Non-Asymptotic Classical Data Compression With Quantum Side Information

Hao-Chung Cheng, Member, IEEE, Eric P. Hanson, Nilanjana Datta, and Min-Hsiu Hsieh, Senior Member, IEEE

Abstract—In this paper, we analyze classical data compression with quantum side information (also known as the classical-quantum Slepian–Wolf protocol) in the so-called large and moderate deviation regimes. In the non-asymptotic setting, the protocol involves compressing classical sequences of finite length $n$ and decoding them with the assistance of quantum side information. In the large deviation regime, the compression rate is fixed, and we obtain bounds on the error exponent function, which characterizes the minimal probability of error as a function of the rate. Devetak and Winter showed that the asymptotic data compression limit for this protocol is given by a conditional entropy. For any protocol with a rate below this quantity, the probability of error converges to zero asymptotically and its speed of convergence is given by the strong converse exponent function. We obtain finite blocklength bounds on this function, and determine exactly its asymptotic value. In the moderate deviation regime for the compression rate, the latter is no longer considered to be fixed. It is allowed to depend on the blocklength $n$, but assumed to decay slowly to the asymptotic data compression limit. Starting from a rate above this limit, we determine the speed of convergence of the error probability to zero and show that it is given in terms of the conditional information variance.

I. INTRODUCTION

SOURCE coding (or data compression) is the task of compressing information emitted by a source in a manner such that it can later be decompressed to yield the original information with high probability. The information source is said to be memoryless if there is no correlation between the successive messages emitted by it. In this case, $n$ successive uses of the source is modeled by a sequence of $n$ independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_n$, each taking values in a finite alphabet $\mathcal{X}$ with probability $p(x)$. Such a source is equivalently modeled by a single random variable $X$ with probability mass function $p(x)$, $x \in \mathcal{X}$, and is called a discrete memoryless source (DMS). Let $H(X)$ denote the Shannon entropy of $X$. Shannon’s Source Coding Theorem [2] tells us that if the messages emitted by $n$ copies of the source are compressed into at least $nH(X)$ bits, then they can be recovered with arbitrary accuracy upon decompression, in the asymptotic limit ($n \to \infty$).

One variant of the above task is that of data compression with classical side information (at the decoder), which is also called Slepian–Wolf coding, first studied by Slepian and Wolf [3]. In this scenario, one considers a memoryless source emitting two messages, $x$ and $y$, which can be considered to be the values taken by a pair of correlated random variables $(X, Y)$. The task is once again to optimally compress sequences $x := (x_1, \ldots, x_n)$ emitted on $n$ copies of the source so that they can be recovered with vanishing probability of error in the asymptotic limit. However, at the recovery (or decompression) step, the decoder also has access to the corresponding sequence $y := (y_1, \ldots, y_n)$. Since $X$ and $Y$ are correlated, the knowledge of $y$ gives information about the sequence $x$, and thus assists in the decoding. Slepian and Wolf showed that as long as the sequences $x$ are compressed into $nR$ bits with $R \geq H(X|Y)$, where $H(X|Y)$ is the conditional entropy of $X$ given $Y$; this task can be accomplished with vanishing probability of error in the asymptotic limit [3]. Here, $R$ is called the rate of the protocol and $n$ is called the coding blocklength. In fact, Slepian and Wolf also considered the case in which $Y$ is compressed and sent to the decoder at a rate $R'$. 

See https://www.ieee.org/publications/rights/index.html for more information.
the decoder attempts to faithfully decode both X and Y. This gives rise to an achievable rate region of pairs \((R, R')\) for which this task is possible. In this work, we do not bound the rate \(R'\); that is, we consider the decoder to receive an uncompressed version of Y.

In this setting, Slepian and Wolf showed that the data compression limit, that is, the minimal rate of asymptotically lossless compression, is given by \(H(X|Y)\). Moreover, Ohama and Han [4] established that the data compression limit for Slepian–Wolf coding satisfies the so-called strong converse property. That is, for any attempted compression to \(nR\) bits with \(R < H(X|Y)\), the probability of error converges to 1 in the asymptotic limit. This protocol has been extended to countably infinite alphabets and a class of information sources with memory by Cover [5], and to various other settings [6], [7].

A quantum generalization of the Slepian–Wolf protocol, which was first introduced by Devetak and Winter [8] is the task of classical data compression with quantum side information. They referred to this task as the Classical-Quantum Slepian–Wolf (CQSW) problem. In this protocol, the correlated pair of random variables \((X, Y)\) is replaced by a classical-quantum (c-q) state \(\rho_{XB}\). Here \(B\) denotes a quantum system which is in the possession of the decoder (say, Bob) and constitutes the quantum side information (QSI), while X is a classical system in the possession of the encoder (say, Alice) and corresponds to a random variable X with probability mass function \(p(x)\), with \(x \in X\), as in the classical setting. Such a c-q state is described by an ensemble \(\{(p(x), \rho_{XB}^x) \}_{x \in X}\) with probability \(p(x)\) the random variable X takes the value x and Bob’s system B is in the state \(\rho_{XB}^x\). In the so-called asymptotic, memoryless setting of CQSW, one considers Alice and Bob to share a large number, \(n\), of identical copies of the c-q state \(\rho_{XB}\). Consequently, Alice knows the sequence \(x := (x_1, x_2, \ldots, x_n)\), whereas the quantum state (i.e. the QSI) \(\rho_{B}^x := \rho_{B}^{x_1} \otimes \rho_{B}^{x_2} \otimes \cdots \otimes \rho_{B}^{x_n}\) is accessible only to Bob. However, Bob has no knowledge of the sequence x. The aim is for Alice to convey the sequence x to Bob using as few bits as possible. Bob can make use of the QSI in order to help him decode the compressed message sent by Alice. Devetak and Winter proved that the data compression limit of CQSW, evaluated in the asymptotic limit \((n \to \infty)\), is given by the conditional entropy \(H(X|B)\) of the c-q state \(\rho_{XB}\).

In this paper we primarily study the CQSW protocol in the non-asymptotic setting in which one no longer takes the limit \(n \to \infty\). This corresponds to the more realistic scenario in which only a finite number of copies of the c-q state \(\rho_{XB}\) are available. First, we focus on the so-called large deviation regime\(^2\), in which the compression rate \(R\) is fixed, and we analyze the optimal probability of error as a function of blocklength \(n\). Specifically, we consider the so-called reliability function or error exponent function \(\varepsilon\) such a state arises when the messages emitted from a quantum DMS are subject to a quantum instrument, and the classical outputs are sent to Alice, and the quantum ones to Bob.

\(^2\)That is, the regime in which the compression rate deviates from the data compression limit by a constant amount. We refer the readers to [9]–[11] for details of different deviation regimes.

(see [12], [13] and references therein), which gives the exponential rate of decay of the minimal probability of error achievable by a Slepian–Wolf protocol, at a fixed rate of compression. In the range \(R > H(X|B)\), we obtain upper and lower bounds on the error exponent function (see Theorems 1 and 2). The lower bound shows that for any \(R > H(X|B)\), the CQSW task can be accomplished with a probability of error which decays to zero exponentially in \(n\). The upper bound puts a limit on how quickly the probability of error can decay.

We term this upper bound the “sphere-packing bound” for CQSW, since it is analogous to the so-called sphere-packing bound obtained in c-q channel coding [14]–[18]. This error exponent analysis contrasts with the so-called small deviations regime, in which one evaluates the minimum compression rate as a function of the coding blocklength, under the constraint that the error probability is below a certain threshold (see [19]–[21] for the classical case, and [22] for the c-q case).

For any protocol with a rate \(R < H(X|B)\), the probability of error converges to one asymptotically and its speed of convergence is given by the strong converse exponent function. We obtain finite blocklength bounds on this function (see Theorems 3 and 4), and determine exactly its asymptotic value (see Corollary 5), in terms of the sandwiched conditional Rényi entropy [23]–[25]. A non-asymptotic study of CQSW in the strong converse domain was also carried out by Tomamichel [26], and by Leditzky et al. [27]. In these works, one-sided bounds were obtained, and hence the asymptotic value of the strong converse exponent was not determined.

The bounds we obtain are expressed in terms of certain entropic exponent functions involving conditional Rényi entropies. To derive these results, we prove and employ properties of these functions. In obtaining the strong converse bounds, we employ variational representations for certain auxiliary exponent functions by making use of those for the so-called log-Euclidean Rényi relative entropies developed in Ref. [28]. Our variational representations are analogous to those obtained by Csiszár and Körner in the classical setting [13], [29]–[31].

We also study the trade-offs between the rate of compression, the minimal probability of error, and the blocklength \(n\). Specifically, we characterize the behaviors of the error probability and the compression rate in the moderate deviation regime [9]–[11], [32]. In contrast to the previously discussed results for which the rate \(R\) was considered to be fixed, here we allow the rate to change with \(n\), approaching \(H(X|B)\), slowly (slower than \(\frac{1}{\sqrt{n}}\)), from above. In this case, we show that the probability of error vanishes asymptotically. In addition, we obtain an asymptotic formula describing the minimum compression rate which converges to \(H(X|B)\) when the probability of error decays sub-exponentially in \(n\). We summarize the error behaviors of different regimes in Table I.

A. Prior Works

Renes and Renner [33] analyzed the protocol in the so-called one-shot setting (which corresponds to the case \(n = 1\)) for a given threshold \(\varepsilon\) on the probability of error. They proved that in this case the classical random
variable $X$ can be compressed to a number of bits given by a different entropic quantity, the so-called smoothed conditional max-entropy, the smoothing parameter being dependent on $\varepsilon$. They also established that this entropic quantity gives the minimal number of bits, up to small additive quantities involving $\varepsilon$. More precisely, the authors established upper and lower bounds on the minimal number of bits in terms of the smoothed conditional max-entropy. The asymptotic result of Devetak and Winter could be recovered from their results by replacing the c-q state $\rho_{XB}$ by its $n$-fold tensor power $\otimes^n\rho_{XB}$ in these one-shot bounds, dividing by $n$, and taking the limit $n \to \infty$. In [22], the authors improved these bounds and established a second order expansion of the so-called minimum code size\(^5\) given an error $\varepsilon$:  

\[
m^* (n, \varepsilon) = nH(X|B)_\rho - \sqrt{nV(X|B)_\rho} \Phi^{-1}(\varepsilon) + O(\log n),
\]

where $V(X|B)_\rho$ is the quantum conditional information variance, and $\Phi$ is the cumulative distribution function of a standard normal distribution.

As regards the strong converse regime, Tomamichel analyzed the maximal success probability (as opposed to the average success probability considered in this work), and derived a lower bound on the strong converse exponent in terms of a min-conditional entropy in Section 8.1.3 of [26]. Later, Leditzky, Wilde, and Datta determined two lower bounds on the strongly converse exponent, which are given in Theorem 6.2 of [27]. The first one bounds the strong converse exponent in terms of a difference of (Petz-) Rényi relative entropies, whereas the second one give a bound in terms of the conditional sandwiched Rényi relative entropy. In fact, the second bound can be written as  

\[
s(n, R) \geq \sup_{\alpha>1} \frac{1}{2\alpha} \left( R - H^*_{\alpha^{-1}}(X|B)_{\rho} \right)
\]

\(^5\)That is, the minimum number of bits required to accomplish the task with blocklength $n$, subject to the constraint that the probability error is at most $\varepsilon$.

where $sc(n, R)$ is the strong converse exponent for blocklength $n$ and rate $R$ (defined in (5)), and $H^*_{\alpha^{-1}}(X|B)_{\rho}$ is the conditional sandwiched $\alpha$-Rényi relative entropy of the source state $\rho_{XB}$ (defined in (15)). This bound is weaker by a factor of $\frac{1}{\alpha}$ to that given in Theorem 3. We note that the first bound is also weaker than Theorem 3. We refer readers to Remark VI.1 for a discussion.

This paper is organized as follows. We introduce the CQSW protocol in Sec. II, and state our main results in Sec. III. The notation and definitions for the entropic quantities and exponent functions are described in Sec. IV. Sec. V presents the error exponent analysis for CQSW as $R > H(X|B)_{\rho}$ (large deviation regime), and we study the optimal success exponent as $R < H(X|B)_{\rho}$ (strong converse regime) in Sec. VI. In Sec. VII we discuss the moderate deviation regime. We conclude this paper in Sec. VIII with a discussion.

\section{Classical Data Compression With Quantum Side Information (Slepian–Wolf Coding)}

Suppose Alice and Bob share multiple (say $n$) identical copies of a classical-quantum (c-q) state  

\[
\rho_{XB} = \sum_{x \in \mathcal{X}} p(x) |x\rangle\langle x| \otimes \rho_B^n
\]

where $\mathcal{X}$ is a finite alphabet and $\rho_B^n$ is a quantum state, of a system $B$ with Hilbert space $\mathcal{H}_B$, in Bob’s possession. The letters $x \in \mathcal{X}$ can be considered to be the values taken by a random variable $X$ with probability mass function $p(x)$. One can associate with $X$ a quantum system (which we also refer to as $X$) whose Hilbert space has an orthonormal basis labeled by $x \in \mathcal{X}$, i.e. $\{|x\rangle\}_{x \in \mathcal{X}}$.

The aim of classical-quantum Slepian–Wolf (CQSW) coding is for Alice to convey sequences $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$ to Bob using as few bits as possible; Bob can employ the corresponding quantum state $\rho_{B^n} = \rho_B^x \otimes \cdots \otimes \rho_B^n$ which is in his possession, and plays the role of quantum side information (PSI), to help decode Alice’s compressed message.
Alice’s encoding (compression) map is given by $E : \mathcal{X}^n \rightarrow \mathcal{W}$, where the alphabet $\mathcal{W}$ is such that $|\mathcal{W}| < |\mathcal{X}|^n$. If Alice’s message was $x$, the compressed message that Bob receives is $E(x) \in \mathcal{W}$. He applies a decoding map $D$ on the pair $(E(x), \rho_{B^n})$ in order to infer Alice’s original message. Thus, Bob’s decoding is given by a map $D : \mathcal{W} \times \mathcal{S}(B^n) \rightarrow \mathcal{X}^n$ where $\mathcal{S}(B^n)$ denotes the set of states on $\mathcal{H}_B^n$. We depict the protocol in Figure 1.

Given $n \in \mathbb{N}$ and $R > 0$, an encoding-decoding pair $(E, D)$ of the form described above is said to form an $(n, R)$-code for the rate $R$. Further, for $R < H(X|B)_\rho$, we obtain bounds on the strong converse exponent $sc(n, R) := \liminf_{n \to \infty} P_e(n, R)$. Hence, $H(X|B)_\rho$ is called the Slepian–Wolf limit. We may illustrate this result by Figure 2 below.

### III. Main Results

The main contributions of this work consist of a refinement of (6). We derive bounds on the speed of convergence of $P_e(n, R)$ to zero for any $R > H(X|B)_\rho$. Further, for $R < H(X|B)_\rho$ we obtain bounds on the strong converse $sc(n, R)$, and determine its exact value in the asymptotic limit. In addition, we analyze the asymptotic behavior of $P_e(n, R)$ in the so-called moderate deviations regime. These results are given by the following theorems, in each of which $\rho_{XB}$ denotes a c-q state (eq. 1), with $X_B$ a variable.

In terms of $P_e^*(n, R)$, Devetak and Winter’s result can be reformulated as

$$
\forall R > H(X|B)_\rho : \limsup_{n \to \infty} P_e^*(n, R) = 0,
$$

$$
\forall R < H(X|B)_\rho : \liminf_{n \to \infty} P_e^*(n, R) > 0.
$$

Thus, the error probability for a random encoding is an average of error probabilities of deterministic encodings. In particular, $\min_{E} P_e((\mathcal{E}, D)) \leq P_e((\mathcal{E}, D))$, so the optimal error probability is achieved for a deterministic code.

The optimal (minimal) rate of data compression evaluated in the asymptotic limit ($n \to \infty$), under the condition that the probability of error vanishes in this limit is called the *data compression limit*. Devetak and Winter [8] proved that it is given by the conditional entropy of $\rho_{XB}$:

$$
H(X|B)_\rho = H(\rho_{XB}) - H(\rho_B)
$$

where $H(\omega) := -\text{Tr}(\omega \log \omega)$ denotes the von Neumann entropy of a state $\omega$.

In this paper, we analyze the Slepian–Wolf protocol primarily in the non-asymptotic scenario ($\text{finite } n$). The two key quantities that we focus on are the following. The *optimal error probability* for a rate $R$ and blocklength $n$ is defined as

$$
P_e^*(n, R) := \inf_{C} \{P_e(C) : C \text{ is an } (n, R)\text{-code for } \rho_{XB}\}.
$$

Similarly, for any $\varepsilon \in (0, 1)$, we define the *optimal rate of compression* at an error threshold $\varepsilon$ and blocklength $n$ by

$$
R^*(n, \varepsilon) := \inf \{R : \exists \text{ an } (n, R)\text{-code } C \text{ with } P_e(C) \leq \varepsilon\}.
$$

In particular, we obtain bounds on the *finite blocklength error exponent*

$$
e(n, R) := \frac{1}{n} \log P_e^*(n, R)
$$

and the *finite blocklength strong converse exponent*

$$
s(n, R) := \frac{1}{n} \log(1 - P_e^*(n, R)).
$$

Given a rate $R \geq H(X|B)_\rho$, the Slepian–Wolf protocol pri-

These results are given by the following theorems, in each of

$$
\forall R > H(X|B)_\rho : \limsup_{n \to \infty} P_e^*(n, R) = 0,
$$

$$
\forall R < H(X|B)_\rho : \liminf_{n \to \infty} P_e^*(n, R) > 0.
$$

Hence, $H(X|B)_\rho$ is called the Slepian–Wolf limit. We may illustrate this result by Figure 2 below.
where

\[ E^i_{\alpha}(R) \equiv E^i_{\alpha}(\rho_{XB}, R) := \sup_{\frac{1}{2} \leq \alpha \leq 1} \frac{1-\alpha}{\alpha} \left( R - H^i_{\alpha}(X|B)_{\rho} \right), \]

\[ H^i_{\alpha}(X|B)_{\rho} := -D_{\alpha}(\rho_{XB}||I_X \otimes \rho_B), \]
and \( D_{\alpha} \) is the \( \alpha \)-Rényi divergence: \( D_{\alpha}(\rho||\sigma) = \frac{1}{\alpha-1} \log \Tr \left[ \rho^\alpha \sigma^{1-\alpha} \right] \).

The proof of Theorem 1 is in Section V-A.

**Theorem 2:** (Sphere-Packing Bound for Slepian–Wolf Coding): Let \( R \in (H(X|B)_{\rho}, H^0_{\alpha}(X|B)_{\rho}) \). Then, there exists \( \alpha_0, K \in \mathbb{N} \), such that for all \( n \geq \alpha_0 \), the finite blocklength error exponent defined in (4) satisfies

\[ c(n, R) \leq E_{\text{op}}(R) + \frac{1}{2} \left( 1 + \left| \frac{\partial E_{\text{op}}(r)}{\partial r} \right|_{r=R} \right) \frac{\log n + K}{n}, \]

where

\[ E_{\text{op}}(R) \equiv E_{\text{op}}(\rho_{XB}, R) := \sup_{0 \leq \alpha \leq 1} \frac{1-\alpha}{\alpha} (R - H^0_{\alpha}(X|B)_{\rho}), \]

and \( H^0_{\alpha}(X|B)_{\rho} := \max_{\sigma \in S(B)} -D_{\alpha}(\rho_{XB}||I_X \otimes \sigma_B) \).

The proof of Theorem 2 is in Section V-B.

On the other hand, for \( R < H(X|B)_{\rho} \), no sequence of codes \( C_n \) can achieve vanishing error asymptotically. For this range, we in fact show that the probability of error converges exponentially quickly to one, as shown by the bounds on \( sc(n, R) \) given in the following theorems.

**Theorem 3:** For all \( R < H(X|B)_{\rho} \), the finite blocklength strong converse exponent defined in (5) satisfies

\[ sc(n, R) \geq E^*_{\text{sc}}(R) > 0, \]

where

\[ E^*_{\text{sc}}(R) \equiv E^*_{\text{sc}}(\rho_{XB}, R) := \sup_{\alpha > 1} \frac{1-\alpha}{\alpha} (R - H^{*,1}_{\alpha}(X|B)_{\rho}), \]

\[ H^{*,1}_{\alpha}(X|B)_{\rho} := \max_{\sigma \in S(B)} -D_{\alpha}^*(\rho_{XB}||I_X \otimes \sigma_B), \]
with \( D_{\alpha}^*(\rho||\sigma) := \frac{1}{\alpha} \log \Tr \left[ \left( \rho^{\frac{\alpha}{2}} \sigma^{1-\alpha} \rho^{\frac{1-\alpha}{2}} \right)^\alpha \right] \) being the sandwiched Rényi divergence [23], [24].

The proof of Theorem 3 is in Section VI-A.

We also obtain an upper bound on \( sc(n, R) \), which, together with Theorem 3 shows that \( E^*_{\text{sc}}(R) \) is the strong converse exponent in the asymptotic limit.

**Theorem 4:** For all \( R < H(X|B)_{\rho} \), the finite blocklength strong converse exponent defined in (5) satisfies

\[ sc(n, R) \leq E^*_{\text{sc}}(R) + \frac{c}{m} \log(m + 1) + O_m \left( \frac{1}{\sqrt{m}} \right) \]
for \( c := \frac{3(\mathcal{H}^2_B + 2(\mathcal{H}^2_B - 1))}{2} \) and any \( m \in \mathbb{N} \), where we denote by \( O_m \left( \frac{1}{\sqrt{m}} \right) \) any term which is bounded by \( C_m \left( \frac{1}{\sqrt{m}} \right) \) for all \( n \) large enough, for some constant \( C_m \) depending only on \( m \).

|\( \mathcal{X} \)|, \( \rho_{XB} \). In particular, taking \( n \to \infty \) then \( m \to \infty \) yields

\[ \lim \sup_{n \to \infty} sc(n, R) \leq E^*_{\text{sc}}(R). \]

The proof of Theorem 4 is in Section VI-B, along with Proposition VI-1, a more detailed version of the result with the constants written explicitly. Note that, together, Theorems 3 and 4 imply the following result.

