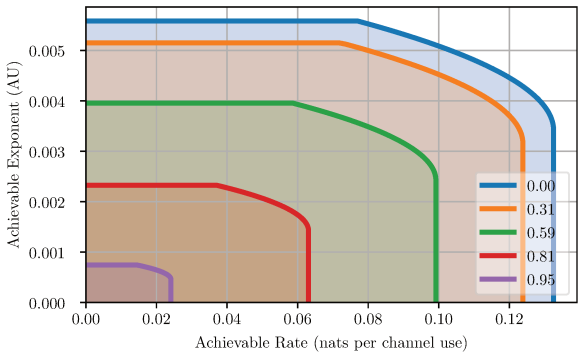
We illustrate Theorem IV.1 with a few numerical examples. Let $\mathcal{X} = \{0, 1\}$, $\Theta = \{0, 1, 2\}$,

$$|\psi_0\rangle_A \triangleq \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \text{ and } |\psi_1\rangle_A \triangleq R(\phi)|\psi_0\rangle,$$

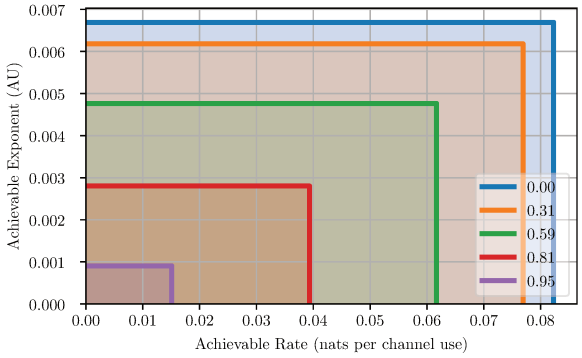
where $R(\phi)$ is a rotation matrix on $\mathbb{R}^2$ with angle ϕ. Then, we construct $\rho_A^{x,\theta}$ according to Table I and II as follows:

$$\rho_A^{x,\theta} = p_\theta^x(0)|\psi_0\rangle\langle\psi_0|_A + p_\theta^x(1)|\psi_1\rangle\langle\psi_1|_A. \tag{9}$$

The construction for $\rho_B^{x,\theta}$ is the same. The numerical result of the rate/detection-error exponent region corresponding to the c-q channel defined by (9) and Table I is shown in Fig. 2a, where different curves correspond to different rotation angles and labels indicate the inner product $|\langle\psi_0|\psi_1\rangle| = \cos\phi$. This example is one in which there exist tradeoffs, and the inner products control the areas of the tradeoff regions. Alternatively, if we choose the c-q channel to be defined by (9) and Table II, Fig. 2b shows an example in which there exist no tradeoff, since the optimal error exponent and the optimal rate can be simultaneously achieved by the same P_X.



(a) Region corresponding to (9) and Table. I.



(b) Region corresponding to (9) and Table. II.

Fig. 2: Illustrations of rate/detection-error exponent region.

V. Proof of Theorem IV.1

A. Achievability Proof

In the achievability, we show that it suffices to use a sequence of *constant-composition* codes to achieve the optimal trade-off between reliable rate and Chernoff information. The rate/detection-error exponent region is then characterized by the type of these constant-composition codewords. Let all the codewords of $\mathcal{C}$ have type P_X.

a) Rate analysis: We first characterize the performance of this constant-composition code in terms of code rate and error probability. [22, Theorem 1] provides us with an explicit construction such that, if all the codewords belong to the type P_X with code rate $R < \min_{\theta\in\Theta} \mathbb{I}\left(P_X, \mathcal{N}_{X\to B}^{(\theta)}\right)$, then there exists a sequence of codes such that $P_c^{(n)} \leqslant \epsilon$ for any $\epsilon > 0$ and n large enough.

b) Error exponent analysis: We show that the detection-error exponent only depends on the type of the codewords. We first recall a one-shot bound on the Bayesian probability of error for discriminating between r quantum states from [16, Theorem 2].

Theorem V.1 ([16]). *Let $\sigma_1, \cdots, \sigma_r \in \mathcal{P}(\mathcal{H})$. For all $1 \leqslant i \leqslant r$, let $\sigma_i = \sum_{k=1}^{T_i} \lambda_{ik} Q_{ik}$ be the spectral decomposition of σ_i, and $T \triangleq \max\{T_1, \cdots, T_r\}$. There exist a function $f(r, T)$ and a POVM $\{\Pi_i\}_{i\in[1;r]}$ such that*

$$\begin{aligned} &P_e\left(\{\sigma_1, \cdots, \sigma_r\}; \{\Pi_1, \cdots, \Pi_r\}\right) \\ &\quad\leqslant f(r, T) \sum_{(i,j):i<j} \sum_{k,\ell} \min\{\lambda_{ik}, \lambda_{j\ell}\} \mathrm{tr}\left(Q_{ik} Q_{jl}\right) \end{aligned} \tag{10}$$

and we have $f(r, T) < 10(r-1)^2 T^2$.

For a fixed message w and its codeword $f(w) = \mathbf{x} \in \mathcal{T}^n_{P_X}$, applying Theorem V.1 to the operators

$$\sigma_\theta \triangleq p_\theta \bigotimes_{k=1}^{n} \rho_A^{x_k,\theta} \tag{11}$$

for $\theta \in \Theta$, we obtain that there exists a POVM $\{\Pi_\theta^{(w)}\}$ such that for any $s \in [0,1]$,

$$\begin{aligned}
&P_e(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta^{(w)}\}_\theta) \\
&\leqslant f(|\Theta|,T) \sum_\theta \sum_{\theta'>\theta} \sum_{k,\ell} \min(p_\theta \lambda_{\theta k}, p_{\theta'} \lambda_{\theta',\ell}) \mathrm{tr}\left(Q_{\theta k} Q_{\theta' \ell}\right) \\
&\leqslant f(|\Theta|,T) \sum_\theta \sum_{\theta'>\theta} \sum_{k,\ell} \max(p_\theta, p_{\theta'}) \lambda_{\theta k}^s \lambda_{\theta',\ell}^{1-s} \mathrm{tr}\left(Q_{\theta k} Q_{\theta' \ell}\right) \\
&= f(|\Theta|,T) \sum_\theta \sum_{\theta'>\theta} \max(p_\theta, p_{\theta'}) \mathrm{tr}\left((\rho_{A^n}^{\mathbf{x},\theta})^s (\rho_{A^n}^{\mathbf{x},\theta'})^{1-s}\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&P_e^*(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta^{(w)}\}_\theta) \leqslant P_e(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta^{(w)}\}_\theta) \\
&\leqslant f(|\Theta|,T) \binom{|\Theta|}{2} \max(\{p_\theta\}_{\theta\in\Theta}) \times \\
&\quad \max_\theta \max_{\theta' \neq \theta} \inf_{s\in[0,1]} \mathrm{tr}\left((\rho_{A^n}^{\mathbf{x},\theta})^s (\rho_{A^n}^{\mathbf{x},\theta'})^{1-s}\right).
\end{aligned}$$

Note that $f(|\Theta|,T) < 10(|\Theta|-1)^2 T^2$, where T in our case will be determined by the number of different types conditioned on the given input codeword $\mathbf{x}$. Since the channel input $\mathbf{x}$ is classical, one can therefore conclude that $T \leqslant (n+1)^{|\mathcal{H}||\mathcal{X}|}$ by the type counting lemma [23, Lemma 2.2], and it immediately follows that this bound is universal for any choice of input $\mathbf{x}$ and grows polynomially with n.