**Corollary 5:** For all \( R < H(X|B)_{\rho} \), the strong converse exponent defined in (5) satisfies

\[ \lim_{n \to \infty} sc(n, R) = E^*_{\text{sc}}(R). \]

Lastly, we consider the case where the rate depends on \( n \) as \( R_n := H(X|B)_{\rho} + a_n \), where \( a_n \) is a moderate sequence, that is, a sequence of real numbers satisfying

\[ \begin{align*}
(i) & \quad a_n \to 0, \quad \text{as} \quad n \to \infty, \\
(ii) & \quad a_n \sqrt{n} \to \infty, \quad \text{as} \quad n \to \infty. \end{align*} \]

In this case, we have the following asymptotic result.

**Theorem 6:** Assume that the c-q state \( \rho_{XB} \) has strictly positive conditional information variance \( V(X|B)_{\rho} \), where

\[ V(X|B)_{\rho} := V(\rho_{XB}||I_X \otimes \rho_B) \]
with \( V(\rho||\sigma) := \Tr[\rho (\log \rho - \log \sigma^2) - D(\rho||\sigma)^2] \). Then for any sequence \( (a_n)_{n \in \mathbb{N}} \) satisfying Eq. (8),

\[ \lim_{n \to \infty} \frac{1}{na^2_n} \log P^*_{\text{e}}(n, R_n) = -\frac{1}{2V(X|B)_{\rho}} \]

for \( R_n := H(X|B)_{\rho} + a_n \).

**Theorem 7:** Assume that the c-q state \( \rho_{XB} \) has \( V(X|B)_{\rho} > 0 \). Then for any sequence \( (a_n)_{n \in \mathbb{N}} \) satisfying Eq. (8), and \( \varepsilon_n := e^{-na_n} \), we have the asymptotic expansion

\[ R^*(n, \varepsilon_n) = H(X|B)_{\rho} + \sqrt{2V(X|B)_{\rho}a_n} + o(a_n). \]

The proof of Theorem 7 is in Section VII-B.

**IV. Preliminaries and Notation**

Throughout this paper, we consider a finite-dimensional Hilbert space \( \mathcal{H} \). The set of density operators (i.e. positive semi-definite operators with unit trace) on \( \mathcal{H} \) is defined as \( S(\mathcal{H}) \). The quantum systems, denoted by capital letter (e.g. \( A, B \)), are modeled by finite-dimensional Hilbert spaces (e.g. \( \mathcal{H}_A, \mathcal{H}_B \); \( n \) copies of a system \( A \) is denoted by \( A^n \), and is modeled by the \( n \)-fold tensor product of the Hilbert spaces, \( \mathcal{H}^n = \mathcal{H}_A^n \). For \( \rho, \sigma \in S(\mathcal{H}) \), we denote by \( \rho \ll \sigma \) if the support of \( \rho \) is contained in the support of \( \sigma \). The identity operator on \( \mathcal{H} \) is denoted by \( \mathbb{I}_\mathcal{H} \). The subscript will
be removed if no confusion is possible. We use \( \text{Tr}[\cdot] \) as the standard trace function. For a bipartite state \( \rho_{AB} \in S(AB) \), \( \text{Tr}_B[\rho_{AB}] \) denotes the partial trace with respect to system \( B \). We denote by \( |t|^+ := \max\{0, t\} \). The indicator function \( 1_{\{A\}} \) is defined as follows: \( 1_{\{A\}} = 1 \) if the event \( A \) is true; otherwise \( 1_{\{A\}} = 0 \).

For a positive semi-definite operator \( X \) whose spectral decomposition is \( X = \sum_i a_i P_i \), where \( (a_i) \) and \( (P_i) \) are the eigenvalues and eigoprojections of \( X \), its power is defined as: \( X^p := \sum_{i:a_i\neq 0} a_i^p P_i \). In particular, \( X^0 \) denotes the projection onto \( \text{supp}(X) \), where we use \( \text{supp}(A) \) to denote the support of the operator \( A \). Further, \( A \perp B \) means \( \text{supp}(A) \cap \text{supp}(B) = \emptyset \). Additionally, we define the pinching map with respect to \( X \) by \( P_X(A) = \sum_i P_i A P_i \). The \( \exp \) and \( \log \) are performed on base 2 throughout this paper.

### A. Entropic Quantities

For any pair of density operators \( \rho \) and \( \sigma \), we define the quantum relative entropy, Petz’s quantum Rényi divergence [34], sandwiched Rényi divergence [23], [24], and the log-Euclidean Rényi divergence [28], [35], respectively, as:

\[
\begin{align*}
D(\rho||\sigma) := & \text{Tr}[\rho (\log \rho - \log \sigma)]; \\
D_\alpha(\rho||\sigma) := & \frac{1}{\alpha - 1} \log Q_\alpha(\rho||\sigma), \\
Q_\alpha(\rho||\sigma) := & \text{Tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]; \\
D_\alpha^*(\rho||\sigma) := & \frac{1}{\alpha - 1} \log Q_\alpha^*(\rho||\sigma), \\
Q_\alpha^*(\rho||\sigma) := & \text{Tr}\left[\left(\rho^{\frac{1}{\alpha}} \sigma^{\frac{1}{1-\alpha}}\right)^\alpha\right].
\end{align*}
\]

Replacing \( \sigma \) by \( \sigma + \varepsilon + I \) and then taking \( \varepsilon \to 0 \), one can define the above four quantities without the support condition.

We define the quantum relative entropy variance [22], [36] by

\[
V(\rho||\sigma) := \text{Tr}[\rho (\log \rho - \log \sigma)^2] - D(\rho||\sigma)^2.
\]

The above quantity is non-negative. Further, it follows that \( V(\rho||\sigma) > 0 \) implies \( D(\rho||\sigma) > 0 \).

For \( \rho_{AB} \in S(AB) \), \( \alpha \geq 0 \) and \( t = \{0\}, \{1\}, \) or \( \{\ast\} \), the quantum conditional Rényi entropies are given by

\[
H_\alpha^R(A|B)_\rho := \max_{\sigma_B \in S(B)} -D_\alpha(\rho_{AB}||I_A \otimes \sigma_B),
\]

\[
H_\alpha^L(A|B)_\rho := -D_\alpha^*(\rho_{AB}||I_A \otimes \rho_B).
\]

When \( \alpha = 1 \) and \( t = \{0\}, \{1\}, \) or \( \{\ast\} \) in Eq. (15), both quantities coincide with the usual quantum conditional entropy:

\[
H_1^R(A|B)_\rho = H_1^L(A|B)_\rho = H(A|B)_\rho := H(AB)_\rho - H(B)_\rho,
\]

where \( H(A)_\rho := -\text{Tr}[\rho_A \log \rho_A] \) denotes the von Neumann entropy. Similarly, for \( \rho_{AB} \in S(AB) \), we define the conditional information variance:

\[
V(A|B)_\rho := V(\rho_{AB}||I_A \otimes \rho_B).
\]

It is not hard to verify from Eq. (14) that \( V(A|B)_\rho > 0 \) implies \( H(A|B)_\rho > 0 \).

**Lemma IV.1** ([37], [28, Lemma 3.3, Lemma 3.12, Theorem 3.16, Corollary 3.27], [38, Corollary 2.2]): Let \( \rho, \sigma \in S(H) \). Then,

\[
\alpha\rightarrow \log Q_\alpha(\rho||\sigma) \quad \text{and} \quad \alpha\rightarrow \log Q_\alpha^*(\rho||\sigma)
\]

are convex on \( [0, 1] \); \( \alpha\rightarrow D_\alpha(\rho||\sigma) \) is continuously monotone increasing on \( [0, 1] \).

Moreover, \( \forall \alpha \in (0, 1), (\rho, \sigma) \mapsto Q_\alpha(\rho||\sigma) \) is jointly concave \n\( \forall \alpha \in [0, 1], \sigma \mapsto D_\alpha(\rho||\sigma) \) is convex and lower semi-continuous.

**Proposition IV.2** (Properties of \( \alpha \)-Rényi Conditional Entropy): Given any classical-quantum state \( \rho_{XB} \in S(XB) \), the following holds:

(a) The map \( \alpha \mapsto H_1^R(X|B)_\rho \) is continuous and monotonically decreasing on \( [0, 1] \).

(b) The map \( \alpha \mapsto \frac{1-a}{a} H_1^L(X|B)_\rho \) is concave on \( (0, 1) \).

The proof is provided in Appendix .

Given two states \( \rho \) and \( \sigma \), one can define an associated binary hypothesis testing problem of determining which of the two states was given via a binary POVM. Such a POVM is described by an operator \( Q \) (associated, say, with the outcome \( \rho \)) that such \( 0 \leq Q \leq 1 \), called the test. Two types of errors are possible; the probability of measuring \( \rho \) and reporting the outcome \( \sigma \) is given by \( \text{Tr}[Q|\sigma] \) and called the type-I error, while the probability of measuring \( \sigma \) and reporting the outcome \( \rho \) is given by \( \text{Tr}[Q|\rho] \) and is called the type-II error. The hypothesis testing relative entropy (e.g. as defined in [39]) is defined by

\[
D_H^R(\rho||\sigma) = -\log \inf_{Q^0 \leq Q \leq 1} \text{Tr}[Q|\sigma] \quad \text{and} \quad D_H^L(\rho||\sigma) = -\log \inf_{Q^0 \leq Q \leq 1} \text{Tr}[Q|\rho]
\]

and characterizes the minimum type-II error incurred via a test which has type-I error at most \( \varepsilon \). The hypothesis testing relative entropy satisfies the data-processing inequality

\[
D_H^R(\rho||\sigma) \geq D_H^R(\Phi(\rho)||\Phi(\sigma)), \quad \varepsilon \in (0, 1)
\]

for any completely positive and trace-preserving (CPTP) map \( \Phi \) [39]. This quantity has an interpretation as a relative entropy as it satisfies the following asymptotic equipartition property:

\[
\lim_{n \to \infty} \frac{1}{n} D_H^R(\rho^\otimes n||\sigma^\otimes n) = D(\rho||\sigma)
\]

which was proven in two steps, by [40] and [35].

\footnote{It was shown in [28, Corollary 3.27] that the map \( \sigma \mapsto D_\alpha(\rho||\sigma) \) is lower semi-continuous on \( S(H) \) for all \( \alpha \in (0, 1) \). The argument can be extended to the range \( \alpha \in [0, 1] \) by the same method in [28, Lemma Corollary 3.27].}
We can consider a related quantity, $\tilde{a}_\mu(\|\cdot\|)$ which denotes the minimum type-I error such that the type-II error does not exceed $\mu$. That is,
\[
\tilde{a}_\mu(\|\cdot\|) = \inf_{T:0\leq\|T\|\leq1} \text{Tr}[(1-T)\rho] = \exp(-D^\mu_H(\sigma||\rho)). \tag{21}
\]
By (20), for any CPTP $\Phi$ we have
\[
\tilde{a}_\mu(\|\cdot\|) = \exp(-D^\mu_H(\sigma||\rho)) \leq \exp(-D^\mu_H(\Phi(\sigma)||\Phi(\rho))). \tag{22}
\]
We also consider the so-called max-relative entropy, given by
\[
D_{\text{max}}(\rho||\sigma) := \inf\{\gamma: \rho \leq 2^\gamma\sigma\}. \tag{23}
\]
In establishing the exact strong converse exponent, we will employ a smoothed variant of this quantity,
\[
D^\delta_{\text{max}}(\rho||\sigma) := \min_{\rho \in B_\delta(\rho)} D_{\text{max}}(\rho||\sigma), \tag{24}
\]
where
\[
B_\delta(\rho) = \{\bar{\rho} \in D(H) : d_{\text{op}}(\bar{\rho}, \rho) \leq \delta\}
\]
is the $\varepsilon$-ball in the distance of optimal purifications$^5$ [41], $d_{\text{op}}$, defined by
\[
d_{\text{op}}(\rho, \sigma) = \min_{\psi_\rho, \psi_\sigma} \frac{1}{2} \| \psi_\rho \langle \psi_\sigma | - | \psi_\sigma \rangle \psi_\rho \|_1 \tag{25}
\]
where the minimum is over purifications $\psi_\rho$ of $\rho$ and $\psi_\sigma$ of $\sigma$. By equation (2) of [42], the distance $d_{\text{op}}$ satisfies
\[
\frac{1}{2} d_{\text{op}}(\rho, \sigma)^2 \leq \frac{1}{2} \| \rho - \sigma \|_1 \leq d_{\text{op}}(\rho, \sigma). \tag{26}
\]
It was shown in [42] that $D^\delta_{\text{max}}(\rho||\sigma)$ satisfies an asymptotic equipartition property. In fact, Theorem 14 of [42] gives finite $n$ upper and lower bounds on $\frac{1}{n} D^\delta_{\text{max}}(\rho^{\otimes n}||\sigma^{\otimes n})$ which converge to $D(\rho||\sigma)$. We will only need the upper bound, namely
\[
\frac{1}{n} D^\delta_{\text{max}}(\rho^{\otimes n}||\sigma^{\otimes n}) \leq D(\rho||\sigma)
\]
\[
+ \frac{4\sqrt{2}(\log n)\log 1 - \sqrt{1 - \delta^2}}{n}, \tag{27}
\]
where $n := 1 + \text{Tr}[\rho^{3/2} \sigma^{-1/2} + \rho^{3/2} \sigma^{-1/2}]$.

The smoothed max-relative entropy satisfies the following simple but useful relation (which is a one-shot analog of Lemma 5.8 of [28]).

**Lemma IV.3:** For $\delta \geq 0$, if $a \geq D^\delta_{\text{max}}(\rho||\sigma)$ then
\[
\text{Tr}[(\sigma - e^{a} \rho)_{+}] \leq 2\delta.
\]

**Proof of Lemma IV.3:** By the definition (23), for any $a \geq D_{\text{max}}(\rho||\sigma)$, we have $\sigma - e^{a} \rho \leq 0$, and therefore $\text{Tr}[(\sigma - e^{a} \rho)_{+}] = 0$, which proves the result for $\delta = 0$. For $\delta > 0$ and $a \geq D^\delta_{\text{max}}(\rho||\sigma)$ there exists a density matrix $\tilde{\sigma}$ with $\sigma - e^{a} \rho \leq 0$ and $d_{\text{op}}(\tilde{\sigma}, \sigma) \leq \delta$. Setting $P_+ = \{\sigma \leq e^{a} \rho\}$, we have
\[
\text{Tr}[P_+(\sigma - e^{a} \rho)] = \text{Tr}[P_+(\sigma - \tilde{\sigma})] + \text{Tr}[P_+(\tilde{\sigma} - e^{a} \rho)].
\]

Since $0 \leq a \leq 1$, we have $\text{Tr}[P_+(\sigma - e^{a} \rho)] \leq \text{Tr}[(\sigma - e^{a} \rho)_{+}] \leq 0$. Thus,
\[
\text{Tr}[P_+(\sigma - e^{a} \rho)] \leq \text{Tr}[P_+(\sigma - \tilde{\sigma})] \leq \|\sigma - \tilde{\sigma}\|_1 \leq 2d_{\text{op}}(\sigma, \tilde{\sigma}) \leq 2\delta
\]
using (26) in the second to last inequality.

**B. Error Exponent Function**

For $t \in \{0, 1\}$, or $\{\theta\}$, we define
\[
E^t_\theta(R) := \sup_{s \geq 0} \left\{ E^t_\theta(s) + sR \right\}; \tag{28}
\]
\[
E^t_{\text{op}}(R) := \sup_{s \geq 0} \left\{ E^t_{\text{op}}(s) + sR \right\}; \tag{29}
\]
\[
E^t_\infty(R) := \sup_{s \geq 0} \left\{ E^t_\infty(s) + sR \right\}; \tag{30}
\]
\[
E^t_0(R) := -s \cdot H_{\alpha}^{-1}(X|B)_\rho \left| a = \frac{1}{1+X} \right. \tag{31}
\]
\[
E^t_0(s) := -s \cdot H_{\alpha}^{-1}(X|B)_\rho \left| a = \frac{1}{1+X} \right. \tag{32}
\]

It is known that [28, Proposition 3.20]
\[
D^t_\alpha(\cdot||\cdot) \leq D_\alpha(\cdot||\cdot) \leq D^t_\alpha(\cdot||\cdot), \quad \alpha \in [0, 1)
\]
\[
D^t_\alpha(\cdot||\cdot) \leq D^t_{\alpha}(\cdot||\cdot) \leq D_\alpha(\cdot||\cdot), \quad \alpha \in (1, +\infty],
\]
which implies that
\[
E_{\text{op}}(R) \leq E^t_{\text{op}}(R)
\]
\[
E_{\theta}(R) \leq E^t_{\theta}(R)
\]
\[
E_{\infty}(R) \leq E^t_{\infty}(R)
\]

Further, $H_{\alpha}^{-1}(X|B)_\rho \leq H_{\alpha}^{-1}(X|B)_\rho \leq a \in [1/2, +\infty)$ [49, Corollary 4], [25, Eq. (5.65)]. For $R \in [H_{\alpha}^{-1}(X|B)_\rho, H_{\alpha}^{\frac{1}{2}}(X|B)_\rho]$, together with Proposition IV.5-a below, we have
\[
E^t_{\theta}(R) \leq E_{\theta}(R) = E_{\text{op}}(R) \leq E^t_{\text{op}}(R) = E^t_{\theta}(R)
\]

In Section V-A, we obtain an achievable bound of the optimal error in terms of $E^t_{\theta}$. We conjecture that it can be further improved to $E_{\theta}$.

In the following, we collect some useful properties of the auxiliary functions $E_0(s)$ and $E_0^t(s)$.
Proposition IV.4 (Properties of Auxiliary Functions): Let \( \rho_{XB} \) be a classical-quantum state with \( H(X|B)_\rho > 0 \), the auxiliary functions \( E_0(s), E_0^1(s), \) and \( E_0^2(s) \) admit the following properties.

(a) (Continuity) The function \( s \mapsto E_0(s) \) is continuous for all \( s \in (-1, +\infty) \); the function \( s \mapsto E_0^1(s) \) is continuous for all \( s \in [0, +\infty) \); and the function \( s \mapsto E_0^2(s) \) is continuous for all \( s \in (-1, 1] \).

(b) (Negativity) For \( s \geq 0 \),
\[
E_0(s) \leq 0; \quad E_0^1(s) \leq 0; \quad \text{and} \quad E_0^2(s) \leq 0
\]
with \( E_0(0) = E_0^1(0) = E_0^2(s) = 0 \).

(c) (Concavity) All the functions \( s \mapsto E_0(s), s \mapsto E_0^1(s), \) and \( s \mapsto E_0^2(s) \) are concave in \( s \) for all \( s \in (-1, +\infty) \).

(d) (First-order Derivative)
\[
\left. \frac{\partial E_0^1(s)}{\partial s} \right|_{s=0} = \left. \frac{\partial E_0(s)}{\partial s} \right|_{s=0} = \left. \frac{\partial E_0^2(s)}{\partial s} \right|_{s=0} = -H(X|B)_\rho.
\]

(e) (Second-order Derivative)
\[
\left. \frac{\partial^2 E_0^1(s)}{\partial s^2} \right|_{s=0} = \left. \frac{\partial^2 E_0(s)}{\partial s^2} \right|_{s=0} = \left. \frac{\partial^2 E_0^2(s)}{\partial s^2} \right|_{s=0} = -V(X|B)_\rho.
\]

The proof is provided in Appendix A.

Proposition IV.5 below discusses the properties of the exponent functions. See Figure 3 for the illustration.

Proposition IV.5 (Properties of the Exponent Function): Let \( \rho_{XB} \) be a classical-quantum state with \( H(X|B)_\rho > 0 \), the following holds.

(a) \( E_{sp}(\cdot) \) is convex, differentiable, and monotonically increasing on \([0, +\infty)\). Further,
\[
E_{sp}(R) = \begin{cases} 0, & R \leq H_1^1(X|B)_\rho \\ E_s(R), & H_1^1(X|B)_\rho \leq R \leq H_1^{1/2}(X|B)_\rho \\ +\infty, & R > H_0^1(X|Y)_\rho \end{cases}
\]

(b) Define
\[
F_R(\alpha, \sigma_B) := \frac{1 - \alpha}{\alpha} (R + D_{\rho_{XB}}(I_X \otimes \sigma_B))
\]
on \((0, 1) \times S(B)\), and \( F_R(1, \sigma_B) := 0 \) for all \( \sigma_B \in S(B) \). For \( R \in (H_1^1(X|B)_\rho, H_0^1(X|B)_\rho) \), there exists a unique saddle-point \((\alpha^*, \sigma^*) \in (0, 1) \times S(B)\) of \( F_R(\cdot, \cdot) \) such that
\[
F_R(\alpha^*, \sigma^*) = \sup_{\alpha \in [0,1]} \inf_{\sigma_B \in S(B)} F_R(\alpha, \sigma_B) = \inf_{\sigma_B \in S(B)} \sup_{\alpha \in [0,1]} F_R(\alpha, \sigma_B) = E_{sp}(R).
\]

(c) Any saddle-point \((\alpha^*, \sigma^*)\) of \( F_R(\cdot, \cdot) \) satisfies \( I_X \otimes \sigma^* \gg \rho_{XB} \).

The proof is provided in Appendix A.

In Proposition IV.6 below, we show that the exponent functions defined in terms of \( D^0 \) admit the variational representations, analogous to those introduced in the classical case by Csiszár and J. Körner’s [13], [29], [31].
A. Achievability

Let us recall Theorem 1.

Theorem 1: For any rate $R \geq H(X|B)_n$, and any blocklength $n \in \mathbb{N}$, the finite blocklength error exponent defined in (4) satisfies

$$e(n, R) \geq E^*_1(R) - \frac{2}{n},$$

where

$$E^*_1(R) = E^*_1(\rho_{XB}, R) := \sup_{\alpha \leq 1} \frac{1-\alpha}{\alpha} \left( R - H^*_2(\rho_{XB})_{\rho} \right),$$

$$H^*_2(\rho_{XB})_{\rho} := -D_{\alpha}(\rho_{XB} \| I_X \otimes \rho_B),$$

and was later studied by Renes and Renner [33].