From (11), we obtain

$$\mathrm{tr}\left((\rho_{A^n}^{\mathbf{x},\theta})^s (\rho_{A^n}^{\mathbf{x},\theta'})^{1-s}\right) = \prod_{k=1}^{n} \mathrm{tr}\left((\rho_A^{x_k,\theta})^s (\rho_A^{x_k,\theta'})^{1-s}\right).$$

Observe that

$$\min_{\theta'} p_{\theta'} P_e^*(\{\rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta^{(w)}\}_\theta) \leqslant P_e(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta^{(w)}\}_\theta).$$

Since $\mathbf{x}$ has type P_X, we get, $\forall s \in [0,1]$

$$\begin{aligned}
&-\frac{1}{n} \log P_e^*(\{\rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta^{(w)}\}_\theta) \\
&\geqslant \min_\theta \min_{\theta'\neq\theta} \sup_{s\in[0,1]} -\sum_{k=1}^{n} \log \mathrm{tr}\left((\rho_A^{x_k,\theta})^s (\rho_A^{x_k,\theta'})^{1-s}\right) - \epsilon \\
&= \min_\theta \min_{\theta'\neq\theta} \sup_{s\in[0,1]} -\sum_{u} P_X(u) \log \mathrm{tr}\left((\rho_A^{u,\theta})^s (\rho_A^{u,\theta'})^{1-s}\right) - \epsilon \\
&= \min_\theta \min_{\theta'\neq\theta} \sup_{s\in[0,1]} \sum_u P_X(u)(1-s)\mathbb{D}_s\left(\rho_A^{u,\theta} \| \rho_A^{u,\theta'}\right) - \epsilon,
\end{aligned}$$

where we have used the fact that there exists a universal $\epsilon > 0$ (i.e., no dependency on $\mathbf{x}$) such that for n large enough,

$$\frac{\log f(|\Theta|,T)}{n} + \frac{\log \binom{|\Theta|}{2}}{n} - \min_{\theta'} \frac{\log p_{\theta'}}{n} \leqslant \epsilon.$$

Thus,

$$\begin{aligned}
&-\frac{1}{n} \log P_e^*(\{\rho_{A^n}^{\mathbf{x},\theta}\}_{\theta\in\Theta}; \{\Pi_\theta^{(w)}\}_{\theta\in\Theta}) \\
&\quad \geqslant \min_\theta \min_{\theta'\neq\theta} \sup_{s\in[0,1]} \sum_u P_X(u)(1-s)\mathbb{D}_s\left(\rho_A^{u,\theta} \| \rho_A^{u,\theta'}\right) - \epsilon.
\end{aligned}$$

Therefore, as we are using a constant-composition code with type P_X, $E_d^{(n)} \geqslant \phi(P_X) - \epsilon$.

c) Rate/detection-error exponent region: Taking the union over all possible $P_X \in \mathcal{P}_\mathcal{X}$ and choosing $\epsilon > 0$ arbitrarily small for n large enough, we obtain the result.

B. Converse Proof

Assume that the rate/detection-error exponent pair (R, E) is achievable. That is, for all $\epsilon > 0$, there exists $N_\epsilon > 0$ and a code $\mathcal{C}$ with block length $n \geqslant N_\epsilon$ such that

$$P_c^{(n)} \leqslant \epsilon, E_d^{(n)} \geqslant E - \epsilon, \text{ and } \frac{\log M}{n} \geqslant R - \epsilon.$$

a) Rate analysis: We first identify a set $\mathcal{T}$ of types to which exponentially and significantly many codewords of $\mathcal{C}$ belong. Since the channel input X is classical, the type counting lemma for classical types asserts that there are at most a polynomial number of classical types of sequence X^n [23, Lemma 2.2]. Then, there exists a set of types $\mathcal{T}$ such that, for all $P_X \in \mathcal{T}$, the subcode $\mathcal{C}_{P_X} \triangleq \{f(w) : \hat{p}_{f(w)} = P_X\} \subset \mathcal{C}$ satisfies $\max_{\theta\in\Theta} \max_{w\in f^{-1}(\mathcal{C}_{P_X})} \mathrm{tr}\left((\mathbf{I} - \Lambda_w)\rho_{B^n}^{\theta,f(w)}\right) \leqslant \epsilon$ and $\frac{\log M_{P_X}}{n} > \frac{\log M}{n} - \delta \geqslant R - \epsilon - \delta$ for some $\delta > 0$ vanishing with ϵ. Fix any $P_X \in \mathcal{T}$. Let