We first present a one-shot lower bound on $P_\varepsilon$, the finite blocklength error exponent defined in (4) satisfies

$$P_\varepsilon(1, \log |W|) \leq P_e(C) \leq 2 \text{Tr} [\rho_{XB} (I_X - \Lambda_{XB})] + \frac{4}{|W|} \text{Tr} [I_X \otimes \rho_B \Lambda_{XB}].$$

Now, we choose, for $x \in \mathcal{X}$,

$$\Lambda_x := \left\{ p(x) \rho_x^{\varepsilon} - \frac{1}{|W|} \rho_B \geq 0 \right\}.$$

We invoke Audenaert et al.'s inequality [50], [51]: for every $A, B \geq 0$ and $s \in [0, 1]$,

$$\text{Tr} [A - B \geq 0] + (B - A \leq 0) \leq \text{Tr} [A^{1-s} B^s].$$

Letting $A = \rho_{XB}, B = \frac{1}{|W|} I_X \otimes \rho_B$, and noting that $\Lambda_{XB} = \left\{ \rho_{XB} - \frac{1}{|W|} I_X \otimes \rho_B \geq 0 \right\}$, we have one-shot achievability:

$$P_\varepsilon(1, \log |W|) \leq 4 \min_{s \in [0, 1]} |W|^{-s} \text{Tr} \left[ \rho_{XB}^{1-s} (I_X \otimes \rho_B)^{s} \right]. \quad (38)$$

Finally, we consider the $n$-tuple case. Note that $\rho_{X^n B^n} = \rho_{X B}^{\otimes n}$, and let $|W| = \exp(n R)$. Eqs. (38) and (32) lead to

$$P_\varepsilon(n, R) \leq 4 \exp \left\{ -n E^*_1(R) \right\},$$

which completes the proof.

B. Optimality

The main result of this section is the finite blocklength bound for the optimal error probability—Theorem 2. We call this the sphere-packing bound for Slepian–Wolf coding with quantum side information, as a counterpart of the sphere-packing bound in classical-quantum channel coding [14], [16]. The proof technique relies on a one-shot converse bound in Proposition V.1 below (adapting the technique of [39] to the case with side information), and a sharp $n$-shot expansion from [9], [16] of a hypothesis testing quantity.

Proposition V.1 (One-Shot Converse Bound for Error): Consider a Slepian–Wolf coding with a joint classical-quantum state $\rho_{XB} \in \mathcal{S}(XB)$ and the index size $|W| < |X'|$. Then,

$$-\log P_\varepsilon(1, \log |W|) \leq \min_{\sigma_B \in \mathcal{S}(B)} - \log \hat{\alpha}_{\rho_{XB} \parallel \sigma_B} (\bar{\tau}_X),$$

where $\bar{\tau}_X$ denotes the uniform distribution on the input alphabet $X'$; and $\hat{\alpha}_{\rho_{XB} \parallel \sigma_B}$ is defined by (21).

Proof of Proposition V.1: Let $W$ be a set and $C$ be a code with encoder $E : X \rightarrow W$, and decoder $D$. To bound the optimal error probability, we may reduce to the case of deterministic $E$, as discussed below (2). The decoder is a family of POVMs $D = \{ D_w \}_{w \in W}$, where $D_w = \{ \Pi_{x_w} \}_{x \in X}$.

Let $\sigma_B \in \mathcal{S}(B)$. We will consider a two-outcome hypothesis test between $\rho_{XB}$ and $\bar{\tau}_X \otimes \sigma_B$, where $\bar{\tau}_X = \frac{1}{|X'|}$. Let us define the test

$$\Pi_{XB} := \sum_{x \in X} \frac{|x \rangle \langle x|}{|W|} \otimes \Pi_{\bar{x}(x)}. \quad (39)$$

Then $0 \leq \Pi_{XB} \leq 1$ and moreover,

$$\text{Tr} [\rho_{XB} \Pi_{XB}] = \sum_{x \in X} \frac{1}{|X'|} \sum_{w \in W} \sum_{x \in X} \text{Tr} [\Pi_{x.w} \sigma] \leq \frac{1}{|X'|} \sum_{w \in W} \text{Tr} [\sigma_B] = \frac{|W|}{|X'|} \leq 1.$$
with $\rho^n \ll \sigma^n$, and a sequence of positive numbers $(r_n)_{n \in \mathbb{N}}$. Denote by
\[
\phi_n(r_n |\rho^n|\sigma^n) := \sup_{\alpha \in (0,1)} \frac{1 - \alpha}{\alpha} \left( \frac{1}{n} D_\alpha(\rho^n||\sigma^n) - r_n \right), \tag{40}
\]
and
\[
s^n_\alpha := \arg \max_{s \geq 0} \left\{ \frac{s}{n} D_{\frac{\alpha}{\alpha}}(\rho^n||\sigma^n) - sr_n \right\}.
\]

If $\nabla_V(\rho^n||\sigma^n) \geq \nu$ for some $\nu > 0$, then there exist $N_1 \in \mathbb{N}$, $K_1 > 0$ such that for all $n \geq N_1$, we have
\[
-\log \tilde{\alpha}_{\exp(-nr)}(\rho^n||\sigma^n) \leq n \phi_n(r_n - \gamma_n |\rho^n|\sigma^n)
\]
\[
+ \frac{1}{2} \log (ns^n_\alpha) + K_1, \tag{41}
\]
where $\gamma_n := \frac{\log n + C}{n}$. Moreover, if there exists $\varepsilon > 0$ such that for all $r \in (r_n - \varepsilon, r_n]$, $\phi_n(r |\rho^n|\sigma^n) \in [\nu, +\infty)$, \tag{42}

for some $\nu > 0$, then there exist $N_2 \in \mathbb{N}$, $K_2 > 0$ such that for all $n \geq N_2$, we have
\[
-\log \tilde{\alpha}_{\exp(-nr)}(\rho^n||\sigma^n) \leq n \phi_n(r_n |\rho^n|\sigma^n)
\]
\[
+ \frac{1}{2} (1 + s^n_\alpha) \log n + K_2. \tag{43}
\]

With Propositions V.1 and V.2, we are able to show the $n$-shot bound given by Theorem 2, which we recall here:

**Theorem 2 (Sphere-Packing Bound for Slepian–Wolf Coding):** Let $R \in (H(X|B)_\nu, H_0(X|B)_\nu)$. Then, there exist $N_0, K \in \mathbb{N}$, such that for all $n \geq N_0$, the finite blocklength error exponent defined in (4) satisfies
\[
e(n, R) \leq E_{sp}(R) + \frac{1}{2} \left( 1 + \frac{\partial E_{sp}(R)}{\partial r} \right) \log n + K, \tag{44}
\]
where
\[
E_{sp}(R) \equiv E_{sp}(\rho_{X^n,B^n}) := \sup_{0 \leq \alpha \leq 1} \frac{1 - \alpha}{\alpha} \left( R - H_0(X|B)_\alpha \right).
\]

and
\[
H_0(X|B)_\alpha := \max_{p_{\theta B^n}} E_{sp}(\rho_{X^n,B^n}) = D_\alpha(\rho_{X^n,B^n} || I_X \otimes \sigma_B).
\]

**Proof of Theorem 2:** The proof is split into two parts.

We first invoke a one-shot converse bound in Proposition V.1 to relate the optimal error of Slepian–Wolf coding to a binary hypothesis testing problem. Second, we employ a sharp converse Hoeffding bound in Proposition V.2 to asymptotically expand the optimal type-I error, which yields the desired result.

Applying Proposition V.1 with an $n$-shot extension $\rho_{X^n,B^n}$ of the c-q state $\rho_{X^n,B^n}$, $|W| = \exp(nR)$, and $\tau_{X^n} = \frac{1}{|I_X|} I_{X^n}$ gives
\[
\log \left( \frac{1}{P_e(n, R)} \right) \leq \min_{\phi_B \in S(B^n)} -\log \tilde{\alpha}_{\exp(-nr)}(\rho_{X^n,B^n} || \tau_{X^n} \otimes \sigma_B^n)
\]
\[
\leq -\log \tilde{\alpha}_{\exp(-nr)}(\rho_{X^n,B^n} || \tau_{X^n} \otimes (\sigma_B^n) \otimes n),
\]
\[
= -\log \tilde{\alpha}_{\exp(-nr)}(\rho_{X^n,B} || (\tau_{X} \otimes \sigma_B^n)^n), \tag{44}
\]
where we invoke the saddle-point property in Proposition IV.5-b to denote by
\[
\sigma_B^n := \arg \min_{\sigma_B \in S(B)} \sup_{\alpha \in (0,1]} \frac{1 - \alpha}{\alpha} \left( R + D_\alpha(\rho_{X^n,B} || I_X \otimes \sigma_B) \right).
\]

Next, we show that Eq. (42) is satisfied, and thus we can exploit Proposition V.2 to expand the right-hand side of Eq. (44). Let $r = \log |X| - R$, and note that item c in Proposition IV.5 implies
\[
\rho_{X,B} \ll \tau_{X} \otimes \sigma_B^n.
\]

One can verify that
\[
\phi_n(r |\rho^n_{X,B} || (\tau_X \otimes \sigma_R^n)^n) = \sup_{\alpha \in (0,1]} \frac{1 - \alpha}{\alpha} \left( D_\alpha(\rho_{X^n,B} || \tau_{X} \otimes \sigma_R^n) - r \right), \tag{45}
\]
\[
= \sup_{\alpha \in (0,1]} \frac{1 - \alpha}{\alpha} \left( D_\alpha(\rho_{X^n,B} || I_X \otimes \sigma_R^n) - \log |X| - r \right),
\]
\[
E_{sp}(R) = 0, \tag{46}
\]
\[
E_{sp}(R) > 0, \tag{47}
\]
where $\phi_n$ is defined in Eq. (40); equality (46) follows from the saddle-point property, item b in Proposition IV.5, and the definition of $E_{sp}(R)$ in Eq. (29); the last inequality (47) is due to item a in Proposition IV.5 and the given range of $R$. Further, the positivity of $\phi_n(r |\rho^n_{X,B} || (\tau_X \otimes \sigma_R^n)^n)$ implies that $r > D_0(\rho_{X^n,B} || I_X \otimes \sigma_R^n)$. By choosing $\varepsilon = r - D_0(\rho_{X^n,B} || I_X \otimes \sigma_R^n) > 0$, we apply Eq. (43) in Proposition V.2 on Eq. (44) to obtain
\[
\log \left( \frac{1}{P_e(n, R)} \right) \leq n \phi_n(r |\rho^n_{X,B} || (\tau_X \otimes \sigma_R^n)^n)
\]
\[
+ \frac{1}{2} \left( 1 + \frac{\partial \phi_n(r |\rho^n_{X,B} || (\tau_X \otimes \sigma_R^n)^n)}{\partial r} \right) \log n + K, \tag{48}
\]
where $K > 0$ is some finite constant independent of $n$. Finally, combining Eqs. (46) and (48) completes the proof.

VI. OPTIMAL SUCCESS EXPONENT AT A FIXED RATE BELOW THE SLEPIAN–WOLF LIMIT (STRONG CONVERSE REGIME)

In this section, we investigate the case of rate below the Slepian–Wolf limit, i.e. $R < H(X|B)_\nu$, which is analogous to the strong converse in channel coding [52–55]. We establish both the finite blocklength converse and achievability bound in Section VI-A and VI-B. As a result of Theorem 3 and Theorem 4 below, we are able to show that in the strong converse regime, $E_{sc}^*(R)$ characterizes the exponential decay of the probability of success:
\[
\lim_{n \to \infty} sc(n, R) = \lim_{n \to \infty} -\frac{1}{n} \log \left( 1 - P_e^*(n, R) \right) = E_{sc}^*(R),
\]
where $E_{sc}^*(R)$ is defined by (7) below, and $P_e^*(n, R)$ is defined by (3).

**Remark VI.1:** Theorem 3 together with Corollary 5 imply that the established $n$-shot strong converse converse bound is stronger than the results obtained in Ref. [27] whenever $E_{sc}^*(R)$ is finite. To see this, let us call the exponent of
the bounds in [27, Theorem 6.2] by $E_{LWD}(R)$. Namely, $E_{LWD}(R) \leq \text{sc}(n, R)$ for all $n$. Now, we assume for some rate $R > 0$ that $E_{sc}(R) < E_{LWD}(R)$. Then, it holds that $E_{sc}^*(R) < E_{LWD}(R) \leq \text{sc}(n, R)$ for all $n$. Taking $n$ to infinity and invoking Corollary 5, we have

$$E_{sc}^*(R) < E_{LWD}(R) \leq \lim_{n \to \infty} \text{sc}(n, R) = E_{sc}^*(R),$$

which is a contradiction as long as $E_{sc}^*(R)$ is finite. Therefore, we conclude that $E_{LWD}(R) \leq E_{sc}^*(R)$ for any $R$ such that $E_{sc}^*(R)$ is finite.

A. Establishing a Strong Converse Exponent

Let us recall Theorem 3.

Theorem 3: For all $R < H(X|B)_p$, the finite blocklength strong converse exponent defined in (5) satisfies

$$\text{sc}(n, R) \geq E_{sc}^*(R) > 0,$$

where

$$E_{sc}^*(R) \equiv E_{sc}^*(\rho_{XB}, R) := \sup_{\alpha > 1} \frac{1 - \alpha}{\alpha} (R - H_{\alpha}^+(X|B)_p),$$

and $H_{\alpha}^+(X|B)_p := \max_{\sigma_B \in S(B)} -D_{\alpha}^*(\rho_{XB} \| \sigma_B)$, with $D_{\alpha}^*(\rho_1 \| \rho_2) := \frac{1}{\alpha - 1} \log \text{Tr} \left( \rho_1 \rho_2^\alpha \right)$ being the sandwiched Rényi divergence [23], [24].

Remark VI.2: The main ingredient in the proof of Theorem 3 is the monotonicity of the quantum sandwiched Rényi divergence [23], [24]. This approach was first considered by Sheverdaev [56] in the classical case (using Hölder’s inequality), and then by Augustin in his thesis [57]. Later, Nagao [58] adopted that approach for the quantum Petz Rényi divergence. Polyanskiy and Verdú [59] formally wrote down the argument as a monotonicity of any generalised divergence. Now, it is a standard approach employed in the above-mentioned quantum information-theoretic tasks; see e.g. [24], [28].

Proof of Theorem 3: Let $R \geq 0$. We first claim that any one-shot code $C$ with $R < H(X|B)_p$ satisfies

$$-\log (1 - P_e^*(1, R)) \leq E_{sc}^*(R),$$

where $\rho_{XB}$ is a c-q state defined by (1).

Let $C$ be a code with encoder $E : \mathcal{X} \to \mathcal{W}$ (which is deterministic, without loss of generality) satisfying $\log |W| = R$, and decoder $D$. Let also $\sigma_B \in S(B)$ be a state. We will use the test $\Pi_{XB}$, satisfying $0 \leq \Pi_{XB} \leq \mathbb{1}_{XB}$, defined in (39) above in terms of $\mathcal{C}$ and $\sigma_B$, for which it was shown that $P_e(C) = 1 - \text{Tr} [\rho_{XB} \Pi_{XB}]$ and

$$\text{Tr} [(\tau_X \otimes \sigma_B) \Pi_{XB}] \leq \frac{|W|}{|\mathcal{X}|}$$

for $\tau_X = \frac{1}{|\mathcal{X}|}$. Then, for all $\alpha > 1$,

$$(1 - P_e(C))^\alpha \left( \frac{|W|}{|\mathcal{X}|} \right)^{1-\alpha} \leq (\text{Tr} [\rho_{XB} \Pi_{XB}])^\alpha \left( \text{Tr} [\tau_X \otimes \sigma_B \Pi_{XB}] \right)^{1-\alpha}$$

$$\leq (\text{Tr} [\rho_{XB} \Pi_{XB}])^\alpha \left( \text{Tr} [\tau_X \otimes \sigma_B \Pi_{XB}] \right)^{1-\alpha}$$

$$+ (\text{Tr} [\rho_{XB} (I - \Pi_{XB})])^\alpha \left( \text{Tr} [\tau_X \otimes \sigma_B (I - \Pi_{XB})] \right)^{1-\alpha}. $$

(50)

Consider the measure-and-prepare map $\Phi : S(\mathcal{H}_X \otimes \mathcal{H}_B) \to S(\mathcal{C}^2)$ given by

$$\Phi : \eta_{XB} \mapsto \text{Tr} [\eta_{XB} \Pi_{XB}] |0\rangle \langle 0| + \text{Tr} [\eta_{XB} (I - \Pi_{XB})] |1\rangle \langle 1|. $$

(Not that it can be viewed as a two-outcome measurement.) Then we can recognize the right-hand side of (50) as $Q_{\alpha}^*(\Phi(\rho_{XB})) \| \Phi(\tau_X \otimes \sigma_B))$. By the monotonicity of $Q_{\alpha}^*$ under CPTP maps,

$$Q_{\alpha}^*(\Phi(\rho_{XB})) \| \Phi(\tau_X \otimes \sigma_B)) \leq Q_{\alpha}^*(\rho_{XB}) \| \tau_X \otimes \sigma_B).$$

Note that this holds for every $\sigma_B \in S(B)$. Thus, it follows for all $\alpha > 1$,

$$\frac{\alpha}{\alpha - 1} \log (1 - P_e(C)) - \log \frac{|W|}{|\mathcal{X}|} \leq \inf_{\sigma_B \in S(B)} D_{\alpha}^*(\rho_{XB} \| \tau_X \otimes \sigma_B),$$

or equivalently

$$\frac{\alpha}{\alpha - 1} \log (1 - P_e(C)) - R \leq -H_{\alpha}^+(X|B).$$

Since $C$ is arbitrary of rate $R$, we obtain Eq. (49). By the additivity of $D_{\alpha}^*$ under tensor products, we find that any $n$-blocklength code with $R < H(X|B)_p$ satisfies

$$\text{sc}(n, R) \geq E_{sc}^*(R).$$

□

B. Matching Bound

The main result in this section is the $n$-shot upper bound on the decay exponent of the probability of success in terms of the $E_{sc}^*$ and additional residual terms, using a proof based on Mosonyi and Ogawa’s proof of an analogous result in study of the transmission of classical information over quantum channels [28]. Let us recall Theorem 4:

Theorem 4: For all $R < H(X|B)_p$, the finite blocklength strong converse exponent defined in (5) satisfies

$$\text{sc}(n, R) \leq E_{sc}^*(R) + \frac{c}{m} \log (m + 1) + O_m \left( \frac{1}{\sqrt{m}} \right)$$

for $c := \frac{3(2|\mathcal{X}| + 2)(2|\mathcal{X}| - 1)}{4}$ and any $m \in \mathbb{N}$, where we denote by $O_m \left( \frac{1}{\sqrt{m}} \right)$ any term which is bounded by $C_m \frac{1}{\sqrt{m}}$ for all $n$ large enough, for some constant $C_m$ depending only on $m$, $|\mathcal{X}|$, and $\rho_{XB}$. In particular, taking $n \to \infty$ then $m \to \infty$ yields

$$\limsup_{n \to \infty} \text{sc}(n, R) \leq E_{sc}^*(R).$$

This is a consequence of the following, more detailed, result.
Proposition VI.1 (n-Shot Strong Converse Matching Bound): Let \( R < H(X|B)_\rho, \) \( m \in \mathbb{N} \) and \( \delta \in (0, \frac{1}{4}) \). For all \( n > m \), we have the bound
\[
sc(n, R) \leq E_c^* (R) + \frac{c}{m} \log (m + 1) + \frac{1}{\sqrt{n-1}} f_1 (m, \delta) + \frac{1}{n-m} f_2 (n, m)
\]
for
\[
E_c^* (R) := \sup_{\alpha > 1} \frac{1 - \alpha}{\alpha} (R - H_{\alpha}^{-1} (X|B)_\rho)
\]
\[
e \frac{3(|H_B| + 2)(|H_B| - 1)}{2}
\]
\[
\sqrt{m} f_1 (m, \delta) := \sqrt{3} \log 2 \left( 4 \cosh (2 + \sqrt{|X|^m}) [\log (1 + 2 \sqrt{|X|^m})^2 + 1] + 4 \sqrt{2} \left( \log \left( 1 + \text{Tr} [\rho_{\text{supp} (\rho_B)} \rho_m^{-1/2} + \rho_m^{1/2}] \right) \right) \right) \log \frac{1}{1 - \sqrt{1 - \delta^2}}
\]
\[
f_2 (n, m) := 1 - \log (1 - 2^{-2(n-1)R})
\]
with \( \rho_m := \mathcal{P}_m (\rho_{\text{sym}}^m) \), where \( \mathcal{P}_m \) is the pinching map associated to \( 1 \otimes \sigma_m \), \( \sigma_m \) is the universal symmetric state on \( \mathcal{H}_B^\otimes m \) given in Lemma VI.7 below, and \( H_{\alpha}^{-1} (X|B)_\rho := \max_{\mu \in S_B} -D^\alpha (\rho_X \| \rho_B \| 1 \otimes \sigma_B) \) for \( D^\alpha \) being sandwiched Rényi divergence, see Eq. (12). In particular,
\[
\lim_{n \to \infty} \sup_{\rho} sc(n, R) \leq E_c^* (R).
\]
To establish Proposition VI.1, we first obtain a bound with \( E_c^0 \) (Proposition VI.2), and then exploit a pinching argument to further relate \( E_c^0 \) to the desired \( E_c^* \). The proof of Proposition VI.2 and Proposition VI.1 are delayed to Section VI-B.1 and VI-B.2, respectively.

Proposition VI.2: For any \( n \in \mathbb{N} \), \( R > 0 \), and \( \delta \in (0, \frac{1}{4}) \), we have
\[
sc(n, R) \leq E_c^0 (R) + \frac{1}{\sqrt{n}} \left[ \sqrt{4 \log 2} (e_2 + 1) + e_1 (\delta) \right] + \frac{1}{n} \left[ 1 - \log (1 - 2^{-nR}) \right]
\]
where
\[
e_1 (\delta) := 4 \sqrt{2} \left( \log \left( 1 + \text{Tr} [\rho_{\text{supp} (\rho_B)} \rho_m^{-1/2} + \rho_m^{1/2}] \right) \right) \log \frac{1}{1 - \sqrt{1 - \delta^2}}
\]
\[
e_2 := 4 \cosh (2 + 2 \sqrt{|X|}) [\log (1 + 2 \sqrt{|X|})]^2.
\]
In particular,
\[
\lim_{n \to \infty} \sup_{\rho} sc(n, R) \leq E_c^0 (R).
\]

1) Proving Proposition VI.2: Let us first describe the strategy we will follow to establish Proposition VI.2. Using the variational representation given in Proposition IV.6, one has
\[
E_c^0 (R) = \min_{\sigma \in S_B (X|B)} \left\{ D (\sigma_X \| \rho_X) + |H(X|B)_\sigma - R^+| \right\}.
\]
The main idea is to use this variational representation to introduce a “dummy state” \( \sigma_{XB} \) for which we can apply Theorem 1. Then we relate the probability of success for this state to that of the source state \( \rho_{XB} \).

To obtain explicit bounds \( sc(n, R) \), we will proceed in the n-shot setting. Let us define an \( n \)-shot analog of the right-hand side for \( \alpha \in (0, 1), \gamma > 0 \), and \( \delta > 0 \) by
\[
F(R, \rho_{XB}, n, \delta, \alpha, \gamma) := \min_{\sigma \in S_B (X|B)} \left\{ \frac{1}{n} D_{\max}^\delta (\sigma_{XB} \| \rho_{XB}^\otimes n) + |H_{\alpha}^\delta (X|B)_\sigma - R + \gamma| \right\}
\]
where \( D_{\max}^\delta (\sigma \| \rho) \) indicates the max-relative entropy smoothed by \( \delta \) (defined by (24)) in the distance of optimal purifications, denoted \( d_{\text{up}} \) (defined in (25)). This quantity is upper bounded by \( E_c^0 \) (with error terms) in Lemma VI.6 below. First, we establish the preliminary result given by Lemma VI.3. Next we bound the strong converse error exponent by \( F(R, n, \delta, \alpha, \gamma) \) in two steps (Lemmas VI.4 and VI.5 below), and use Lemma VI.6 to bound with \( E_c^0 \).

The following result allows us to compare the success probability of the same code \( C \) for \( \rho_{XB} \) and a dummy state \( \sigma_{XB} \in S_{\rho} (X|B) \).

Lemma VI.3: Let \( \sigma_{XB} \in S_{\rho} (X|B) \). For any code \( C \), and any \( a > 0 \) we have
\[
P_a (\rho_{XB}, C) \geq e^{-a} \left( P_a (\sigma_{XB}, C) - \text{Tr} [\sigma_{XB} - e^a \rho_{XB}] \right).
\]
Here, we write \( P_a (\rho_{XB}, C) \) and \( P_a (\sigma_{XB}, C) \) to emphasize the dependency on the state, which is taken to be \( \rho_{XB} \) when not explicitly indicated.

Proof of Lemma VI.3: Let \( C \) be a code with encoder \( E \) and a decoder given by the family of POVMs \( D = \{ D_w \} \subseteq \mathcal{W} \) where \( D_w = \{ \Pi_{x}^{(w)} \}_{x \in X} \). Let \( \Pi_{XB} = \sum_{x \in X} |x \rangle \langle x| \otimes \Pi_{x}^{(E(x))} \). Then
\[
\text{Tr} [\Pi_{XB} \rho_{XB}] = P_a (\rho_{XB}, C), \quad \text{Tr} [\Pi_{XB} \sigma_{XB}] = P_a (\sigma_{XB}, C)
\]
For any self-adjoint operator \( X \) and \( Y \) with \( 0 \leq Y \leq 1 \),
\[
\text{Tr} [X_Y] \geq \text{Tr} [X+Y] \geq \text{Tr} [X+Y] - \text{Tr} [X-Y] = \text{Tr} [XY].
\]
(54)
Since \( 0 \leq \Pi_{XB} = \sum_{x \in X} |x \rangle \langle x| \otimes \Pi_{x}^{(E(x))} \leq \sum_{x \in X} |x \rangle \langle x| \otimes \Pi_{1} = \Pi_{1XB}, \) eq. (54) with \( X = (\sigma_{XB} - e^a \sigma_{XB}) \) and \( Y = \Pi_{XB} \) yields
\[
\text{Tr} \left[ (\sigma_{XB} - e^a \sigma_{XB})_+ \right] \geq \text{Tr} \left[ \Pi_{XB} (\sigma_{XB} - e^a \sigma_{XB}) \right] = P_a (\sigma_{XB}, C) - e^a \left( P_a (\rho_{XB}, C) \right)
\]
which yields the result after dividing by \( e^a \). \( \square \)

In the following, we prove Proposition VI.2 using two lemmas, analogous to Lemmas 5.10 and 5.11 of [28]. Let us define
\[
F_1 (R, n, \delta, \alpha, \gamma) := \min_{\sigma_{XB} : H_{\alpha}^\delta (X|B)_\sigma \leq R - \gamma n} \frac{1}{n} D_{\max}^\delta (\sigma_{XB}^{\otimes n} \| \rho_{XB}^{\otimes n})
\]
\[
F_2 (R, n, \delta, \alpha, \gamma) := \min_{\sigma_{XB} : H_{\alpha}^\delta (X|B)_\sigma \geq R + \gamma n} \frac{1}{n} D_{\max}^\delta (\sigma_{XB}^{\otimes n} \| \rho_{XB}^{\otimes n}) + H_{\alpha}^\delta (X|B)_\sigma - R + \gamma
\]
6By the \( n \)-shot analog we mean the states \( \sigma_{XB} \) and \( \rho_{XB} \) are both replaced by the \( n \)-copy ones, and then we normalize the whole quantity \( D_{\max}^\delta \) by \( n \).
where the minima are restricted to $\sigma_{XB} \in S_p(XB)$. Then
\[
F(R, n, \delta, \alpha, \gamma) = \min \{ F_1(R, n, \delta, \alpha, \gamma), F_2(R, n, \delta, \alpha, \gamma) \}.
\] (55)

First, we will bound the error probability using $F_1$.

Lemma VI.4: Let $\rho_{XB}$ be a c-q source state. For any $\delta \in (0, \frac{1}{2})$, $\gamma > 0$, $\alpha \in [0, 1)$ and $n \geq 1$,
\[
sc(n, R) \leq F_1(R, n, \delta, \alpha, \gamma) - \frac{1}{n} \log \left( 1 - 4e^{-n(1-\alpha) - 2\delta} \right).
\] (56)

The proof proceeds by choosing a dummy state $\sigma_{XB}$ close to $\rho_{XB}$ with low conditional entropy. The probability of success with $\rho_{XB}$ under any code can be related to the probability of success with $\sigma_{XB}$ under the same code by Lemma VI.3; since the states are close, the bound is not too weak. Since Theorem 1 shows there exist good codes for $\sigma_{XB}$ (using that $\sigma_{XB}$ has low conditional entropy so that the error exponent is not too high), we hence obtain bounds on the probability of success for codes with $\rho_{XB}$.

Proof of Lemma VI.4: Let $r > F_1(R, n, \delta, \alpha, \gamma)$. Then there exists $\sigma_{XB}$ such that
\[
\frac{1}{n}D^\delta_n(\sigma_{XB} \| \rho_{XB}) \leq r
\] (57)
and
\[
H^1_\alpha(K|B)_{\sigma} \leq R - \gamma.
\] (58)

By Lemma VI.3 with $a = nr$, we have
\[
P_s(\rho_{XB}^\otimes n, C_n) \geq e^{-nr} (P_s(\sigma_{XB}^\otimes n, C_n) - Tr[(\sigma_{XB}^\otimes n - e^{nr}\rho_{XB}^\otimes n)])
\]
for any code $C_n$. Then, since $\frac{1}{n}D^\delta_n(\sigma_{XB}^\otimes n \| \rho_{XB}^\otimes n) \leq nr$, we have that $Tr[(\sigma_{XB}^\otimes n - e^{nr}\rho_{XB}^\otimes n)] \leq 2\delta$ by Lemma IV.3. Additionally, since $\alpha \mapsto H^1_\alpha(K|B)_{\sigma}$ is monotone decreasing (see Proposition IV.2), we have that $H^1_\alpha(K|B)_{\sigma} \leq H^1_\alpha(K|B)_{\sigma} \leq R$. Thus, we can choose a code $C_n$ with rate $R$ such that
\[
P_s(\sigma_{XB}^\otimes n, C_n) \geq 1 - 4\exp(-nrE^1_\alpha(\sigma_{XB}, R))
\]
by Theorem 1. Moreover,
\[
E^1_\alpha(\sigma_{XB}, R) = \sup_{s \in [0,1]} -s[H^1_{1-s}(K|B)_{\sigma} - R]
\]
\[
\geq (1 - \alpha)[R - H^1_\alpha(K|B)_{\sigma}] \geq (1 - \alpha),
\]
since $\alpha \in [0, 1)$. Therefore,
\[
1 - P_e(\rho_{XB}^\otimes n, C_n) = P_s(\sigma_{XB}^\otimes n, C_n) \geq e^{-nr} \left( 1 - 4e^{-n(1-\alpha) - 2\delta} \right).
\] (59)

Then since $P^*_e(n, R) \leq P_e(\rho_{XB}^\otimes n, C_n)$,
\[
\frac{1}{n} \log(1 - P^*_e(n, R)) \leq r - \frac{1}{n} \log \left( 1 - 4e^{-n(1-\alpha) - 2\delta} \right).
\]
Since this is true for any $r > F_1(R, n, \delta, \alpha, \gamma)$, we have (56).

Next, we bound $sc(n, R)$ by $F_2$.

Lemma VI.5: For all $n \geq 1$, $\delta \in (0, \frac{1}{2})$, $\alpha \in [0, 1)$, and $\gamma > 0$, we have
\[
sc(n, R) \leq F_2(R, n, \delta, \alpha, \gamma)
\]
\[
- \frac{1}{n} \log \left( 1 - 2^{-nR} \left( 1 - 4e^{-n(1-\alpha) - 2\delta} \right) \right).
\] (60)

Let us overview the proof. In this case, in contrast to the previous lemma, choosing a dummy state $\sigma_{XB}$ to achieve the minimum results in a state whose conditional entropy is not bounded above in terms of the rate $R$. Hence, we cannot immediately apply the achievability results for $\sigma$ to obtain a good code for $\rho_{XB}$ as was done in the previous proof. Instead, we choose a code for $\sigma_{XB}$ achieving a worse rate $R_1 \geq R$, and then modify the code by discarding difficult-to-recover sequences in order to use less compressed space and achieve the rate $R$ instead.

Proof of Lemma VI.5: Let $r > F_2(R, \rho_{XB}, n, \delta, \alpha, \gamma)$. Then there exists $\sigma_{XB}$ such that
\[
r > \frac{1}{n}D^\delta_n(\sigma_{XB}^\otimes n \| \rho_{XB}^\otimes n) + H^1_\alpha(K|B)_{\sigma} - R + \gamma
\] (61)
and $R - \gamma \leq H^1_\alpha(K|B)_{\sigma}$. Let
\[
R_1 := H^1_\alpha(K|B)_{\sigma} + \gamma
\]
\[
r_1 := r + R - H^1_\alpha(K|B)_{\sigma} - \gamma
\]
which has $r_1 \geq \frac{1}{n}D^\delta_n(\sigma_{XB}^\otimes n \| \rho_{XB}^\otimes n)$ by (61). We see then that equation (57) holds for the state $\sigma_{XB}$ when $r$ is replaced by $r_1$, and (58) holds for $\sigma_{XB}$ when $R$ is replaced by $R_1$. Therefore, (59) yields that there is a code $C_n$ with rate $R_1$ such that
\[
P_s(\rho_{XB}^\otimes n, C_n) \geq e^{-nr_1} \left( 1 - 4e^{-n(1-\alpha) - 2\delta} \right).
\] (62)

We have $R \leq R_1$ by construction. We aim to construct a code $\tilde{C}_n$ with rate $R$ and a high probability of success by pruning the low probability elements of $\mathcal{W}$, as follows. If $R = R_1$, take $\tilde{C}_n = C_n$. Then (63) holds, and we may continue the proof from there. Otherwise, $R < R_1$. Let $C_n$ have encoder $\mathcal{E}_n : \mathcal{X}^n \rightarrow \mathcal{W}_n$ and decoding POVMs $\mathcal{D}_w = \{ P^w_{\mathcal{S}^e}\}_{w \in \mathcal{X}^n}$, for $w \in \mathcal{W}_n$, where $|\mathcal{W}_n| = 2^nR_1$. Then given a sequence $x \in \mathcal{X}^n$, the probability of correctly decoding the sequence is given by
\[
\text{Tr}[\rho^\otimes n \Pi_{\mathcal{S}^e}(x)]
\]
where $\rho^\otimes n = \rho^\otimes 1 \otimes \rho^\otimes 2 \otimes \cdots \otimes \rho^\otimes n$. Similarly, writing $p(x) = p(x_1) \cdots p(x_n)$, the quantity
\[
P_w := \sum_{x \in \mathcal{X}^n : \mathcal{E}_n(x) = w} p(x) \text{Tr}[\rho^\otimes n \Pi^w_{\mathcal{S}^e}]\]
is the probability of success for sequences which are encoded as $w$. Then
\[
P_s(\rho^\otimes n, C_n) = \sum_{w \in \mathcal{W}_n} P_w.
\]

Therefore, $\mathbf{P} := \left( \frac{P_w}{P(\rho_{XB}^\otimes n, C_n)} \right)_{w \in \mathcal{W}_n}$ is a probability vector of length $2^nR_1$. Let $k = 2^nR_1 - 2^nR$. In order to make a code of rate $R$ from $C_n$, we need to remove $k$ elements from $\mathcal{W}_n$. We will choose the elements of $\mathcal{W}_n$ corresponding to the $k$ smallest entries of the vector $\mathbf{P}$, to keep as much probability
of success as possible. Without loss of generality, assume $P$ is in decreasing order: $P_1 \geq P_2 \geq \ldots$.

Let $\mathcal{R}_n := \{ j : j \geq 2^n R \}$ be the set of indices to remove (note $|\mathcal{R}_n| = k$). Define $W_n = W_n \setminus \mathcal{R}_n$. Choose $w_0 \in W$ to be the element of $W$ with index $2^n R - 1$, that is, with index corresponding to the smallest element of $P$ which has not been removed. Define the encoder

$$\tilde{E}_n(x) = \begin{cases} w_0 & \text{if } E_n(x) \in \mathcal{R}_n \\ E_n(x) & \text{else} \end{cases}$$

Then $\tilde{E}_n$ and the decoding POVMs $\{D_w\}_{w \in \tilde{W}_n}$ forms a code $\tilde{C}_n$ of rate $R$, since $|\tilde{W}_n| = 2^n R - k = 2^n R$.

Let us briefly introduce the majorization pre-order of vectors. Given $x, y \in \mathbb{R}^n$, we say $y$ majorizes $x$, written $x \prec y$, if $\sum_{i=1}^k y_k \geq \sum_{i=1}^k x_k$ for each $k = 1, \ldots, n$, with equality for the case $k = n$, where $y^n$ is the rearrangement of $y$ in decreasing order: $y^n_1 \geq \cdots \geq y^n_n$.

Setting $u = (\frac{1}{2^n R})_{w \in \tilde{W}_n}$ as the uniform distribution, we have

$$u \sim P$$

since the uniform distribution is majorized by every other probability vector. In particular,

$$\sum_{j=1}^{2^n R - k - 1} P_j \geq \frac{2^n R - k - 1}{2^n R} = \frac{2^n R - 1}{2^n R}.$$

The probability of success for any element which $\tilde{E}_n$ does not map to $\omega_0$ is the same as it was under the code $C_n$. Therefore,

$$P_s(\rho_{XB}^{\otimes n}, \tilde{C}_n) = \sum_{w \in \tilde{W}_n \setminus \{\omega_0\}} P_w + \sum_{x \in X^n \setminus \tilde{E}_n(x) = \omega_0} \text{Tr}[\rho_{E_x}^{\otimes n}]$$

$$\geq \sum_{w \in \tilde{W}_n \setminus \{\omega_0\}} P_w = P_s(\rho_{XB}^{\otimes n}, C_n) \sum_{j=1}^{2^n R - k - 1} P_j$$

$$\geq \frac{2^n R - 1}{2^n R} P_s(\rho_{XB}^{\otimes n}, C_n)$$

$$= \frac{2^n R - 1 - 2^n R}{2^n R} P_s(\rho_{XB}^{\otimes n}, C_n)$$

By (62), we have therefore

$$P_s(\rho_{XB}^{\otimes n}, \tilde{C}_n) \geq \frac{2^n R (1 - 2^n R)}{2^n R} e^{-nr_1} \left(1 - 4 e^{-\gamma(1 - \alpha)} - 2 \delta \right).$$

Since $1 - P_e(n, R) \leq P_s(\rho_{XB}^{\otimes n}, \tilde{C}_n)$, by substituting the definition of sc$(n, R)$ from (5) we find

$$sc(n, R) \leq -\frac{1}{n} \log \left( \frac{2^n R (1 - 2^n R)}{2^n R} \right)$$

$$+ r_1 - \frac{1}{n} \log \left(1 - 4 e^{-\gamma(1 - \alpha)} - 2 \delta \right)$$

$$= r_1 + R_1 - R$$

$$\leq r_1 + R_1 - R$$

$$= r_1 + R_1 - R$$

using $r_1 + R_1 - R = r$. Since this inequality holds for all $r > F_2(R, n, \delta, \alpha, \gamma)$ we have (60).

Together, Lemmas VI.4 and VI.5 establish an n-shot bound on sc$(n, R)$ in terms of $F(R, n, \delta, \alpha, \gamma)$, via (55). However, $F(R, n, \delta, \alpha, \gamma)$ itself is an approximation of the asymptotic quantity $E^\gamma_n$. In order to bound $(n, R)$ in terms of $E^\gamma_n$, therefore, we need to investigate this approximation. We can see as $n \to \infty$, $\alpha \to 1$, and $\gamma \to 0$,

$$F(R, n, \delta, \alpha, \gamma) \to E^\gamma_n$$

using the variational characterization of $E^\gamma_n(R)$ given in (53). To find an n-shot bound on the error exponent in terms of $E^\gamma_n$, therefore, we need to find the error terms hidden in this limit. This is accomplished by the following lemma.

**Lemma VI.6:** If $R > 0$, $\delta \in (0, \frac{1}{2})$, $\gamma > 0$, $n \in \mathbb{N}$ and $\alpha \in (\frac{1}{2}, 1)$, then

$$F(R, n, \delta, \alpha, \gamma) \leq E^\gamma_n(R) + \gamma + \frac{1}{\sqrt{n}} e_1(\delta) + (1 - \alpha) e_2 (64)$$

where $e_1(\delta)$ and $e_2$ are defined in (51) and (52).

The key tools in proving this result are the asymptotic equipartition property of $D_{\max}(\sigma \| \rho)$ and continuity bounds for the conditional Rényi relative entropy in order to bound the terms in $F(R, n, \delta, \alpha, \gamma)$, as follows.

**Proof of Lemma VI.6:** Consider the variational representation

$$E^\gamma_n(R) = \min_{\sigma \in S_n(X|B)} \left\{ D(\sigma_{XB} \| \rho_{XB}) + |H(X|B)_\sigma - R|^+ \right\}$$

(65)

given by Proposition IV.6. Let $\sigma^*_X|B$ achieve the minimum in (65). Note $\sigma^*_X|B$ depends on $\rho_{XB}$ and $R$. Then

$$F(R, \rho_{XB}, n, \delta, \alpha, \gamma)$$

$$= \min_{\rho_{XB} \in S_n(X|B)} \left\{ \frac{1}{n} D_{\max}(\sigma_{XB}^{\otimes n} \| \rho_{XB}^{\otimes n}) + |H^1_{\sigma}(X|B)_\sigma - R + \gamma|^+ \right\}$$

$$\leq \frac{1}{n} D_{\max}(\sigma^*_{XB}^{\otimes n} \| \rho_{XB}^{\otimes n}) + |H^1_{\sigma}(X|B)_\sigma - R + \gamma|^+.$$

By (27), we have

$$\frac{1}{n} D_{\max}(\sigma^*_{XB}^{\otimes n} \| \rho_{XB}^{\otimes n})$$

$$\leq D(\sigma^*_X \| \rho_{XB})$$

$$+ \frac{1}{\sqrt{n}} 4 \sqrt{2} (\log \eta^*) \log \frac{1}{1 - \sqrt{1 - \delta_2}}$$

where

$$\eta^* = 1 + \text{Tr}[\sigma^*_{XB}^{3/2} \rho_{XB}^{1/2}] + \text{Tr}[\sigma^*_{XB}^{1/2} \rho_{XB}^{3/2}]$$

We can remove the dependence on $\sigma^*_X$ by the inequality

$$\eta^* \leq 1 + \text{Tr}[P_{\text{supp}}(\sigma) \rho_{XB}^{1/2} + \rho_{XB}^{1/2}]$$

which follows from $\sigma^*_X \leq \rho_{XB}$, the orthogonal projection onto the support of $\sigma$, which holds for all $\alpha > 0$, and since $P_{\text{supp}}(\sigma) \leq P_{\text{supp}}(\sigma)$ as $\sigma^*_X \in S_n(X|B)$. In particular, this removes the dependence on the rate $R$. Therefore we have

$$\frac{1}{n} D_{\max}(\sigma^*_{XB}^{\otimes n} \| \rho_{XB}^{\otimes n}) \leq D(\sigma^*_X \| \rho_{XB}) + \frac{1}{\sqrt{n}} e_1(\delta).$$

(66)
Next, we use a continuity bound in \( \alpha \) for \( H_{\alpha}^{1} \): by Lemma 2.3 of [60], for any \( c > 0 \),
\[
H_{\alpha}^{1}(X|B)_{\sigma^{*}} \leq H(X|B)_{\sigma^{*}} + 4 \cosh c(1-\alpha)(\log \eta')^2
\]
where
\[
\eta' := 1 + e^{\frac{1}{2}H_{\alpha}^{1}(X|B)_{\sigma^{*}} + e^{-\frac{1}{2}H_{\alpha}^{1}(X|B)_{\sigma^{*}}}
\]
which holds when \( |1-\alpha| \leq \min \left\{ \frac{1}{2}, \frac{2 \log \eta'}{c} \right\} \). Note that for any state \( \omega \) and \( \alpha \in [0, 1] \), we have
\[
- \log \min(\text{rank} \omega_X, \text{rank} \omega_B) \leq H_{\alpha}^{1}(X|B)_{\omega} \leq \log \text{rank} \omega_X
\]
by Lemma 5.2 of [25]. Thus,
\[
1 + \frac{1}{\sqrt{\min([X], |B|)}} + \frac{1}{\sqrt{|X|}} \leq \eta' \quad (67)
\]
\[
\leq 1 + \sqrt{|X|} + \min([X], |B|) \leq 1 + 2 \sqrt{|X|}
\]
independently of \( \sigma^{*}_{X,B} \). Then \( \log \eta' > 0 \), and
\[
(\log \eta')^2 \leq (\log(1 + 2 \sqrt{|X|}))^2.
\]
Moreover, since \( \log \eta' > 0 \), we can take \( c = \log \eta' \) to recover for \( \alpha \in (\frac{1}{2}, 1) \),
\[
H_{\alpha}^{1}(X|B)_{\sigma^{*}} \leq H(X|B)_{\sigma^{*}} + 4 \cosh(\eta')(1-\alpha)(\log \eta')^2 \leq H(X|B)_{\sigma^{*}} + (1-\alpha)e_2 (68)
\]
using (67).