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

Observe that the average type

$$\overline{P}_X(x) \triangleq \frac{1}{n} \sum_{i=1}^{n} \widetilde{P}_{X_i}(x) = P_X(x), \tag{12}$$

where $\widetilde{P}_{X_i}$ is the i-th marginal distribution of $\widetilde{P}_X^n$. Then, for any θ,

$$\begin{aligned}
&(1-\epsilon)\log M_{P_X} - h_b(\epsilon) \\
&\overset{(a)}{\leqslant} \chi(\{\widetilde{P}_X^n, \rho_{B^n}^{\mathbf{x},\theta}\}) &(13)\\
&\overset{(b)}{=} \mathbb{H}(\rho_{B^n}^\theta) - \sum_{\mathbf{x}\in\mathcal{X}^n} \widetilde{P}_X^n(\mathbf{x}) \sum_{i=1}^{n} \mathbb{H}\left(\rho_B^{x_i,\theta}\right) &(14)\\
&\overset{(c)}{\leqslant} \sum_{i=1}^{n} \mathbb{H}\left(\rho_B^{\theta,(i)}\right) - \sum_{i=1}^{n} \sum_{x_i\in\mathcal{X}} \widetilde{P}_{X_i}(x_i) \mathbb{H}\left(\rho_B^{x_i,\theta}\right) &(15)\\
&= \sum_{i=1}^{n} \chi(\{\widetilde{P}_{X_i}, \rho_B^{x,\theta}\}) \overset{(d)}{\leqslant} n\chi(\{\overline{P}_X, \rho_B^{x,\theta}\}), &(16)
\end{aligned}$$

where $h_b(\cdot)$ is the binary entropy function, (a) follows from Fano's inequality and the Holevo bound, (b) follows by defining $\rho_{B^n}^\theta$ as the density operator of ensemble $\{\widetilde{P}_X^n, \rho_{B^n}^{\mathbf{x},\theta}\}$ and the definition of Holevo information, (c) follows from the subadditivity of von Neumann entropy and by defining $\rho_B^{\theta,(i)}$ as

the ensemble density operator of $\{\widetilde{P}_{X_i}, \rho_B^{x_i,\theta}\}$, and (d) follows from the concavity of the Holevo information in the input distribution. Then, we obtain

$$\frac{\log M_{P_X}}{n} \leqslant \frac{\chi(\{P_X, \rho_B^{x,\theta}\}) + \frac{C}{n}}{1-\epsilon} = \frac{\mathbb{I}\Big(P_X, \mathcal{N}_{X\to B}^{(\theta)}\Big) + \frac{C}{n}}{1-\epsilon}, \tag{17}$$

where we have used (12) and $h_b(\epsilon) \leqslant \log 2 \triangleq C, \forall \epsilon \in [0,1]$.

Since (17) is valid for any ϵ and θ, the size of subcode $\mathcal{C}_{P_X}$ is upper-bounded by the mutual information in a compound channel sense, i.e.,

$$\frac{\log M_{P_X}}{n} < \min_{\theta} \mathbb{I}(P_X, \mathcal{N}_{X\to B}^{(\theta)}) + \tau \tag{18}$$

for some $\tau > 0$ vanishing with ϵ.

b) Reduction to a Binary Hypothesis Testing: We use the proof technique in [15, Theorem 1] to show that, for any sequence of POVM discriminating the M-ary hypotheses, we can lower bound the detection error by that of a binary hypothesis test between any pair of hypotheses.