Then for any \( R > 0 \), \( \delta \in (0, \frac{1}{2}) \), \( \gamma > 0 \), and \( \alpha \in (\frac{1}{2}, 1) \) such that \( H_{\alpha}^{1}(X|B)_{\sigma^{*}} - R + \gamma \geq 0 \), we have
\[
F = F(R, n, \delta, \alpha, \gamma)
\leq D(\sigma^{*}_{X,B}||\rho_{X,B}) + H_{\alpha}^{1}(X|B)_{\sigma^{*}} - R + \gamma
\]
\[
+ \frac{1}{\sqrt{n}}e_1(\delta) \quad (69)
\]
\[
\leq D(\sigma^{*}_{X,B}||\rho_{X,B}) + H(X|B)_{\sigma^{*}} - R + \gamma
\]
\[
+ \frac{1}{\sqrt{n}}e_1(\delta) + (1-\alpha)e_2
\]
\[
\leq D(\sigma^{*}_{X,B}||\rho_{X,B}) + |H(X|B)_{\sigma^{*}} - R| + \gamma
\]
\[
+ \frac{1}{\sqrt{n}}e_1(\delta) + (1-\alpha)e_2
\]
\[
eq E_{sc}^{\gamma}(R) + \gamma + \frac{1}{\sqrt{n}}e_1(\delta) + (1-\alpha)e_2 \quad (70)
\]
where in (69) we use (66), and in (70) we use (68).

On the other hand, if \( R > 0 \), \( \delta \in (0, \frac{1}{2}) \), \( \gamma > 0 \), and \( \alpha \in (\frac{1}{2}, 1) \), and \( H_{\alpha}^{1}(X|B)_{\sigma^{*}} - R + \gamma < 0 \), then
\[
F = F(R, n, \delta, \alpha, \gamma)
\leq D(\sigma^{*}_{X,B}||\rho_{X,B}) + \frac{1}{\sqrt{n}}e_1(\delta) \quad (71)
\]
\[
= E_{sc}^{\gamma}(R) + \frac{1}{\sqrt{n}}e_1(\delta) \quad (72)
\]
where in (71) we use (66), and in (72) we use (70) we use (68).

With these results, we may prove Proposition VI.2 simply by choosing \( \alpha, \delta, \) and \( \gamma \) to scale appropriately with \( n \), as follows.

\section*{Proof of Proposition VI.2:}

By (55) and Lemmas VI.4 and VI.5, we have the bound
\[
sc(n, R) \leq F(R, n, \delta, \alpha, \gamma) \quad (73)
\]
\[
- \frac{1}{n} \log \left[ (1 - 2^{-nR}) \left( 1 - 4 e^{-n\gamma(1-\alpha) - \delta} \right) \right]
\]
for any \( R > 0 \), \( n \geq 1 \), \( \delta \in (0, \frac{1}{2}) \), \( \alpha \in (0, 1) \), and \( \gamma > 0 \). We choose \( \delta \in (0, 1/4) \) and \( \kappa = 4 \log 2 = 16 \). Then \( 4 e^{-\kappa} + \delta \leq \frac{1}{2} \). For \( x \in [0, \frac{1}{2}] \), we have the bound \( -\log(1-x) \leq 2x \), and therefore
\[
- \log \left( 1 - 4 e^{-\kappa} - \delta \right) \leq 2 \cdot \frac{1}{2} = 1.
\]
Now, choose \( \alpha = 1 - \frac{\sqrt{\kappa}}{n} \), and \( \gamma = \sqrt{\frac{\kappa}{n}} \). Then for \( \delta \in (0, \frac{1}{4}) \) by (73), we have
\[
sc(n, R) \leq F(R, n, \delta, \alpha, \gamma) + \frac{1}{n} \left[ 1 - \log(1 - 2^{-nR}) \right].
\]
Thus, by Lemma VI.6, we have
\[
sc(n, R) \leq E_{sc}^{\gamma}(R) + \frac{\sqrt{\kappa}}{n} [1 + e_2] + \frac{1}{\sqrt{n}} e_1(\delta) + \frac{1}{n} \left[ 1 - \log(1 - 2^{-nR}) \right]
\]
where \( e_1(\delta) \) and \( e_2 \) are defined in (51) and (52). Taking \( n \to \infty \) recovers the asymptotic bound. \( \square \)

2) \section*{Proving Proposition VI.1:}

In order to prove Proposition VI.1, we first need to introduce the idea of universal symmetric state, which is obtained via representation theory (see e.g. [61]–[63], [28, Appendix A]). Briefly, \( \mathcal{H}_{B}^{\otimes m} \) is decomposed into a direct sum of tensor products of irreducible representations \( U_{\lambda} \) of the symmetric group on \( m \) letters and \( V_{\lambda} \) of the special unitary group of dimension \( d := |\mathcal{H}_{B}| \),
\[
\mathcal{H}_{B}^{\otimes m} \cong \bigoplus_{\lambda \in Y_{m}} U_{\lambda} \otimes V_{\lambda}
\]
where the index set \( Y_{m} \) consists of Young diagrams of depth up to \( d \). The state \( \sigma_{a,m} \) is constructed by taking the maximally mixed state on each representation:
\[
\sigma_{a,m} = \bigoplus_{\lambda \in Y_{m}} \frac{1}{\text{dim} U_{\lambda}} \frac{1}{\text{dim} V_{\lambda}} U_{\lambda} \otimes V_{\lambda}
\]

\section*{Lemma VI.7 (\cite{62}, Lemma 2.8 of \cite{28}):}

There is a symmetric state \( \sigma_{a,m} \) on \( \mathcal{H}_{B}^{\otimes m} \) such that for any other symmetric state \( \omega \) on \( \mathcal{H}_{B}^{\otimes m} \), we have that \( \sigma_{a,m} \) commutes with \( \omega \) and obeys the bound
\[
\omega \leq v_{m,d} \sigma_{a,m}
\]
where \( v_{m,d} \) is a number obeying \( v_{m,d} \leq (m+1) \frac{(d+2)(d-1)}{2} \) where \( d = \dim \mathcal{H} \). Additionally, the number of different eigenvalues of \( \sigma_{a,m} \) is bounded by \( v_{m,d} \).

In the following, we employ a pinching argument to relate \( E_{sc}^{\gamma} \) to \( E_{sc}^{\gamma} \) (see Section IV for the definition of pinching). This argument uses the fact that the log-Euclidean Rényi relative entropy \( D_{\alpha}^{\gamma} \) and the sandwiched Rényi entropy \( D_{\alpha}^{\gamma} \) coincide when their arguments commute. This commutativity is ensured by pinching \( \rho_{X,B} \) with the universal symmetric state described above. Then properties of the universal symmetric state are
used to show that the relevant rates evaluated at $\rho_{XB}$ and the pinched version of $\rho_{XB}$ differ only at the lower order terms. This kind of argument has been used previously for similar bounds in [63, Lemma 3] and [28, Lemma 4.10].

**Lemma VI.8:** Let $P_m(\rho_{XB}^m)$ be the map pinching by $I_{Xm} \otimes \sigma_{a,m}$, and define $\rho_m = P_m(\rho_{XB}^m)$. Then for $t < s < 0$,

$$E^{s}_{j}(\rho_m, s) \leq mE^{s}_{0}(s) - s \log v_{m, |\mathcal{H}_B|}. \quad (74)$$

In particular,

$$E^{s}_{\infty}(\rho_m, mR) \leq mE^{s}_{\infty}(R) + 3\log v_{m, |\mathcal{H}_B|}. \quad (75)$$

Here, we write $E^{s}_{j}(\rho_m, s)$ and $E^{s}_{\infty}(\rho_m, mR)$ to emphasize the dependency on $\rho_m$. For $E^{s}_{0}(s)$ and $E^{s}_{\infty}(R)$, the underlying system is given by $\rho_{XB}$.

**Proof of Lemma VI.8:** Recall for $t = \{s\}, \{\bar{s}\}$ the definition

$$E^{s}_{0}(s) = -sH^{\downarrow s}(X|B)_{\rho}. \quad (76)$$

Let us then consider, for $\alpha > 1$, the quantities

$$H^{\downarrow \alpha, \uparrow}(X^m|B^m)_{\rho_m} = \max_{\sigma_{B^m}} -D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m}) \quad (77)$$

and

$$H^{\downarrow, \uparrow}_{\alpha, \uparrow}(X^m|B^m)_{\rho_m} = \max_{\sigma_{B^m}} -D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m}). \quad (78)$$

The symmetric group on $m$ letters has a natural unitary representation on $\mathcal{H}^{\otimes m}$ as follows (see e.g. [64]). For $\pi \in \mathfrak{S}_m$, we define

$$\pi_{\mathcal{H}} := \sum_{i_1, \ldots, i_m} \ket{i_{\pi^{-1}(1)}, \ldots, i_{\pi^{-1}(m)}} \bra{i_1, \ldots, i_m},$$

which is a representation of $\mathfrak{S}_m$ on $\mathcal{H}^{\otimes m}$. Then define a mixed-unitary channel $\text{Sym}_{B^m}$ given by

$$\text{Sym}_{B^m}(\sigma_{B^m}) := \frac{1}{m!} \sum_{\pi \in \mathfrak{S}_m} \pi_{B^m}(\sigma_{B^m}) \pi_{B^m}. \quad (79)$$

Since $D^{(t)}_{\alpha}$ is convex in its second argument for any $\alpha \in [\frac{1}{2}, \infty)$ for both $t = \{s\}$ and $t = \{\bar{s}\}$ by Proposition 3.18 of [28], we have

$$D^{(t)}_{\alpha}(\rho_m||I_{Xm} \otimes \text{Sym}_{B^m}(\sigma_{B^m})) \leq \frac{1}{m!} \sum_{\pi \in \mathfrak{S}_m} D^{(t)}_{\alpha}(\rho_m||I_{Xm} \otimes \pi_{B^m}(\sigma_{B^m}) \pi_{B^m}). \quad (79)$$

But since $\rho_m$ is pinched by $I_{Xm} \otimes \sigma_{a,m}$, the state $\rho_m$ commutes with $I_{Xm} \otimes \sigma_{a,m}$ and therefore with $I_{Xm} \otimes \pi_{B^m}$ for every $\pi \in \mathfrak{S}_m$. Thus,

$$D^{(t)}_{\alpha}(\rho_m||I_{Xm} \otimes \pi_{B^m}(\sigma_{B^m}) \pi_{B^m}) = D^{(t)}_{\alpha}(I_{Xm} \otimes \pi_{B^m}(\sigma_{B^m}) \pi_{B^m})$$

$$= D^{(t)}_{\alpha}(I_{Xm} \otimes \pi_{B^m}(\sigma_{B^m})) \pi_{B^m}(I_{Xm} \otimes \sigma_{B^m}) \pi_{B^m}$$

$$= D^{(t)}_{\alpha}(\rho_m||I_{Xm} \otimes \sigma_{B^m})$$

since $D^{(t)}_{\alpha}$ is invariant under unitary conjugation. Thus, by (79), we have

$$-D^{(t)}_{\alpha}(\rho_m||I_{Xm} \otimes \text{Sym}_{B^m}(\sigma_{B^m})) \geq -D^{(t)}_{\alpha}(\rho_m||I_{Xm} \otimes \sigma_{B^m}).$$

Since $\text{Sym}_{B^m}(\sigma_{B^m})$ is symmetric, we have that $-D^{(t)}_{\alpha}$ is maximized on a symmetric state for $t = \{s\}, \{\bar{s}\}$, so we may restrict to symmetric states in the maximums in (77) and (78).

Moreover, again since $\rho_m$ is pinched by $I_{Xm} \otimes \sigma_{a,m}$, we have that $\rho_m$ commutes with $I_{Xm} \otimes \sigma_{a,m}$, and therefore with $I_{Xm} \otimes \sigma_{B^m}$, for any symmetric state $\sigma_{B^m}$. By commutativity then, for any symmetric $\sigma_{B^m}$, we have

$$D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m}) = D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m}).$$

Therefore,

$$H^{\downarrow \alpha, \uparrow, \bar{s}}_{\alpha}(X^m|B^m)_{\rho_m} = \max_{\sigma_{B^m} \in \mathcal{S}(B^m)} -D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m})$$

$$= \max_{\sigma_{B^m} \in \mathcal{S}(B^m)} -D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m})$$

$$= \max_{\sigma_{B^m} \in \mathcal{S}(B^m)} -D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m})$$

$$= \max_{\sigma_{B^m} \in \mathcal{S}(B^m)} -D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m})$$

and

$$= \max_{\sigma_{B^m} \in \mathcal{S}(B^m)} -D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m})$$

However, we want a quantity in terms of $\rho_{XB}^m$ instead of $\rho_m$. For any symmetric $\sigma_{B^m}$, we have $\sigma_{B^m} \leq v_{m, |\mathcal{H}_B|/\sigma_{a,m}}$ so by Lemma 3.24 of [28],

$$D^{\alpha}_{\rho_m}(\rho_m||I_{Xm} \otimes \sigma_{B^m}) \geq D^{\alpha}_{\rho_{XB}^m}(\rho_{XB}^m||I_{Xm} \otimes \sigma_{a,m})$$

$$= D^{\alpha}_{\rho_{XB}^m}(\rho_{XB}^m||I_{Xm} \otimes \sigma_{a,m}) - \log v_{m, |\mathcal{H}_B|}$$

and then by Lemma 3 of [63], we have

$$D^{\alpha}_{\rho_{XB}^m}(\rho_{XB}^m||I_{Xm} \otimes \sigma_{a,m}) \geq D^{\alpha}_{\rho_{XB}^m}(\rho_{XB}^m||I_{Xm} \otimes \sigma_{a,m}) - 2\log v_{m, |\mathcal{H}_B|}.$$ 

Putting it together, by (80),

$$H^{\downarrow \alpha, \uparrow, \bar{s}}_{\alpha}(X^m|B^m)_{\rho_m} \leq \frac{1}{m^s} \sum_{e \in (-1,0)} \frac{1}{m^s} \sum_{e \in (-1,0)} \frac{1}{m^s} \sum_{e \in (-1,0)}$$

By writing $\alpha = \frac{1}{1+s}$ for $s \in (-1,0)$ and using the definition (76), we recover (74). Then, using $H^{\downarrow \alpha, \uparrow, \bar{s}}_{\alpha}(X^m|B^m)_{\rho_{XB}^m} = mH^{\downarrow \alpha, \uparrow, \bar{s}}_{\alpha}(X|B)_{\rho}$, and the definitions

$$E^{s}_{\infty}(\rho_m, mR) = \sup_{-1<s<0} [E^{s}_{\text{SW}}(s, \rho_m) + smR]$$

and

$$mE^{s}_{\infty}(R) = \sup_{-1<s<0} [mE^{s}_{\infty}(s) + mR]$$

we obtain (75).

Next, we proceed to establish the main result of this section, Proposition VI.1. The proof proceeds by establishing a code for the pinched state $\rho_m$, whose rate is characterized by $E^{s}_{\infty}(R)$ (using Proposition VI.2 and Lemma VI.8) and then using it to establish a code for $\rho_{XB}$ with the same rate.

**Proof of Proposition VI.1:** Let $n, m \in \mathbb{N}$, $n \geq m > 1$, and $\delta \in (0, \frac{1}{2})$. Let $k \in \mathbb{N}$ such that $km \leq n \leq (k+1)m$. We will consider $\rho_m = P_m(\rho_{XB})$, the state $\rho_{XB}^m$ pinched by the universal symmetric state, and apply Proposition VI.2 to obtain a code for $\rho_{XB}^m$ and bounds on its probability of success. From this, we will construct a code for $\rho_{XB}^m$. 


We define
\[ T(m)(\delta, k) := \sqrt{4\log(2(e_2^{(m)} + 1) + e_1^{(m)}(\delta))} \]
\[ + \frac{1}{\sqrt{k}}[1 - \log(1 - 2^{-kR})] \]
where
\[ e_1^{(m)}(\delta) := \frac{2}{\sqrt{\delta}} \left[ \log \left( 1 + \sqrt{\log(1 - 1 - \delta)} \right) \right] \]
\[ e_2^{(m)} := 4 \cosh(1 + 2\sqrt{|X|^m} |\log(1 + 2\sqrt{|X|^m}|)^2). \]

By Proposition VI.2 for any rate \( R > 0 \), we can construct a code \( C_k^{(m)} \) for \( k \) copies of the state \( \rho_m \) with rate \( R_m \) such that
\[ \frac{1}{k} \log P_s(\rho_m, C_k^{(m)}) \leq E_{sc}(\rho_m, mR) + \frac{1}{\sqrt{k}} T(m)(\delta, k). \]

We may use Lemma VI.8 to upper bound \( E_{sc}(\rho_m, mR) \) to obtain
\[ \frac{1}{k} \log P_s(\rho_m, C_k^{(m)}) \leq m E_{sc}(R) + 3 \log v_m |H_B| \]
\[ + \frac{1}{\sqrt{k}} T(m)(\delta, k). \quad (81) \]

From this code, we wish to construct a code \( C_n \) for \( n \) copies of \( \rho_{XB} \) with rate \( R \) that has the same probability of success as \( C_k^{(m)} \) does for \( \rho_m \).

\( C_k^{(m)} \) consists of an encoder \( c_k^{(m)} : X^{km} \rightarrow W_{kmR} \), where \( |W_{kmR}| = 2^{kmR} \), and a decoder \( D_k^{(m)} \) that associates to any element \( w \in W_{kmR} \) a POVM \( \{\Pi_w^{(m)}\}_{w \in X^{km}} \) on \( H_{\rho_B^{km}} \) with outcomes in \( X^{km} \). Note
\[ \rho_m = \sum_{x \in X^{km}} p(x)|x| \otimes \mathcal{P}_m(\rho_B^x) \]
for \( \rho_B^x = \rho_B^{x_1} \otimes \cdots \otimes \rho_B^{x_{km}} \) and \( p(x) = p(x_1) \cdots p(x_{km}) \) when \( x = (x_1, \ldots, x_{km}) \). Moreover,
\[ P_s(\rho_m, c_k^{(m)}) = \sum_{x \in X^{km}} p(x) \text{Tr}[\Pi_x^{(m)}(\rho_B^x) \otimes m(\rho_B^{km})]. \]

We may replace the code \( c_k^{(m)} \) by one with a decoder invariant under the pinching operation,
\[ \Pi_x^{(m)}(\rho_B^x) = \mathcal{P}_m(\Pi_x^{(m)}(\rho_B^x)) \quad (82) \]
since the success probability does not change:
\[ \sum_{x \in X^{km}} p(x) \text{Tr}[\Pi_x^{(m)}(\rho_B^x) \otimes m(\rho_B^{km})] \]
\[ = \sum_{x \in X^{km}} p(x) \text{Tr}[\mathcal{P}_m(\Pi_x^{(m)}(\rho_B^x)) \otimes m(\rho_B^{km})] \]
using that \( \mathcal{P}_m \) consists of a sum of conjugations by orthogonal projections. Moreover,
\[ \sum_{x \in X^{km}} p(x) \text{Tr}[\mathcal{P}_m(\Pi_x^{(m)}(\rho_B^x)) \otimes m(\rho_B^{km})] \]
\[ = \sum_{x \in X^{km}} p(x) \text{Tr}[\Pi_x^{(m)}(\rho_B^x) \otimes m(\rho_B^{km})] \]
and assuming (82), we find
\[ P_s(\rho_m, c_k^{(m)}) = \sum_{x \in X^{km}} p(x) \text{Tr}[\Pi_x^{(m)}(\rho_B^x) \otimes m(\rho_B^{km})]. \quad (83) \]

Let us define the code \( C_n \) as follows. Let \( W_{nk} \cap W_{kmR} \)

have size \( 2^{n+k} \), and define \( \mathcal{E}_n : X^n \rightarrow W_{nk} \) by
\[ \mathcal{E}_n : (x_1, \ldots, x_{km}) \mapsto \mathcal{E}_k^{(m)} (x_1, \ldots, x_{km}). \]

Note then \( C_n \) has rate \( R \). Next, define the decoder for \( w \in W_{kmR} \cap W_{nk} \)

by \( \Pi_w^{(w)} = \Pi_w^{(w)} \otimes m(\rho_B^{km}) \)

where \( \Pi_w^{(w)} \) is the POVM element defined by the decoder of \( c_k^{(m)} \). For \( w \in W_{nk} \cap W_{kmR} \), let \( \Pi_w^{(w)} \) be any POVM. Let us evaluate the success probability of this code. By definition
\[ P_s(\rho_{XB}, C_n) = \sum_{x \in X^n} p(x) \text{Tr}[\Pi_x^{(w)}(\rho_B^x)]. \]

We may split the sum as
\[ P_s(\rho_{XB}, C_n) = \sum_{y \in X^{km}} \sum_{x = (x_1, \ldots, x_{km}) \in X^n \cap \{y \}} p(x_{km+1}) \cdots p(x_n) \text{Tr}[\Pi_y^{(w)}(\rho_B^x)]. \]

Using \( p(x) = p(y)p(x_{km+1}) \cdots p(x_n) \) for \( x = (y_1, \ldots, y_{km}, x_{km+1}, \ldots, x_n) \), and \( \mathcal{E}_n(x) = c_x^{(m)}(y) \),
we have
\[ P_s(\rho_{XB}, C_n) = \sum_{y \in X^{km}} \sum_{x = (x_1, \ldots, x_{km}) \in X^n \cap \{y \}} p(y) \text{Tr}[\Pi_y^{(w)}(\rho_B^x)]. \]

By (84), and evaluating the trace over the last \( n - km \) copies of \( H_B \),
\[ P_s(\rho_{XB}, C_n) = \sum_{y \in X^{km}} \sum_{x_1 = \mathcal{E}_n(x)} p(x_{km+1}) \cdots p(x_n) \text{Tr}[\Pi_y^{(w)}(\rho_B^x)]. \]

Here, we have denoted by \( x_{km+1} : = (x_1, \ldots, x_{km}) \) and \( x_{km} := (x_1, \ldots, x_{km}) \). Then since \( \sum_{x_{km}} p(x_{km+1}) \cdots p(x_n) = 1 \), by (83) we have
\[ P_s(\rho_{XB}, C_n) = P_s(\rho_m, C_k^{(m)}). \] Therefore, using \( n \geq mk \) and (81), we have
\[ \frac{1}{n} \log P_s(\rho_{XB}, C_n) \leq \frac{1}{mk} \log P_s(\rho_{XB}, C_n) \]
\[ \leq E_{sc}(R) + \frac{3}{m} \log v_m |H_B| \]
\[ + \frac{1}{m} \sqrt{k} E_{sc}(\rho_B^x). \]
By using the upper bound on $v_{m, |\mathcal{H}_B|}$ given in Lemma VI.7 and that $1 - P^*_B(n, R) \geq P_s(\rho_X B, C_n)$, we find

$$- \frac{1}{n} \log (1 - P^*_B(n, R)) \leq E^*_m(R) + \frac{3(|\mathcal{H}_B| + 2)(|\mathcal{H}_B| - 1)}{2} \frac{1}{m} \log (m + 1) + \frac{1}{m \sqrt{k}} T^{(m)}(\delta, k).$$

Finally, we take $k = \lfloor \frac{m}{n} \rfloor$ and make use of the inequality $k \geq \frac{n}{m} - 1$ to obtain the result.