For a sequence $\mathbf{x} \in \mathcal{X}^n$, consider a POVM $\{\Pi_\theta^{\mathbf{x}}\}_{\theta\in\Theta}$ and fix a pair of indices $\theta,\ \theta' \in \Theta$. For simplicity, we ignore the dependency of the POVM on $\mathbf{x}$. Let $A,\ B \in \mathcal{P}(\mathcal{H}_A^{\otimes n})$ such that $A + B = \mathbf{I} - \Pi_\theta - \Pi_{\theta'}$. We construct operators $\widetilde{\Pi}_\theta \triangleq \Pi_\theta + A$ and $\widetilde{\Pi}_{\theta'} \triangleq \Pi_{\theta'} + B$ to form a POVM for the binary hypothesis testing for θ against θ'.

We next show that the error of this binary hypothesis test forms a valid lower bound for the original problem. Indeed, since $\widetilde{\Pi}_\theta \succeq \Pi_\theta$, we have by monotonicity that $\text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \widetilde{\Pi}_\theta)\right) \leqslant \text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \Pi_\theta)\right)$. Then,

$$\begin{aligned}
&P_e\left(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}; \{\Pi_\theta\}_\theta\right) && (19)\\
&\geqslant p_\theta \text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \Pi_\theta)\right) + p_{\theta'}\text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta'}(\mathbf{I} - \Pi_{\theta'})\right) && (20)\\
&\geqslant p_\theta \text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \widetilde{\Pi}_\theta)\right) + p_{\theta'}\text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta'}(\mathbf{I} - \widetilde{\Pi}_{\theta'})\right) && (21)\\
&\geqslant \min\{p_\theta, p_{\theta'}\}\left(\text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \widetilde{\Pi}_\theta)\right) + \text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta'}(\mathbf{I} - \widetilde{\Pi}_{\theta'})\right)\right) && (22)\\
&\overset{(a)}{=} \min\{p_\theta, p_{\theta'}\}\left(\text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \widetilde{\Pi}_\theta)\right) + \text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta'}\widetilde{\Pi}_\theta\right)\right), && (23)
\end{aligned}$$

where (a) follows since $\{\widetilde{\Pi}_\theta, \mathbf{I} - \widetilde{\Pi}_\theta\}$ is a POVM by construction.

c) Reduction to classical hypothesis testing: We now construct a lower bound on the error probability for any binary hypothesis testing between $\rho_{A^n}^{\mathbf{x},\theta}$ and $\rho_{A^n}^{\mathbf{x},\theta'}$. Let $\{\Lambda_\theta, \mathbf{I} - \Lambda_\theta\}$ be *any* POVM for discriminating between $\rho_{A^n}^{\mathbf{x},\theta}$ and $\rho_{A^n}^{\mathbf{x},\theta'}$. Observe that

$$\begin{aligned}
&\text{tr}\left((\mathbf{I} - \Lambda_\theta)\rho_{A^n}^{\mathbf{x},\theta}\right) + \text{tr}\left(\Lambda_\theta \rho_{A^n}^{\mathbf{x},\theta'}\right)\\
&\geqslant 1 - \max_{\mathbf{0}\leqslant\Lambda\leqslant\mathbf{I}} \text{tr}\left(\Lambda\left(\rho_{A^n}^{\mathbf{x},\theta'} - \rho_{A^n}^{\mathbf{x},\theta}\right)\right) && (24)\\
&= \text{tr}\left((\mathbf{I} - \Lambda^*)\rho_{A^n}^{\mathbf{x},\theta}\right) + \text{tr}\left(\Lambda^*\rho_{A^n}^{\mathbf{x},\theta'}\right), && (25)
\end{aligned}$$

where we note that $\max_{\mathbf{0}\leqslant\Lambda\leqslant\mathbf{I}} \text{tr}\left(\Lambda\left(\rho_{A^n}^{\mathbf{x},\theta'} - \rho_{A^n}^{\mathbf{x},\theta}\right)\right) = \frac{1}{2}\left\|\rho_{A^n}^{\mathbf{x},\theta'} - \rho_{A^n}^{\mathbf{x},\theta}\right\|_1$ and the optimizer $\Lambda^* \triangleq \{\rho_{A^n}^{\mathbf{x},\theta'} - \rho_{A^n}^{\mathbf{x},\theta} > 0\}$, which is the Holevo-Helstrom test [24, Chapter IV.2] and [25], is indeed a Projection-Valued Measure (PVM). Without loss of generality, we shall develop a lower bound on the error probability for any PVM $\{\Gamma_\theta, \mathbf{I} - \Gamma_\theta\}$, which then also holds for the optimal test Λ^*.