**VII. Moderate Deviation Regime**

In this section, we investigate the scenario in which the compression rate is not fixed but instead approaches the compression limit (from above) at a speed slower than $O(1/\sqrt{n})$. In particular, in Section VII-A, we apply the technique employed in Section V to show that in this scenario, the optimal error decays subexponentially (Theorem 6). In Section VII-B, we prove that, on the other hand, for errors decaying subexponentially the associated optimal compression rate converges to the compression limit at a speed slower than $O(1/\sqrt{n})$ (Theorem 7).

**A. Optimal Error When the Rate Approaches the Slepian–Wolf Limit Moderately Quickly**

Recall that a sequence of real numbers $(a_n)_{n \in \mathbb{N}}$ satisfying (8) is called a moderate deviation. We invoke Theorem 6:

**Theorem 6:** Assume that the c-q state $\rho_{XB}$ has strictly positive conditional information variance $V(X|B)_\rho$, where

$$V(X|B)_\rho := V(\rho_{XB}||I_X \otimes \rho_B)$$

with $V(\rho||\sigma) := \text{Tr}[\rho (\log \rho - \log \sigma)] - D(\rho||\sigma)^2$. Then for any sequence $(a_n)_{n \in \mathbb{N}}$ satisfying Eq. (8),

$$\lim_{n \to \infty} \frac{1}{n a_n^2} \log \left( \frac{1}{P^*_e(n, R_n)} \right) = - \frac{1}{2V(X|B)_\rho}$$

for $R_n := H(X|B)_\rho + a_n$.

**Proof of Theorem 6:** We shorthand $H = H(X|B)_\rho$, $V = V(X|B)_\rho$ for notational convenience. We first show the achievability, i.e. the "$\leq" in Eq. (9). Let $(a_n)_{n \geq 1}$ be any sequence of real numbers satisfying Eq. (8). For every $n \in \mathbb{N}$, Theorem 1 implies that there exists a sequence of $n$-block codes with rates $R_n = H + a_n$ so that

$$P^*_e(n, R_n) \leq 4 \exp \left\{ - n \max_{0 \leq s \leq 1} \left\{ E^1_0(s) + s R_n \right\} \right\}.$$  

Applying Taylor’s theorem to $E^1_0(s)$ at $s = 0$ together with Proposition IV.4 gives

$$E^1_0(s) = - s H - \frac{s^2}{2} V + \frac{s^3}{6} \frac{\partial^3 E^1_0(s)}{\partial s^3} \bigg|_{s = \bar{s}},$$

for some $\bar{s} \in [0, s]$. Now, let $s_n = a_n/V$. Then, $s_n \leq 1$ for all sufficiently large $n$ by the assumption in Eq. (8) and $V > 0$. For all $s_n \leq 1$,

$$\max_{0 \leq s \leq 1} \left\{ E^1_0(s) + s R_n \right\} \geq E^1_0(s_n) + s_n R_n.$$  

Then Eq. (86) yields

$$\max_{0 \leq s \leq 1} \left\{ E^1_0(s) + s R_n \right\} \geq \frac{a_n}{V} (-H + R_n) - \frac{a_n^2}{2V} + \frac{a_n^3}{6V^3} \frac{\partial^3 E^1_0(s)}{\partial s^3} \bigg|_{s = \bar{s}}$$

where $\bar{s} \in [0, s_n]$. Then, since $R_n = H + a_n$,

$$\max_{0 \leq s \leq 1} \left\{ E^1_0(s) + s R_n \right\} \geq \frac{a_n^2}{2V} - \frac{a_n^3}{6V^3} \frac{\partial^3 E^1_0(s)}{\partial s^3} \bigg|_{s = \bar{s}} \geq \frac{a_n^2}{2V} - \frac{a_n^3}{6V^3} \bar{s},$$

where

$$\bar{s} = \max_{s \in [0, 1]} \left| \frac{\partial^3 E^1_0(s)}{\partial s^3} \right|.$$  

This quantity is finite due to the compactness of $[0, 1]$ and the continuity, item a in Proposition IV.4. Therefore, substituting Eq. (87) into Eq. (85) gives, for all sufficiently large $n \in \mathbb{N}$,

$$\frac{1}{n a_n^2} \log \left( \frac{1}{P^*_e(n, R_n)} \right) \geq - \frac{4}{n a_n^2} + \frac{1}{2V} \left( 1 - \frac{\bar{s} a_n}{3V^2} \right).$$

Recall Eq. (8) and let $n \to \infty$, which completes the lower bound:

$$\lim_{n \to \infty} \frac{1}{n a_n^2} \log \left( \frac{1}{P^*_e(n, R_n)} \right) \geq \frac{1}{2V}.$$  

We move on to show the converse, i.e. the "$\geq" in Eq. (9). Let $N_1 \in \mathbb{N}$ be an integer such that $R_n = H + a_n \in (H_1(X|B)_\rho, H_0(X|B)_\rho)$ for all $n \geq N_1$. We denote by $(\alpha_{R_n}, \sigma^*_{R_n})$ the unique saddle-point of

$$\sup_{\alpha \in [0, 1]} \inf_{\sigma \in S(B)} \left\{ \frac{1 - \alpha}{\alpha} (R_n \rho_{XB}) \right\} \leq \frac{1}{\alpha} (R_n \rho_{XB} \otimes \sigma^*_{R_n}) \leq \frac{1}{\alpha} (R_n \rho_{XB} \otimes \sigma^*_{R_n} \otimes \sigma^*_{R_n}).$$

By invoking the one-shot converse bound, Proposition V.1, with $M = \exp(n R_n)$, we obtain for all $n \geq N_1$,

$$\log \left( \frac{1}{P^*_e(n, R_n)} \right) \leq \min_{\sigma \in S(B)} \left( \log \frac{1}{\rho_{XB}} \right) \rho_{XB} \sigma^*_{R_n} \leq - \log \frac{1}{\alpha} (R_n \rho_{XB} \otimes \sigma^*_{R_n} \otimes \sigma^*_{R_n}).$$

Next, we verify that we are able to employ Eq. (41) in Proposition V.2 to asymptotically expand Eq. (89). Equation (93) in Proposition VII.1 below shows that

$$\lim_{n \to \infty} \sigma^*_{R_n} = \lim_{n \to \infty} \frac{\text{Tr} \left( \frac{\sigma_{R_n}}{\rho_{XB}} \right) \frac{1}{\sigma_{R_n}} \sigma_{R_n}}{\text{Tr} \left( \frac{\sigma_{R_n}}{\rho_{XB}} \right) \frac{1}{\sigma_{R_n}} \sigma_{R_n}} = \rho_B.$$
Since \( V = V(\rho_{XB}\|\mathbb{1}_X \otimes \rho_B) > 0 \), by the continuity of \( V(\cdot \| \cdot) \) (c.f. (13)), for every \( \kappa \in (0, 1) \) there exists \( N_2 \in \mathbb{N} \) such that for all \( n \geq N_2 \),
\[
V(\rho_{XB}\|\tau_X \otimes \sigma^*_R) = V(\rho_{XB}\|\mathbb{1}_X \otimes \sigma^*_R) \geq (1 - \kappa)V =: \nu > 0.
\]

Hence, we apply Eq.(41) in Proposition V.2 with \( r_n = \log |X| - R_n, \rho = \rho_{XB} \) and \( \sigma = \tau_X \otimes \sigma^*_R \) to obtain for all \( n \geq \max\{N_1, N_2\} \),
\[
- \log \tilde{a}_{exp(-nr_n)}(\rho^\otimes n\|\sigma^\otimes n) \leq n \sup_{\alpha \in \mathbb{R}} \frac{1 - \alpha}{\alpha} (D_\alpha(\rho\|\sigma) - r_n + \gamma_n) + \log(s_n^* \sqrt{n}) + K,
\]
\[
= nE_{sp}(H + a_n + \gamma_n) + \log(s_n^* \sqrt{n}) + K, \tag{90}
\]

for some constant \( K > 0 \), and \( s_n^* := (1 - \alpha(r_n))/\alpha(r_n) \). Now, let \( \delta_n := a_n + \gamma_n \), and notice that \( \gamma_n = O(\log n/n) = o(a_n) \). We invoke Proposition VII.1 below to have
\[
\limsup_{n \to \infty} E_{sp}(H(X|B)_\rho + \delta_n) = \limsup_{n \to \infty} E_{sp}(H(X|B)_\rho + \delta_n) \leq \frac{1}{2V}. \tag{91}
\]

Moreover, Eq. (93) in Proposition VII.1 below gives that \( \lim_{n \to \infty} \frac{1}{na_n^2} = 1/V \). Combining Eqs. (8), (89), (90) and (91) to conclude our claim
\[
\limsup_{n \to \infty} \frac{1}{na_n^2} \log \left( \frac{1}{P^*_n(n, R_n)} \right) \leq \limsup_{n \to \infty} \frac{-\log \delta_n}{na_n^2} - \frac{1}{na_n^2} \log \left( \frac{1}{P^*_n(n, R_n)} \right) \leq \frac{1}{2V} + \limsup_{n \to \infty} \frac{\log(s_n^* \sqrt{n})}{na_n^2} \leq \frac{1}{2V} + \limsup_{n \to \infty} \frac{\log(s_n^* \sqrt{n})}{na_n^2} \leq \frac{1}{2V}, \tag{92}
\]
where the last line follows from \( \lim_{n \to \infty} n\delta_n^2 = +\infty \). Hence, Eq (88) together with Eq. (92) completes the proof.

**Proposition VII.1 (Error Exponent Around Conditional Entropy):** Let \( (\delta_n)_{n \in \mathbb{N}} \) be a sequence of positive numbers with \( \lim_{n \to \infty} \delta_n = 0 \). The following hold:
\[
\limsup_{n \to \infty} E_{sp}(H(X|B)_\rho + \delta_n) \leq \frac{1}{2V(X|B)_\rho};
\]
\[
\limsup_{n \to \infty} \frac{s_n^*}{\delta_n} = \frac{1}{V(X|B)_\rho}; \tag{93}
\]

where
\[
s_n^* := \arg \max_{s \geq 0} \left\{ s \left( H(X|B)_\rho + \delta_n \right) - sH(X|B)_\rho \right\}.
\]

The proof of Proposition VII.1 is provided in Appendix.

**Remark VII.1:** Following the same approach as in the proofs of the achievable bounds given in Theorem 6, Theorem 3, and Proposition IV.4, one can obtain the optimality part of the moderate deviations when the compression rate approaches \( H(X|B)_\rho \) from below. Namely,
\[
\lim_{n \to \infty} \frac{1}{na_n^2} \log [1 - P^*_n(n, H(X|B)_\rho - a_n)] \leq -\frac{1}{2V(X|B)_\rho}.
\]

However, to establish the moderate deviations for the other direction, one would need a refined analysis of the matching bound, Theorem 4, i.e. we need the high-order terms of the form like the result presented in Theorem 2. We note that such bound is missing even in the classical case.

**B. Optimal Rate When the Error Approaches Zero Moderately Quickly**

Theorem 9 of [22] provides bounds on \( R^*(1, \varepsilon) \). By applying these bounds to \( \rho_{XB}^\otimes n \) and slightly reformulating them, we find that for any \( n \in \mathbb{N} \) and any \( \alpha \in (0, 1) \), we have
\[
-\frac{1}{n} D_H^c(\rho_{XB}^\otimes n(\mathbb{1}_X \otimes \rho_B)^\otimes n) \leq -\frac{1}{n} R^*(n, \varepsilon)
\]
\[
\leq \frac{1}{n} D_H^c(\rho_{XB}^\otimes n(\mathbb{1}_X \otimes \rho_B)^\otimes n) + \frac{1}{n} \log \left( \frac{8}{(1 - \alpha)^2 \varepsilon} \right). \tag{94}
\]

By combining this result with a moderate-deviations expansion of the hypothesis testing relative entropy developed in [10], we obtain the moderate deviations result for \( R^*(n, \varepsilon) \) given by Theorem 7. Let us recall the statement of the result before proceeding to the proof.

**Theorem 7:** Assume that the c-q state \( \rho_{XB} \) has \( V(X|B)_\rho > 0 \). Then for any sequence \( (a_n)_{n \in \mathbb{N}} \) satisfying Eq. (8), and \( \varepsilon_n := e^{-na_n^2} \), we have the asymptotic expansion
\[
R^*(n, \varepsilon_n) = H(X|B)_\rho + \sqrt{2V(X|B)_\rho} a_n + o(a_n). \tag{10}
\]

This result relies heavily on the following expansion of the hypothesis testing relative entropy.

**Proposition VII.2 (Theorem 1 of [10]):** For any moderate sequence \( a_n \) and \( \varepsilon_n := e^{-na_n^2} \), and quantum states \( \rho \) and \( \sigma \) with \( \rho \ll \sigma \), we have
\[
\frac{1}{n} D_H^c(\rho^\otimes n\|\sigma^\otimes n) = D(\rho\|\sigma) - 2\sqrt{V(\rho\|\sigma)} a_n + o(a_n). \tag{95}
\]

**Proof of Theorem 7:** We may extend Proposition VII.2 to unnormalized \( \sigma \geq 0 \) simply by factoring out the trace of \( \sigma \) from the second slot of the hypothesis testing relative entropy using that
\[
D_H^c(\rho\|\lambda \sigma) = D_H^c(\rho\|\sigma) - \log \lambda;
\]
\[
D(\rho\|\lambda \sigma) = D(\rho\|\sigma) - \log \lambda,
\]
and
\[
V(\rho\|\lambda \sigma) = V(\rho\|\sigma)
\]
for \( \lambda > 0 \). Therefore,
\[
\frac{1}{n} D_H^c(\rho^\otimes n(\lambda \sigma)^\otimes n) = \frac{1}{n} D_H^c(\rho^\otimes n\|\sigma^\otimes n) - \log \lambda
\]
\[
= D(\rho\|\sigma) - 2\sqrt{V(\rho\|\sigma)} a_n + o(a_n) - \log \lambda.
\]
\[
= D(\rho\|\lambda \sigma) - 2\sqrt{V(\rho\|\lambda \sigma)} a_n + o(a_n)
\]
and thus the relation (95) holds for unnormalized \( \sigma \geq 0 \).
Next, we consider (94) with $\varepsilon = \varepsilon_n := e^{-na_n^2}$, yielding

$$\frac{1}{n} D^n_{\alpha}(\rho_{XB}^{\otimes n}(1_X \otimes \rho_B)^{\otimes n}) \leq \frac{1}{n} \log R^*(n, \varepsilon) \leq \frac{1}{n} D^n_{\alpha}(\rho_{XB}^{\otimes n}(1_X \otimes \rho_B)^{\otimes n}) + \frac{1}{n} \log \frac{8}{(1 - \alpha)^2} \varepsilon_n.$$ (96)

We next need to apply Proposition VII.2 to the hypothesis testing relative entropy on each side. While this application on the left-hand side is immediate, for the right-hand side, we need to check that $(b_n)$ satisfying

$$\alpha \varepsilon_n = e^{-nb_n^2}$$

is a moderate sequence. We define $b_n := \sqrt{a_n^2 + \frac{1}{\alpha} \log \frac{1}{\alpha}}$. Since for any $x, y > 0$ we have

$$\sqrt{x + y} \leq \sqrt{x} + \sqrt{y},$$

we therefore obtain

$$0 \leq b_n \leq a_n + \frac{1}{\sqrt{n}} \log \frac{1}{\alpha} \xrightarrow{n \to \infty} 0$$

by taking $x = a_n^2$ and $y = \frac{1}{n} \log \frac{1}{\alpha}$ in (97). Since $nb_n^2 = na_n^2 + \log \frac{1}{\alpha} \xrightarrow{n \to \infty} \infty$, the sequence $(b_n)$ is indeed moderate. Thus, Proposition VII.2 yields

$$\frac{1}{n} D^n_{\alpha}(\rho_{XB}^{\otimes n}(1_X \otimes \rho_B)^{\otimes n}) = D(\rho_{XB} \| 1_X \otimes \rho_B) - \sqrt{2V(\rho_{XB} \| 1_X \otimes \rho_B)b_n + o(b_n)}.$$ (98)

Now, we have that

$$b_n - a_n \leq \frac{1}{\sqrt{n}} \log \frac{1}{\alpha} = o(a_n)$$

since for any $\delta > 0$,

$$\frac{1}{\sqrt{n}} \log \frac{1}{\alpha} \leq \delta a_n \iff \sqrt{n}a_n \leq \log \frac{1}{\alpha}$$

which occurs for all $n$ sufficiently large because $\sqrt{n}a_n \to \infty$.

Moreover, if $f_n = o(b_n)$, since $b_n - a_n = o(a_n)$, we have $f_n = o(a_n)$. Therefore, the error terms $f_n$ hidden in the $o(b_n)$ of (98) are in fact $o(a_n)$. Moreover, we may write

$$\sqrt{2Vb_n} = \sqrt{2V \alpha_n} + \sqrt{2V(b_n - a_n)} = \sqrt{2V \alpha_n} + o(a_n)$$

with $V := V(\rho_{XB} \| 1_X \otimes \rho_B)$. Thus, (98) yields

$$\frac{1}{n} D^n_{\alpha}(\rho_{XB}^{\otimes n}(1_X \otimes \rho_B)^{\otimes n}) = D(\rho_{XB} \| 1_X \otimes \rho_B) - \sqrt{2V \alpha_n} + o(a_n)$$

for any $\alpha \in (0, 1)$.

The second term of the right-hand side of (96) is

$$\frac{1}{n} \log \frac{8}{(1 - \alpha)^2} \varepsilon_n = \frac{1}{n} \log \frac{8}{(1 - \alpha)^2} - \frac{1}{n} \log \varepsilon = \frac{1}{n} \log \frac{8}{(1 - \alpha)^2} + a_n^2.$$ Since both $\frac{1}{n} = o(a_n^2)$ and $a_n^2 = o(a_n)$, the second term on the right-hand side of (96) is $o(a_n)$. Thus, we may conclude

$$\frac{1}{n} \log R^*(n, \varepsilon) \leq -D(\rho_{XB} \| 1_X \otimes \rho_B) + \sqrt{2V(\rho_{XB} \| 1_X \otimes \rho_B)}a_n + o(a_n).$$

This precisely matches the bound obtained by applying Proposition VII.2 to the left-hand side of (96), and therefore we obtain (10).

Remark VII.2: We established Theorem 6 using a strong large deviation technique (see Section V and [9]), whereas Theorem 7 was proven via a small deviation technique similar to that used in [10]. We note that these two results are independent and do not directly imply each other. On the other hand, to derive one result from the other, one would need an additional Taylor’s expansion on maps $R \mapsto P^R_v(n, R)$ and $\varepsilon \mapsto R^*(n, \varepsilon)$. This is highly non-trivial, however.

VIII. DISCUSSION

In this paper, we study the CQSW protocol, which is the task of classical data compression with quantum side information associated to a c-q state $\rho_{XB}$, for which the asymptotic data compression limit was shown to be $H(X | B)_\rho$ [8]. We focus primarily on the non-asymptotic (i.e. finite $n$) scenario, and obtain results for both the large and moderate deviation regimes. In the large deviation regime, the compression rate $R$ is fixed. We derive lower and upper bounds on the error exponent function for the range $R > H(X | B)_\rho$ (Theorems 1 and 2), and corresponding bounds for the strong converse exponent for the range $R < H(X | B)_\rho$ (Theorems 3 and 4). Comparing the finite blocklength lower bound on the strong converse exponent given in Theorem 3 with the bounds given given by Equation (6.19) in Theorem 6.2 of [27] and Equation (8.6) of Section 8.1.3 of [26] remains an open question.

In addition, we analyze two scenarios in the moderate deviation regime. In the first, the rate depends on $n$ and slowly decays to the $H(X | B)_\rho$ from above, and we characterize the speed of convergence of the optimal error probability to zero in terms of the conditional entropy variance (Theorem 6). In the second, we obtain an expansion for the minimal rate possible to accomplish the CQSW protocol when the error probability is less than a threshold value which decays slowly with $n$ (Theorem 7).

There seems to be an interesting duality between our results on the error exponents for the CQSW protocol and those for classical-quantum channel coding [14], [16], [45], [46]. In the former, the entropic error exponent functions arising in our bounds involve the difference between the compression rate and a conditional Rényi entropy, while in the latter they involve the difference between the Rényi capacity and the transmission rate. The above duality mimics the connections found between the tasks of Slepian Wolf coding and classical channel coding [65]–[67]. We summarize such connections in Table II below.

Besides investigating this duality in detail [17], other open problems include extending variable-rate Slepian–Wolf
coding [12], [29], [30], [68], [69], and the universal coding scenario [4], [12], [31], [61] to the CQSW setting.

**APPENDIX**

**Proposition IV.2 (Properties of $\alpha$-Rényi Conditional Entropy):** Given any classical-quantum state $\rho_{XB} \in \mathcal{S}(XB)$, the following holds:

(a) The map $\alpha \mapsto H_1^\alpha(X|B)_{\rho}$ is continuous and monotonically decreasing on $[0, 1]$.

(b) The map $\alpha \mapsto \frac{1 - \alpha}{\alpha} H_1^\alpha(X|B)_{\rho}$ is strictly concave on $(0, 1)$.

**Proof of Proposition IV.2:**

a Fix an arbitrary sequence $(\alpha_k)_{k \in \mathbb{N}}$ such that $\alpha_k \in [0, 1]$ and $\lim_{k \to +\infty} \alpha_k = \alpha_{\infty} \in [0, 1]$. Let

$$\sigma^*_k \in \text{arg min}_{\sigma \in \mathcal{S}(H)} D_{\alpha_k}(\rho_{XB} \| I_X \otimes \sigma), \quad \forall k \in \mathbb{N} \cup \{+\infty\}.$$ 

The definition in Eq. (15) implies that

$$\limsup_{k \to +\infty} H_1^{\alpha_k}(X|B)_{\rho} = -\liminf_{k \to +\infty} D_{\alpha_k}(\rho_{XB} \| I_X \otimes \sigma^*_k) \leq -D_{\alpha_{\infty}}(\rho_{XB} \| I_X \otimes (\lim_{k \to +\infty} \sigma^*_k)) \leq -\min_{\sigma \in \mathcal{S}(H)} D_{\alpha_{\infty}}(\rho_{XB} \| I_X \otimes \sigma) = H_1^{\alpha_{\infty}}(X|B)_{\rho},$$

where, in order to establish (99), we used the lower semi-continuity of the map $\sigma \mapsto D_{\alpha}(\rho_{XB} \| I_X \otimes \sigma)$ (see Eq. (19) in Lemma IV.1) and the continuity of $\alpha \mapsto D_{\alpha}(\rho_{XB} \| I_X \otimes \sigma^*_k)$ (Eq. (18) in Lemma IV.1).

Next, we let

$$\sigma_k := (1 - \varepsilon_k) \sigma^*_\infty + \varepsilon_k \frac{1}{d}, \quad \forall k \in \mathbb{N},$$

where $(\varepsilon_k)_{k \in \mathbb{N}}$ is an arbitrary positive sequence that converges to zero. Then, it follows that

$$\liminf_{k \to +\infty} H_1^{\alpha_k}(X|B)_{\rho} \geq -\limsup_{k \to +\infty} D_{\alpha_k}(\rho_{XB} \| I_X \otimes \sigma_k) = -D_{\alpha_{\infty}}(\rho_{XB} \| I_X \otimes \sigma^*_\infty) \quad (100) = H_1^{\alpha_{\infty}}(X|B)_{\rho}.$$ 

Here, equality (100) holds because $I_X \otimes \sigma_k \ngeq \rho_{XB}$ for all $k \in \mathbb{N} \cup \{+\infty\}$. Thus, the map $(\alpha_k, \sigma_k) \mapsto D_{\alpha_k}(\rho_{XB} \| I_X \otimes \sigma_k)$ is continuous for $k \in \mathbb{N} \cup \{+\infty\}$. Hence, we prove the continuity.

Now, we show the monotonicity. For all $\sigma_B \in \mathcal{S}(B)$, Eq. (19) in Lemma IV.1 implies that $-D_{\alpha}(\rho_{XB} \| I \otimes \sigma_B)$ is monotonically decreasing in $\alpha \geq 0$. Since $H_1^\alpha(X|B)_{\rho}$ is the pointwise supremum of the above function, we conclude that $H_1^\alpha(X|B)_{\rho}$ is monotonically decreasing in $\alpha \geq 0$. Hence, item (a) is proven.

b This proof follows directly from item c in Proposition IV.4, Eq. (30), and the substitution $\alpha = 1/(1 + s)$.

Let us recall Proposition IV.4.

**Proposition IV.4 (Properties of Auxiliary Functions):** Let $\rho_{XB}$ be a classical-quantum state with $H(X|B)_{\rho} > 0$, the auxiliary functions $E_0(s), E_0^1(s)$, and $E_0^2(s)$ admit the following properties.

(a) (Continuity) The function $s \mapsto E_0(s)$ is continuous for all $s \in (-1, +\infty)$; the function $s \mapsto E_0^1(s)$ is continuous for all $s \in [0, +\infty)$; and the function $s \mapsto E_0^2(s)$ is continuous for all $s \in (-1, +\infty)$.

(b) (Negativity) For $s \geq 0$,

$$E_0(s) \leq 0, \quad E_0^1(s) \leq 0; \quad \text{and} \quad E_0^2(s) \leq 0$$

with $E_0(0) = E_0^1(0) = E_0^2(0) = 0$.

(c) (Concavity) All the functions $s \mapsto E_0(s)$, $s \mapsto E_0^1(s)$, and $s \mapsto E_0^2(s)$ are concave in $s$ for all $s \in (-1, +\infty)$. 

}\[\begin{array}{|c|c|c|}
\hline
\text{Bounds/Settings} & \text{Slepian–Wolf Coding with Quantum Side Information} & \text{Classical–Quantum Channel Coding} \\
\hline
\text{Achievability} (R < C_W \text{ or } R > H(X|B)_{\rho}) & \mathcal{E}^\alpha_c(R) := \sup_{\frac{1}{\alpha} \leq k \leq 1} \frac{1 - \alpha}{\alpha} \left(R - H^\alpha_1(X|B)_{\rho}\right) & \mathcal{E}^\alpha_c(R) := \sup_{\frac{1}{\alpha} \leq k \leq 1} \frac{1 - \alpha}{\alpha} \left(\max_{\rho \in \mathcal{P}(X)} I^\alpha_1(\rho, W) - R\right) \\
\hline
\text{Optimality} (R < C_W \text{ or } R > H(X|B)_{\rho}) & \mathcal{E}^\alpha_{\text{opt}}(R) := \sup_{s \geq 0} \{E_0(s) - sR\} & \mathcal{E}^\alpha_{\text{opt}}(R) := \sup_{s \geq 0} \left\{\max_{\rho \in \mathcal{P}(X)} E_0(s, \rho) - sR\right\} \\
\hline
\text{Strong Converse} (R > C_W \text{ or } R < H(X|B)_{\rho}) & \mathcal{E}^\alpha_{\text{sc}}(R) := \sup_{\alpha > 0} \left\{\frac{1 - \alpha}{\alpha} \left(R - H^\alpha_1(X|B)_{\rho}\right)\right\} & \mathcal{E}^\alpha_{\text{sc}}(R) := \sup_{\alpha > 0} \left\{\frac{1 - \alpha}{\alpha} \left(C_{\alpha, W} - R\right)\right\} \\
\hline
\text{Auxiliary Function} & E_0(s) := -\log Tr_B \left[\left(Tr_X(\rho_{XB})\right)^{(1+s)}\right] & E_0(s, \rho) := -\log Tr \left[\left(\sum_{x \in X} P(x) \cdot \rho_2^{(1+s)}\right)^{(1+s)}\right] \\
\hline
\end{array}\]
\[ \frac{\partial E_0^1(s)}{\partial s} \bigg|_{s=0} = \frac{\partial E_0(s)}{\partial s} \bigg|_{s=0} = \frac{\partial^2 E_0^1(s)}{\partial s^2} \bigg|_{s=0} = -H(X|B)_p. \]

(e) (Second-order Derivative)
\[ \frac{\partial^2 E_0^1(s)}{\partial s^2} \bigg|_{s=0} = \frac{\partial^2 E_0(s)}{\partial s^2} \bigg|_{s=0} = -V(X|B)_p. \]

Proof of Proposition IV.4:

(a) (Continuity) Since \( E_0(s) \) admits a closed-form
\[ -\log \text{Tr} \left[ \left( \text{Tr}_X \rho_{XB}^{1+s} \right) \right], \quad \forall s > -1. \]

It is clearly continuous for all \( s > -1 \).

Likewise, since \( E_0^1(s) = -\log \text{Tr} \left[ \rho_{XB}^{1+s} (I_X \otimes \rho_B)^s \right] \), it is continuous for all \( s \geq 0 \).

The continuity of \( s \mapsto E_0^1(s) \) for \( s \in (-1, 1) \) follows from similar techniques by Tomamichel and Hayashi [63, Proposition 11].

(b) (Negativity) The negativity of \( E_0(s), E_0^1(s), \) and \( E_0^2(s) \) directly follows from the non-negativity of the conditional Rényi entropy and the definition, Eqs. (30) and (33).

(c) (Concavity) For \( s \mapsto E_0^1(s) \), the claim follows from the concavity of the map \( s \mapsto sD_{1-s}(\cdot, \cdot) \), Eq. (17) in Lemma IV.1.

Let us now consider \( s \mapsto E_0(s) \). The concavity for \( s \geq 0 \) can be proved with the geometric matrix means in [46]. Here, we present another proof by the following matrix inequality.

Lemma A.1 ([71, Corollary 3.6]): Let \( A_i \) be \( m \times m \) positive semi-definite matrix and \( Z_i \) be \( n \times m \) matrix for \( i = 1, \ldots, k \). Then, for all unitarily invariant norms \( \| \cdot \| \) and \( \gamma > 0 \), the map
\[ (p, t) \mapsto \left\| \left( \sum_{i=1}^{k} Z_i^* A_i^{1/p} Z_i \right)^{\gamma p} \right\| \]
is jointly log-convex on \( (0, +\infty) \times (-\infty, +\infty) \).

Let \( \rho_{XB} = \sum_{x \in X} p(x)|x\rangle\langle x| \otimes W_x \), \( t = \gamma = 1, i = x, k = |X|, A_i = p(x)W_x \), and \( Z_i = I_n \otimes m \). We obtain the log-convexity of the map by applying Lemma A.1:
\[ p \mapsto \text{Tr} \left( \sum_{x \in X} (p(x)W_x)^{1/s} \right)^p, \quad \forall p > 0, \]

which is exactly the concavity of the map \( s \mapsto E_0(s) \) for all \( s > -1 \).

The concavity of \( s \mapsto E_0^1(s) \) on \( s > -1 \) can be obtained using the technique as in [48].

(d) (First-order Derivative) By the definitions of \( E_0(s) \) and \( E_0^1(s) \),
\[ \frac{\partial E_0(s)}{\partial s} \bigg|_{s=0} = -H^1(X|B)_p - s \frac{\partial H^1(X|B)_p}{\partial s} \bigg|_{s=0} = -H(X|B)_p; \]
\[ \frac{\partial E_0^1(s)}{\partial s} \bigg|_{s=0} = -H^1(X|B)_p - s \frac{\partial H^1(X|B)_p}{\partial s} \bigg|_{s=0} = -H(X|B)_p \]

Likewise, one can verify that
\[ \frac{\partial E_0^1(s, \rho_{XB})}{\partial s} \bigg|_{s=0} = D_{1-s}(\rho_{XB}||I_X \otimes \rho_B) - sD_{1-s}(\rho_{XB}||I_X \otimes \rho_B) \bigg|_{s=0} = D(\rho_{XB}||I_X \otimes \rho_B) = -H(X|B)_p. \]

(e) (Second-order derivative) We first consider \( E_0(s) \).

Similar to Item d, it follows that
\[ \frac{\partial^2 E_0(s)}{\partial s^2} \bigg|_{s=0} = -2 \frac{\partial H^1(X|B)_p}{\partial s} - \frac{\partial^2 H^1(X|B)_p}{\partial s^2} \bigg|_{s=0}. \]

The above equation indicates that we need to evaluate the first-order derivative of \( H^1(X|B)_p \) at 0. In the following, we directly deal with the closed-form expression, Eq. (31).

To ease the burden of derivations, we denote some notation:
\[ f(s) := \text{Tr}_X \rho_{XB}^{1/(1+s)}, \]
\[ g(s) := f(s)^{1+s}, \]
\[ F(s) := \text{Tr}[g(s)], \]

Then,
\[ \frac{\partial E_0(s)}{\partial s} = -\frac{F'(s)}{F(s)}, \]
\[ \frac{\partial^2 E_0(s)}{\partial s^2} = -\frac{F''(s)}{F(s)} - \left( \frac{\partial E_0(s)}{\partial s} \right)^2. \]

Direct calculation shows that
\[ f'(s) = -\frac{1}{(1+s)^2} \text{Tr}_X \left[ \rho_{XB}^{1/(1+s)} \log \rho_{XB} \right], \]
\[ f''(s) = \frac{1}{(1+s)^3} \text{Tr}_X \left[ \rho_{XB} \log \rho_{XB} \cdot \left( 2 + \frac{\log \rho_{XB}}{(1+s)} \right) \right]. \]
Note that\(^\text{7}\) \(g(s) = e^{(1+s) \log f(s)}\). By applying the chain rule of the Fréchet derivatives, one can show
\[
g'(s) = \nabla \exp[\log g(s)] (1+s) \nabla \log f(s) (f'(s) + \log f(s)) = (1+s) \nabla \log f(s) (f'(s) + \log f(s)). \tag{108}\]

Moreover, Eq. (101) gives
\[
\frac{\partial H^1 s(X|B)\rho}{\partial \alpha} \bigg|_{\alpha=0} = -\frac{1}{2} V(X|B)\rho.
\]

For \(E_0\), continuing from item d, one obtains
\[
\frac{\partial^2 E_0(s)}{\partial s^2} \bigg|_{s=0} = -2 D'_1 -(\rho_{X|B}\| I_X \otimes \rho_B) + s D''_1 -(\rho_{X|B}\| I_X \otimes \rho_B)\bigg|_{s=0} = V(\rho_{X|B}| I_X \otimes \rho_B) \tag{117}\]

where in equality (117) we use the fact \(D'_1 (\cdot (\cdot)\| I_X \otimes \rho_B = V(\cdot | (\cdot))/2 [72, \text{Theorem 2}]\).

The second-order derivative of \(E_0\) follows from similar techniques by Tomamichel and Hayashi [63, Proposition 11].

**Lemma A.2 (\[73, \text{Theorem 3.23}\]):** Let \(A, X\) be \(d \times d\) Hermitian matrices, and \(t \in \mathbb{R}\). Assume \(f : I \to \mathbb{R}\) is a continuously differentiable function. Then
\[
\frac{d}{dt} \nabla f(A + t X) \bigg|_{t = t_0} = \nabla f'(A + t_0 X). \tag{119}\]

**Proposition IV.5 (Properties of the Exponent Function):** Let \(\rho_{X|B}\) be a classical-quantum state with \(H(X|B)\rho > 0\), the following holds.

\(\text{(a)}\) \(E_{\varphi}(\cdot)\) is convex, differentiable, and monotonically increasing on \([0, +\infty]\).
\[
E_{\varphi}(R) = \begin{cases} 0, & R \leq H^1 s(X|Y)\rho \\ E_{\varphi}(R), & H^1 s(X|Y)\rho \leq R \leq H^1 s(1/2)(X|Y)\rho \\ +\infty, & R > H^1 s(1/2)(X|Y)\rho \end{cases} \tag{118}\]

\(\text{(b)}\) Define
\[
F_R(\alpha, \sigma_B) := \frac{1 - \alpha}{\alpha} (R + D_\alpha (\rho_{X|B}\| I_X \otimes \sigma_B)) \tag{119}\]

on \((0,1) \times S(B), \) and \(F_R(1, \sigma_B) = 0\) for all \(\sigma_B \in S(B).\) For \(R \in (H^1 s(1/2)(X|Y)\rho, H^1 s(1/2)(X|Y)\rho)\), there exists a unique saddle-point \((\alpha^*, \sigma^*) \in (0,1) \times S(B)\) of \(F_R(\cdot, \cdot)\) such that
\[
F_R(\alpha^*, \sigma^*) = \sup_{\alpha \in [0,1]} \inf_{\sigma_B \in S(B)} \ F_R(\alpha, \sigma_B) = \inf_{\sigma_B \in S(B)} \sup_{\alpha \in [0,1]} \ F_R(\alpha, \sigma_B) = E_{\varphi}(R). \tag{120}\]

\(\text{(c)}\) Any saddle-point \((\alpha^*, \sigma^*)\) of \(F_R(\cdot, \cdot)\) satisfies
\[
1_X \otimes \sigma^* \gg \rho_{X|B} \tag{121}\]

**Proof of Proposition IV.5:**

\(\text{(a)}\) Item a in Proposition IV.2 shows that the map \(\alpha \mapsto H^1 s(X|B)\rho\) is monotonically decreasing on \([0,1]\). Hence, from the definition:
\[
E_{\varphi}(R) := \sup_{\alpha \in [0,1]} \frac{1 - \alpha}{\alpha} (R - H^1 s(X|B)\rho), \tag{122}\]
it is not hard to verify that $E_{\text{up}}(R) = +\infty$ for all $R > H_0^1(H|B)_\rho$; finite for all $R < H_0^1(H|B)_\rho$; and $E_{\text{sw}}(R) = 0$, for all $R \geq H_0^1(H|B)_\rho$. Moreover, $E_{\text{up}}(R) = E_{\text{sw}}(R)$ for $R \in [H_0^1(X|Y)_\rho, H_1^1/2(X|Y)_\rho]$ by the definition in Eq. (28).
For every $\alpha \in (0, 1]$, the function $\frac{1}{\alpha} \left( H_0^1 H_0^1 \right)$ is non-decreasing, convex, and continuous function in $R \in \mathbb{R}_>0$. Since $E_{\text{up}}(R)$ is the pointwise supremum of the above functions, $E_{\text{up}}(R)$ is non-decreasing, convex, and lower semi-continuous for all $R \geq 0$. Furthermore, since a convex function is continuous on the interior of the interval if it is finite [74, Corollary 6.3.3], thus $E_{\text{up}}(R)$ is continuous for all $R \geq H_0^1(H|B)_\rho$, and continuous from the left at $R = H_0^1(H|B)_\rho$.

b Before commencing the proof, we remark that the saddle-point property in Proposition (IV.5)-b is a stronger result than the min-max identity in (34). In other words, the identity that the infimum and supremum is exchangeable does not guarantee the existence of the saddle points. To show such existence, we need to check if the underlying objective function satisfies the so-called saddle-element (see Definition A.3 below). We explain this in the following.

Let

$$ S_\rho(B) := \{ \sigma_B \in S(B) : \rho_{XB} \not\perp X \otimes \sigma_B \}. $$

Fix an arbitrary $R \in (H_0^1(X|B)_\rho, H_0^1(H|B)_\rho)$. In the following, we first prove the existence of a saddle-point of $F_R(\cdot, \cdot)$ on $(0, 1) \times S_\rho(B)$. Ref. [75, Lemma 36.2] states that $(\alpha^*, \sigma^*)$ is a saddle point of $F_R(\cdot, \cdot)$ if and only if the supremum in

$$ \sup_{\alpha \in (0, 1]} \inf_{\sigma \in S_\rho(B)} F_R(\alpha, \sigma) $$

is attained at $\alpha^* \in (0, 1]$, the infimum in

$$ \inf_{\sigma \in S_\rho(B)} \sup_{\alpha \in (0, 1]} F_R(\alpha, \sigma) $$

is attained at $\sigma^* \in S_\rho(B)$, and the two extrema in Eqs. (119), (120) are equal and finite. We first claim that, for all $\alpha \in (0, 1]$, $E_{\text{up}}(R)$ equals finite.

To see this, observe that for any $\alpha \in (0, 1)$, Eqs. (11) yield

$$ \forall \sigma \in S(B) \backslash S_\rho(B), \quad D_\alpha(\rho_{XB} \| I_X \otimes \sigma) = +\infty, $$

which, in turn, implies

$$ \forall \sigma \in S(B) \backslash S_\rho(B), \quad F_R(\alpha, \sigma) = +\infty. $$

Further, Eq. (121) holds trivially when $\alpha = 1$. Hence, Eq. (121) yields

$$ \sup_{\alpha \in (0, 1]} \inf_{\sigma \in S_\rho(B)} F_R(\alpha, \sigma) = \sup_{\alpha \in (0, 1]} \inf_{\sigma \in S(B)} F_R(\alpha, \sigma) $$

Owing to the fact $R < H_0^1(X|B)_\rho$, and Eq. (29), we have

$$ \text{E}_{\text{up}}(R) = \sup_{\alpha \in (0, 1]} \inf_{\sigma \in S(B)} F_R(\alpha, \sigma) < +\infty, $$

which guarantees the supremum in the right-hand side of Eq. (124) is attained at some $\alpha \in (0, 1]$. Namely, there exists some $\alpha_R \in (0, 1]$ such that

$$ \sup_{\alpha \in (0, 1]} \inf_{\sigma \in S_\rho(B)} F_R(\alpha, \sigma) = \max_{\alpha \in [\alpha_R, 1]} \inf_{\sigma \in S(B)} F_R(\alpha, \sigma) < +\infty. $$

Thus, we complete our claim in Eq. (119). It remains to show that the infimum in Eq.(120) is attained at some $\sigma^* \in S_\rho(B)$ and the supremum and infimum are exchangeable. To achieve this, we will show that $([\alpha_R, 1], S_\rho(B), F_R)$ is a closed saddle-element (see Definition A.3 below) and employ the boundedness of $[\alpha_R, 1] \times S_\rho(B)$ to conclude our claim.

Definition A.3 (Closed Saddle-Element [75]): We denote by $\rho \in \mathcal{S}$ the relative interior and the closure of a set, respectively. Let $A, B$ be subsets of a real vector space, and $F : A \times B \rightarrow \mathbb{R} \cup \{\pm\infty\}$. The triple $(A, B, F)$ is called a closed saddle-element if for any $x \in \rho \in \mathcal{S}(\rho)$ (resp. $y \in \rho \in \mathcal{S}(\rho)$),

(i) $B$ (resp. $A$) is convex.

(ii) $F(x, \cdot)$ (resp. $F(\cdot, y)$) is convex (resp. concave) and lower (resp. upper) semi-continuous.

(iii) Any accumulation point of $B$ (resp. $A$) that does not belong to $\rho \in \mathcal{S} \rho \in \mathcal{S}$, say $y_0$ (resp. $x_0$) satisfies $\lim_{x \rightarrow x_0} F(x, y) = +\infty$ (resp. $\lim_{x \rightarrow x_0} F(x, y) = -\infty$).

Fix an arbitrary $\alpha \in \rho \in \mathcal{S}(\rho)$ and $y_0$ (resp. $x_0$) satisfies $\lim_{x \rightarrow x_0} F(x, y) = +\infty$ (resp. $\lim_{x \rightarrow x_0} F(x, y) = -\infty$).
Now we verify that $(\bar{\alpha}_R, 1, F_R(\cdot, \sigma))$ satisfies the three items in Definition A.3. Fix an arbitrary $\sigma \in \mathcal{S}_\rho(B)$. (i) The set $(0, 1)$ is obviously convex. (ii) From Eq. (18) in Lemma IV.1, the map $\alpha \mapsto F_R(\alpha, \sigma)$ is continuous on $(0, 1)$. Further, it is not hard to verify that $F_R(1, \sigma) = 0 = \lim_{\alpha \searrow 1} F_R(\alpha, \sigma)$ from Eqs. (129), (118), and (11). Item b in Proposition IV.2 implies that $\alpha \mapsto F_R(\alpha, \sigma)$ on $[\bar{\alpha}_R, 1)$ is concave. Moreover, the continuity of $\alpha \mapsto F_R(\alpha, \sigma)$ on $[\bar{\alpha}_R, 1]$ guarantees the concavity of $\alpha \mapsto F_R(\alpha, \sigma)$ on $[\bar{\alpha}_R, 1)$. (iii) Since $[\bar{\alpha}_R, 1]$ is closed, there is no accumulation point of $[\bar{\alpha}_R, 1]$ that does not belong to $[\bar{\alpha}_R, 1]$.