Lemma V.2. *Fix $\theta, \theta' \in \Theta$. For any sequence $\mathbf{x} \in \mathcal{X}^n$ and any PVM $\{\Gamma_\theta, \mathbf{I} - \Gamma_\theta\}$, we have, for any $\xi > 0$ and n large enough,*

$$\begin{aligned}
&\alpha_n + \beta_n \geqslant \left(\frac{1}{2} - \xi\right) \times\\
&\exp\left(-n \sup_{s\in[0,1]} \sum_u \hat{p}_{\mathbf{x}}(u)(1-s)\mathbb{D}_s\Big(\rho_A^{u,\theta}\Big\|\rho_A^{u,\theta'}\Big)\right), \quad (26)
\end{aligned}$$

where $\alpha_n \triangleq \text{tr}\left((\mathbf{I} - \Gamma_\theta)\rho_{A^n}^{\mathbf{x},\theta}\right)$ and $\beta_n \triangleq \text{tr}\left(\Gamma_\theta \rho_{A^n}^{\mathbf{x},\theta'}\right)$.

Proof: The proof follows from the techniques developed in [14, Theorem 2.2] to construct a classical binary hypothesis as a lower bound on error and [11, Appendix A] to analyze this lower bound. ∎

Thus, by combining (26) with (23) and (25),

$$\begin{aligned}
&P_e\left(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_\theta; \{\Pi_\theta\}_\theta\right) \geqslant \max_\theta \max_{\theta\neq\theta'} \min\{p_\theta, p_{\theta'}\}\left(\frac{1}{2} - \xi\right) \times\\
&\exp\left(-n \sup_{s\in[0,1]} \sum_u \hat{p}_{\mathbf{x}}(u)(1-s)\mathbb{D}_s\Big(\rho_A^{u,\theta}\Big\|\rho_A^{u,\theta'}\Big)\right). \quad (27)
\end{aligned}$$

Note that $\sum_{\theta\in\Theta} \text{tr}\left(p_\theta \rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \Pi_\theta)\right) \leqslant \max_{\theta\in\Theta} \text{tr}\left(\rho_{A^n}^{\mathbf{x},\theta}(\mathbf{I} - \Pi_\theta)\right)$. Then, we conclude that for any $\mathbf{x} \in \mathcal{X}^n$ and POVM $\{\Pi_\theta\}_{\theta\in\Theta}$,

$$\begin{aligned}
&-\frac{1}{n}\log P_e^*(\{\rho_{A^n}^{\mathbf{x},\theta}\}_{\theta\in\Theta}, \{\Pi_\theta\}_{\theta\in\Theta})\\
&\leqslant -\frac{1}{n}\log P_e(\{p_\theta \rho_{A^n}^{\mathbf{x},\theta}\}_{\theta\in\Theta}, \{\Pi_\theta\}_{\theta\in\Theta})\\
&\leqslant \min_\theta \min_{\theta'\neq\theta} \sup_{s\in[0,1]} \sum_u \hat{p}_{\mathbf{x}}(u)(1-s)\mathbb{D}_s\Big(\rho_A^{u,\theta}\Big\|\rho_A^{u,\theta'}\Big) + \delta\\
&= \phi(\hat{p}_{\mathbf{x}}) + \delta.
\end{aligned}$$