We are at the position to prove the saddle-point property. The closed saddle-element, along with the boundedness of $\mathcal{S}_\rho(B)$ and Rockafellar’s saddle-point result [75, Theorem 8], [76, Theorem 37.3] imply that

$$
-\infty < \inf_{\alpha \in [\bar{\alpha}_R, 1], \sigma \in \mathcal{S}_\rho(B)} F_R(s, \sigma) = \min_{\sigma \in \mathcal{S}_\rho(B)} \sup_{\alpha \in [\bar{\alpha}_R, 1]} F_R(s, \sigma). \tag{130}
$$

Then Eqs. (125) and (130) lead to the existence of a saddle-point of $F_R(\cdot, \cdot)$ on $(0, 1) \times \mathcal{S}_\rho(B)$.

Next, we prove the uniqueness. The rate $R$ and item a in Proposition IV.5 shows that

$$
\sup_{0 < \alpha \leq \bar{\alpha}} \min_{\sigma \in \mathcal{S}(B)} F_R(\alpha, \sigma) \in \mathbb{R}_{\geq 0}. \tag{131}
$$

Note that $\alpha^* = 1$ will not be a saddle point of $F_R(\cdot, \cdot)$ because $F_R(1, \sigma) = 0$, $\forall \sigma \in \mathcal{S}(B)$, contradicting Eq. (131).

Now, fix $\alpha^* \in (0, 1)$ to be a saddle-point of $F_R(\cdot, \cdot)$. Eq. (19) in Lemma IV.1 implies that the map $\sigma \mapsto D_{\alpha^*}(\rho_{XB} \| \mathbf{1}_X \otimes \sigma)$ is strictly convex, and thus the minimizer of Eq. (131) is unique. Next, let $\sigma^* \in \mathcal{S}_\rho(B)$ be a saddle-point of $F_R(\cdot, \cdot)$. Then,

$$
F_R(\alpha, \sigma^*) = \frac{1 - \alpha}{\alpha} (R - H_0(\mathbf{1}_B)_\rho).
$$

Item b in Proposition IV.2 then shows that $\frac{1 - \alpha}{\alpha} H_0(\mathbf{1}_B)_\rho$ is strictly concave on $(0, 1)$, which in turn implies that $F_R(\cdot, \cdot)$ is also strictly concave on $(0, 1)$. Hence, the maximizer of Eq. (131) is unique, which completes item b of Proposition IV.5.

As shown in the proof of item b, $\alpha^* = 1$ is not a saddle point of $F_R(\cdot, \cdot)$ for any $R < H_0(\mathbf{1}_B)_\rho$. We assume $(\alpha^*, \sigma^*)$ is a saddle-point of $F_R(\cdot, \cdot)$ with $\alpha^* \in (0, 1)$, it holds that

$$
F_R(\alpha^*, \sigma^*) = \min_{\sigma \in \mathcal{S}(B)} F_R(\alpha^*, \sigma) = \frac{1 - \alpha^*}{\alpha^*} R + \frac{1 - \alpha^*}{\alpha^*} \min_{\sigma \in \mathcal{S}(B)} D_{\alpha^*}(\rho_{XB} \| \mathbf{1}_X \otimes \sigma). \tag{132}
$$

It is known [47], [49, Lemma 1], [25, Lemma 5.1] that the minimizer of Eq. (132) is

$$
\sigma^* = \left( \frac{\text{Tr}_{XB}}{\text{Tr}[(\text{Tr}_{XB})]} \right)^{\frac{1}{\alpha^*}}.
$$

From this expression, it is clear that $\mathbf{1}_X \otimes \sigma^* \gg \rho_{XB}$, and thus item c is proved.

**Proof of Proposition IV.6:** We only provide the proof for Eq (37) since Eqs. (35) and (36) follow similarly.

Starting with the left-hand side,

$$
\sup_{-1 < s < 0} \left\{ E_\rho(s, \rho_{XB}) + sR \right\} = \sup_{-1 < s < 0} \left\{ -s H_{\rho_{XB}}^{\perp}(\mathbf{1}_B)_{\rho} + sR \right\} = \sup_{-1 < s < 0} \left\{ -s \max_{s \in \mathbb{R}} -D_{\rho}(\rho_{XB} \| \mathbf{1}_X \otimes \tau_B) + sR \right\} = \sup_{-1 < s < 0} \left\{ sD_{\rho}(\rho_{XB} \| \mathbf{1}_X \otimes \tau_B) + sR \right\}.
$$

Next, Lemma A.5 gives the variational representation

$$
D_{\rho}(\rho_{XB} \| \mathbf{1}_X \otimes \tau_B) = \max_{\sigma \in \mathcal{S}(B)} \left[ D(\sigma_{XB} \| \mathbf{1}_X \otimes \tau_B) + \frac{1}{s} D(\sigma_{XB} \| \rho_{XB}) \right]
$$

and shows that the optimizer is a state $\sigma^*_{XB} \in \mathcal{S}_\rho(XB)$. Therefore, we can write

$$
D_{\rho}(\rho_{XB} \| \mathbf{1}_X \otimes \tau_B) = \max_{\sigma \in \mathcal{S}(B)} \left[ D(\sigma_{XB} \| \mathbf{1}_X \otimes \tau_B) + \frac{1}{s} D(\sigma_{XB} \| \rho_{XB}) \right]
$$

Substituting this variational representation into our expansion of $E_{\infty}(R)$, we have

$$
E_{\infty}(R) = \sup_{-1 < s < 0} \min_{\tau_{XB} \in \mathcal{S}(B)} \left\{ s \left[ D(\sigma_{XB} \| \mathbf{1}_X \otimes \tau_B) + \frac{1}{s} D(\sigma_{XB} \| \rho_{XB}) \right] + sR \right\} = \sup_{-1 < s < 0} \min_{\tau_{XB} \in \mathcal{S}(B)} \left\{ D(\sigma_{XB} \| \mathbf{1}_X \otimes \tau_B) + \frac{1}{s} D(\sigma_{XB} \| \rho_{XB}) + R \right\}.
$$

Consider the function

$$
G(\tau_B, \sigma_{XB}) = D(\sigma_{XB} \| \mathbf{1}_X \otimes \tau_B) + \frac{1}{s} D(\sigma_{XB} \| \rho_{XB})
$$

**Lemma A.4 (Properties of G):** Let $s \in (-1, 0)$. As a function of states $\tau_B$ in $\mathcal{S}(B)$ and $\sigma_{XB}$ in $\mathcal{S}(XB)$,

1) $G(\tau_B, \sigma_{XB})$ is convex and lowersemicontinuous in $\tau_B$.
2) $G(\tau_B, \sigma_{XB})$ is concave and continuous in $\sigma_{XB}$.

**Proof:** Since the only dependence in $\tau_B$ is in the second argument of the relative entropy, $G$ is convex and lowersemicontinuous in $\tau_B$. To prove the second part,
we expand $G$ as
\[
G(\tau_B, \sigma_{XB}) = -H(\sigma_{XB}) - \text{Tr}[\sigma_{XB} \log(1_X \otimes \tau_B)]
\]
\[\quad - \frac{1}{s} H(\sigma_{XB}) - \text{Tr}[\sigma_{XB} \log \rho_{XB}] \]
\[= \left(1 + \frac{1}{s}\right) H(\sigma_{XB})
\]
\[\quad - \text{Tr}[\sigma_{XB}(\log(1_X \otimes \tau_B) + \log \rho_{XB})].\]

Since $s \in (-1, 0)$, the coefficient of $H(\sigma_{XB})$ is positive. The second term is linear in $\sigma_{XB}$, so $G$ is concave and continuous in $\sigma_{XB}$. \hfill \square

By these properties of $G$, the compactness and convexity of $S(B)$ and $S_p(B)$, we may apply the min-max theorem given by Lemma 2.7 of [28] to find
\[
\min_{\tau_B} \max_{\sigma_{XB} \in S_p(B, X)} G(\tau_B, \sigma_{XB}) = \max_{\sigma_{XB} \in S_p(B, X)} \min_{\tau_B} G(\tau_B, \sigma_{XB}).
\]

Therefore,
\[
E_{\infty}^\rho(R) = \sup_{-1 < s < 0} \max_{\sigma_{XB} \in S_p(B, X)} \min_{\tau_B} \left\{\frac{1}{s} D(\sigma_{XB}&||\rho_{XB}) + R \right\}
\]
\[\quad = \sup_{-1 < s < 0} \min_{\sigma_{XB} \in S_p(B, X)} \left\{-s \left[\max_{\tau_B} - D(\sigma_{XB}||1_X \otimes \tau_B)\right] + \frac{1}{s} D(\sigma_{XB}||\rho_{XB}) + sR\right\}
\]
\[\quad = \sup_{-1 < s < 0} \min_{\sigma_{XB} \in S_p(B, X)} \left\{-s \left[H(\sigma_{XB}) - D(\sigma_{XB}||\rho_{XB})\right] + sR\right\}
\]
\[\quad = \sup_{0 < s < 1} \min_{\sigma_{XB} \in S_p(B, X)} \left\{\frac{1}{s} D(\sigma_{XB}||\rho_{XB}) + s \left[H(\sigma_{XB}) - D(\sigma_{XB}||\rho_{XB})\right] + sR\right\}
\]
\[\quad \leq \min_{\sigma_{XB} \in S_p(B, X)} \sup_{0 < s < 1} \left\{\frac{1}{s} D(\sigma_{XB}||\rho_{XB}) + s \left[H(\sigma_{XB}) - D(\sigma_{XB}||\rho_{XB})\right] + sR\right\}
\]
\[\quad = \min_{\sigma_{XB} \in S_p(B, X)} \left\{\frac{1}{s} D(\sigma_{XB}||\rho_{XB}) + s \left[H(\sigma_{XB}) - D(\sigma_{XB}||\rho_{XB})\right] + sR\right\}.
\]

Proposition VII.1 (Error Exponent around Conditional Entropy): Let $(\delta_n)_{n \in \mathbb{N}}$ be a sequence of positive numbers with $\lim_{n \to +\infty} \delta_n = 0$. The following hold:
\[
\limsup_{n \to +\infty} \frac{E_{\infty}^\rho (H(X|B)_\rho + \delta_n)}{\delta_n^2} \leq \frac{1}{2V(X|B)_\rho};
\]
\[
\limsup_{n \to +\infty} \frac{s_n^*}{\delta_n} = \frac{1}{V(X|B)_\rho},
\]
where
\[
s_n^* := \arg \max_{s \geq 0} \left\{s \left(H(X|B)_\rho + \delta_n\right) - sH^1(\frac{1}{\delta_n})(X|B)_\rho\right\}.
\]

Proof of Proposition VII.1: For notational convenience, we denote by $H := H(X|B)_\rho$, $V := V(X|B)_\rho$. Thus,
\[
E_{\infty}^\rho (R) = \sup_{s \geq 0} \left\{sR + E_0(s)\right\},
\]
Let a critical rate to be
\[
\tau_c := \frac{\partial E_0(s)}{\partial s} \bigg|_{s=1}.
\]
Let $N_0$ be the smallest integer such that $H(X|B)_\rho + \delta_n < \tau_c$, for all $n \geq N_0$. Since the map $r \mapsto E_{\infty}^\rho (r)$ is non-increasing by item a in Proposition IV.5, the maximization over $s$ in Eq. (135) can be restricted to the set $[0, 1]$ for any rate below $\tau_c$, i.e.,
\[
E_{\infty}^\rho (H + \delta_n) = \max_{0 \leq s \leq 1} \left\{s \left(H + \delta_n\right) + E_0(s)\right\}.
\]
For every $n \in \mathbb{N}$, let $s_n^*$ attain the maxima in Eq. (136) at a rate of $H + \delta_n$. It is not hard to observe that $s_n^*>0$ for all $n \geq N_0$ since $s_n^* = 0$ if and only if $H + \delta_n < H$, which violates the assumption of $\delta_n > 0$ for finite $n$. Now, we will show Eq. (134) and
\[
\lim_{n \to +\infty} s_n^* = 0.
\]
Let $(s_{nk})_{k \in \mathbb{N}}$ be arbitrary subsequences. Since $[0, 1]$ are compact, we may assume that
\[
\lim_{k \to +\infty} s_{nk}^* = s_0,
\]
for some $s_0 \in [0, 1]$. Since $s \mapsto E_0(s)$ is strictly concave from item c in Proposition IV.4, the maximizer $s_n^*$ must satisfy
\[
\frac{\partial E_0(s)}{\partial s} \bigg|_{s=s_n^*} = -(H + \delta_n),
\]
which together with item a in Proposition IV.4 implies
\[
\lim_{k \to +\infty} \frac{\partial E_0(s)}{\partial s} \bigg|_{s=s_{nk}^*} = \frac{\partial E_0(s)}{\partial s} \bigg|_{s=s_0} = -H.
\]
On the other hand, item d in Proposition IV.4 gives
\[
\frac{\partial^2 E_0(s)}{\partial s^2} \bigg|_{s=0} = -V < 0,
\]
Since item e in Proposition IV.4 guarantees
which implies that the first-order derivative $\partial E_0(s)/\partial s$ is strictly decreasing around $s = 0$. Hence, we conclude $s_0 = 0$. Because the subsequence is arbitrary, Eq. (137) is shown.

Next, from Eqs. (138) and Eqs. (139), the mean value theorem states that there exists a number $s_{nk} \in (0, s^*_n)$, for each $k \in \mathbb{N}$, such that

$$-\frac{\partial^2 E_0(s)}{\partial s^2} \bigg|_{s=s_{nk}} = -H + (H + \delta_{nk}) = \delta_{nk},$$

(140)

When $k$ approaches infinity, items a and e in Proposition IV.4 give

$$\lim_{k \to +\infty} \frac{\partial^2 E_0(s)}{\partial s^2} \bigg|_{s=s_{nk}} = 0.$$

Combining Eqs. (140) and (141) leads to

$$\lim_{k \to +\infty} \frac{\partial s^*_{nk}}{\partial \delta_{nk}} = \frac{1}{\mathcal{V}}.$$

Since the subsequence was arbitrary, the above result establishes Eq. (134).

Finally, denote by

$$\Upsilon = \max_{s \in [0,1]} \left| \frac{\partial^3 E_0(s)}{\partial s^3} \right| < +\infty.$$

For every sufficiently large $n \geq N_0$, we apply Taylor’s theorem to the map $s^*_n \to E_0(s^*_n)$ at the original point to obtain

$$E_{sp}(H + \delta_n) = s^*_n (H + \delta_n) + E_0(s^*_n) + \frac{1}{2} s^*_n \frac{\partial^2 E_0(s, P_n)}{\partial s^2} \bigg|_{s=s^*_n} + \frac{1}{6} s^*_n \frac{\partial^3 E_0(s, P_n)}{\partial s^3} \bigg|_{s=s^*_n},$$

(142)

where $s^*_n$ is some number in $(0, s^*_n)$. Then, Eqs. (134), (137), (142), and the assumption $\lim_{n \to +\infty} \delta_n = 0$ imply that the desired inequality

$$\lim_{n \to +\infty} \frac{E_{sp}(H + \delta_n)}{2 \delta_n^2} \leq \frac{1}{2\mathcal{V}}.$$

\[\square\]

ACKNOWLEDGMENT

Eric P. Hanson would like to thank the Institut Henri Poincaré, where part of this research was carried out, for its support and hospitality.

REFERENCES

[1] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, “Error exponents and strong converse exponents for classical data compression with quantum side information,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2018, pp. 2162–2166.

[2] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.

[3] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” IEEE Trans. Inf. Theory, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[4] Y. Oohama and T. Sun Han, “Universal coding for the Slepian-Wolf data compression system and the strong converse theorem,” IEEE Trans. Inf. Theory, vol. 40, no. 6, pp. 1908–1919, Nov. 1994.

[5] T. Cover, “A proof of the data compression theorem of Slepian and Wolf for ergodic sources (Corresp.),” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 226–228, Mar. 1975.

[6] A. Wyner, “On source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol. IT-21, no. 3, pp. 294–300, May 1975.

[7] R. Ahlswede and J. Körner, “Source coding with side information and a converse for degraded broadcast channels,” IEEE Trans. Inf. Theory, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.

[8] I. Devetak and A. Winter, “Classical data compression with quantum side information,” Phys. Rev. A, Gen. Phys., vol. 68, no. 4, Oct. 2003, Art. no. 042301.

[9] H.-C. Cheng and M.-H. Hsieh, “Moderate deviation analysis for classical-quantum channels and quantum hypothesis testing,” IEEE Trans. Inf. Theory, vol. 60, no. 2, pp. 1385–1403, Feb. 2018.

[10] C. T. Chubb, V. Y. F. Tan, and M. Tomamichel, “Moderate deviation analysis for classical communication over quantum channels,” Commun. Math. Phys., vol. 355, no. 3, pp. 1283–1315, Nov. 2017.

[11] C. Rouze and N. Datta, “Finite blocklength and moderate deviation analysis of hypothesis testing of correlated quantum states and application to classical-quantum channels with memory,” IEEE Trans. Inf. Theory, vol. 64, no. 1, pp. 593–612, Jan. 2018.

[12] R. G. Gallager, Source Coding With Side Information and Universal Coding. Accessed: 1976. [Online]. Available: http://web.mit.edu/gallager/www/papers/paper5.pdf

[13] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[14] M. Dalai, “Lower bounds on the probability of error for classical and classical-quantum channels,” IEEE Inf. Theory, vol. 59, no. 12, pp. 8027–8056, Dec. 2013.

[15] M. Dalai and A. Winter, “Constant compositions in the sphere packing bound for classical-quantum channels,” IEEE Trans. Inf. Theory, vol. 63, no. 9, pp. 5603–5617, Sep. 2017.

[16] H.-C. Cheng, M.-H. Hsieh, and M. Tomamichel, “Quantum sphere-packing bounds with polynomial prefactors,” IEEE Trans. Inf. Theory, vol. 65, no. 5, pp. 2872–2898, May 2019.

[17] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh, “Duality between source coding with quantum side information and cq channel coding,” in Proc. IEEE Int. Symp. Inf. Theory, Jul. 2019, pp. 1142–1146, doi: 10.1109/isit.2019.8849457.

[18] H.-C. Cheng, “Error exponent analysis in quantum information theory,” Ph.D. dissertation, School Comput. Sci., Dept. Eng. Inf. Technol., Univ. Technol. Sydney, Ultimo, NSW, Australia, 2018.

[19] R. Nomura and T. S. Han, “Second-order Slepian-Wolf coding theorems,” J. Math. Phys., vol. 54, no. 12, Art. no. 122203.

[20] V. Y. F. Tan and O. Kosut, “On the dispersions of three network information theory problems,” IEEE Trans. Inf. Theory, vol. 60, no. 2, pp. 881–903, Feb. 2014.

[21] V. Y. F. Tan, “Asymptotic estimates in information theory with non-vanishing error probabilities,” Found. Trends Commun. Inf. Theory, vol. 10, no. 4, pp. 1–184, 2014.

[22] M. Tomamichel and M. Hayashi, “A hierarchy of information quantities for finite block length analysis of quantum tasks,” IEEE Trans. Inf. Theory, vol. 59, no. 11, pp. 7693–7710, Nov. 2013.

[23] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, “On quantum Rényi entropies: A new generalization and some properties,” J. Math. Phys., vol. 54, no. 12, Dec. 2013, Art. no. 122203.

[24] M. M. Wilde, A. Winter, and D. Yang, “Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy,” Commun. Math. Phys., vol. 331, no. 2, pp. 593–622, Oct. 2014.

[25] M. Tomamichel, Quantum Information Processing With Finite Resources. Springer, 2016.

[26] M. Tomamichel, “A framework for non-asymptotic quantum information theory,” Ph.D. dissertation, Dept. Phys., ETH Zurich, Zürich, Switzerland, Mar. 2012.

[27] F. Ledizky, M. M. Wilde, and N. Datta, “Strong converse theorems using Rényi entropies,” J. Math. Phys., vol. 57, no. 8, Aug. 2016, Art. no. 082202.

[28] M. Mosonyi and T. Ogawa, “Strong converse exponent for classical-quantum channel coding,” Commun. Math. Phys., vol. 355, no. 1, pp. 373–426, Oct. 2017.
Eric P. Hanson received the bachelor’s degree in mathematics and physics from McGill University, Canada, in 2014, the master’s degree in mathematics from McGill University, in 2016, and the Ph.D. degree from Cambridge University, U.K., in 2020. This research is carried out during his Ph.D. degree, in 2020. His scientific interests include quantum information theory, convex optimization, and distance measures for distributions and signals.

Nilanjana Datta received the Ph.D. degree in mathematical physics from ETH Zurich, Switzerland, in 1996. From 1997 to 2000, she was a Post-Doctoral Researcher with the Dublin Institute of Advanced Studies, C.N.R.S. Marseille, and EPFL, Lausanne. In 2001, she joined the University of Cambridge, U.K., as a Lecturer in Mathematics with the Pembroke College, and a member of the Statistical Laboratory, Centre for Mathematical Sciences. She is currently a Reader in Quantum Information Theory with the Department of Applied Mathematics and Theoretical Physics, University of Cambridge. Her scientific research interests include quantum information theory and mathematical physics. She is a Fellow of the Pembroke College.

Min-Hsiu Hsieh (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2008. From 2008 to 2010, he was a Researcher with the ERATO-SORST Quantum Computation and Information Project, Japan Science and Technology Agency, Tokyo, Japan. From 2010 to 2012, he was a Post-Doctoral Researcher with the Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, U.K. He is currently an Australian Research Council (ARC) Future Fellow and an Associate Professor with the Centre for Quantum Software and Information (UTS:QSI), Faculty of Engineering and Information Technology (FEIT), University of Technology Sydney (UTS). His scientific research interests include quantum Shannon theory, entanglement theory, and quantum coding theory.