d) Rate/detection-error exponent region: Since for any codeword $f(w) \in \mathcal{C},\ w \in [1; M]$ and the collection of POVMs $\{\{\Pi_\theta^{(w)}\}_\theta\}_w$, $E - \epsilon \leqslant E_d^{(n)} = -\frac{1}{n}\log \max_{\theta\in\Theta}\max_{w\in[1;M]} \text{tr}\left((\mathbf{I} - \Pi_\theta^{(w)})\rho_{A^n}^{f(w),\theta}\right) \leqslant \phi(\hat{p}_{f(w)}) + \delta$. Choose now $P_X^* \in \mathcal{T}$ such that $P_X^* \triangleq \text{argmin}_{P_X\in\mathcal{T}} \phi(P_X)$. We then obtain that for all $\epsilon > 0$, there exist $\tau, \delta > 0$ vanishing with ϵ,

$$\begin{aligned}
R &\leqslant \min_\theta \mathbb{I}\Big(P_X^*, \mathcal{N}_{X\to B}^{(\theta)}\Big) + \tau + \epsilon + \delta\\
E &\leqslant \phi(P_X^*) + \epsilon + \delta.
\end{aligned}$$

Since ϵ can be chosen arbitrarily small as the block length goes to infinity, E is upper-bounded by $\phi(P_X)$ and the rate is achieved by P_X for some $P_X \in \mathcal{P}_X$. Taking the union over all possible P_X completes the result.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] M. Ahmadipour, M. Wigger, and M. Kobayashi, "Joint sensing and communication over memoryless broadcast channels," in *Proc. of IEEE Information Theory Workshop*, Apr. 2021.

[7] M. Ahmadipour, M. Kobayashi, M. Wigger, and G. Caire, "An Information-Theoretic Approach to Joint Sensing and Communication," *arXiv preprint*, vol. 2107.14264, Jul. 2021.

[8] 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.

[9] 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.

[10] H. Wu and H. Joudeh, "On joint communication and channel discrimination," Feb. 2022, arXiv:2202.01112.

[11] 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.

[12] 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.

[13] U. Pereg, C. Deppe, and H. Boche, "Classical state masking over a quantum channel," *Physical Review A*, vol. 105, no. 2, p. 022442, Feb. 2022.

[14] M. Nussbaum and A. Szkoła, "The Chernoff lower bound for symmetric quantum hypothesis testing," *The Annals of Statistics*, vol. 37, no. 2, pp. 1040–1057, Apr. 2009.

[15] ——, "Asymptotically Optimal Discrimination between Pure Quantum States," in *Theory of Quantum Computation, Communication, and Cryptography*. Springer, Berlin, Heidelberg, 2011, pp. 1–8.

[16] K. Li, "Discriminating quantum states: The multiple chernoff distance," *The Annals of Statistics*, vol. 44, no. 4, Aug. 2016.

[17] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, "Amortized channel divergence for asymptotic quantum channel discrimination," *Letters in Mathematical Physics*, vol. 110, no. 8, pp. 2277–2336, Aug. 2020.

[18] F. Salek, M. Hayashi, and A. Winter, "Usefulness of adaptive strategies in asymptotic quantum channel discrimination," *Physical Review A*, vol. 105, no. 2, p. 022419, Feb. 2022.

[19] D. Petz, "Quasi-entropies for finite quantum systems," *Reports on Mathematical Physics*, vol. 23, no. 1, pp. 57–65, Feb. 1986.

[20] H. Nagaoka and T. Ogawa, "Strong converse and Stein's lemma in quantum hypothesis testing," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2428–2433, Feb. 2000.

[21] M. Mosonyi and T. Ogawa, "Strong Converse Exponent for Classical-Quantum Channel Coding," *Communications in Mathematical Physics*, vol. 355, no. 1, pp. 373–426, Oct. 2017.

[22] M. Hayashi, "Universal Coding for Classical-Quantum Channel," *Communications in Mathematical Physics*, vol. 289, no. 3, pp. 1087–1098, Aug. 2009.

[23] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. Cambridge: Cambridge University Press, Jan. 2011, vol. 9780521196.

[24] C. W. Helstrom, *Quantum detection and estimation theory*, 1st ed. New York, NY, USA: Academic Press, 1976.

[25] A. S. Holevo, "Investigations in the General Theory of Statistical Decisions," *Proceedings of the Steklov Institute of Mathematics*, vol. 124, no. 3, 1978.