Moderate deviation expansion for fully quantum tasks

Navneeth Ramakrishnan,1,2,∗ Marco Tomamichel,2,3 and Mario Berta1,4,5,†

1 Department of Computing, Imperial College London, London, UK
2 Center for Quantum Technologies, National University of Singapore, Singapore
3 Department of Electrical and Computer Engineering, National University of Singapore, Singapore
4 Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, USA
5 AWS Center for Quantum Computing, Pasadena, USA

The moderate deviation regime is concerned with the finite block length trade-off between communication cost and error for information processing tasks in the asymptotic regime, where the communication cost approaches a capacity-like quantity and the error vanishes at the same time. We find exact characterisations of these trade-offs for a variety of fully quantum communication tasks, including quantum source coding, quantum state splitting, entanglement-assisted quantum channel coding, and entanglement-assisted quantum channel simulation. The main technical tool we derive is a tight relation between the partially smoothed max-information and the hypothesis testing relative entropy. This allows us to obtain the expansion of the partially smoothed max-information for i.i.d. states in the moderate deviation regime.

I. OVERVIEW

Fundamental tasks in information theory are often characterized in the asymptotic setting: One considers i.i.d. copies of a resource such as a channel or a state, takes the limit of the number of copies to infinity, and then enforces that the error vanishes in the asymptotic limit to obtain an optimal coding rate [57]. While such results are fundamental theoretical bounds, practical settings involve only finitely many copies of the resource — the so-called finite block length setting. It is therefore of interest to also understand the quantitative trade-off between the error and the rate for a given block length.

In the moderate deviation setting one obtains a rate and an error in terms of the block length \( n \), such that the rate approaches a capacity-like quantity and the error vanishes as the code length \( n \) increases. In contrast, second-order (small deviation) analyses [33, 42, 51, 60, 63] result in a rate that approaches the capacity but the error is only bounded by some constant. As for error exponent (large deviation) analyses [7, 20, 25, 30, 45], these result in an error that vanishes exponentially with \( n \) but the difference between the rate and the capacity is a constant term. The moderate deviation regime is natural to consider since it achieves both these goals simultaneously and several classical information-theoretic tasks have been analysed in this setting. Classical channel coding has been studied in [3, 51], while asymmetric binary hypothesis testing has been studied in [55]. Moderate deviation analyses for several classical-quantum coding tasks and quantum hypothesis testing are found in [15, 16, 18, 54]. Other settings such as majorization-based resource interconversion have also been explored [19]. Recently, moderate deviation analyses have been done for private communication over wiretap channels and entropy accumulation [59], quantum soft covering [14, 58], and privacy amplification against quantum side information [58, 59].

A moderate sequence is defined as a sequence of positive numbers \( \{a_n\} \) such that \( \lim_{n \to \infty} a_n = 0 \) and \( \lim_{n \to \infty} a_n \sqrt{n} = \infty \) [18]. We furthermore introduce the strict moderate deviation regime,

∗ Part of this work appeared in the Information Theory Workshop 2021 [52].
† This work was completed prior to MB joining the AWS Center for Quantum Computing.
where we impose additionally that the moderate sequence satisfies \( \lim_{n \to \infty} (na_n^2)^{-1} \log n = 0 \). An example of a strictly moderate sequence is \( \{n^{-\alpha}\} \) for \( n \in \mathbb{N} \) and \( \alpha \in (0, \frac{1}{2}) \) and in fact, this definition has been used in classical works to define a moderate sequence [36, 37, 65]. Within the moderate deviation regime, one can consider the low-error case where the error in the protocol is chosen to be \( \varepsilon_n = e^{-a_n^2} \) and the high-error case where the error is chosen to be \( \varepsilon_n = 1 - e^{-a_n^2} \).

As an illustrative example of our work, consider the task of quantum source coding: Alice holds a quantum state \( \rho_A^n \) and the goal is to send this state to Bob using a noiseless channel while preserving correlations with a purifying reference quantum register. A fundamental result in quantum information theory [56] states that the asymptotic rate at which this can be achieved is given by the von Neumann entropy \( S(A)_\rho \). We consider the setting for finite \( n \in \mathbb{N} \) and aim to achieve quantum source coding such that the output of the protocol \( \tilde{\rho}_B^n \) is \( \varepsilon_n \)-close in purified distance to the ideal output state \( \rho_B^{\otimes n} = I_{A^n \rightarrow B^n} (\rho_A^{\otimes n}) \). We show in Theorem 4.3 that the minimal quantum communication cost \( q^*_n(\rho_A^{\otimes n}) \) of \( \varepsilon_n \)-error quantum source coding of \( \rho_A^{\otimes n} \) with \( \varepsilon_n = e^{-na_n^2} \) for a moderate sequence \( \{a_n\} \) is given by

\[
\frac{1}{n} q^*_n(\rho_A^{\otimes n}) = S(A)_\rho + 2a_n \sqrt{V(A)_\rho} + o(a_n),
\]

where \( V(A)_\rho = \text{Tr} (\rho (\log \rho)^2) - S(A)_\rho^2 \) denotes the quantum varentropy and \( o(a_n) = f(n) \) for functions with \( \lim_{n \to \infty} f(n)a_n^{-1} = 0 \).

We show similar moderate deviations analyses for quantum state splitting [2, 6] (Theorem III.6), entanglement-assisted quantum channel coding [8] (Theorem VI.2), and quantum channel simulation [8, 10, 13] (Theorem V.3). For quantum source coding, both second-order [1] and error exponent [31, 33] analyses already exist and it is possible to derive our moderate deviation results from these results. Similarly, second-order analysis [23] for entanglement-assisted channel coding may also be used to derive our moderate deviation results.

Some of our technical discussion may be of independent interest: First, we use a fidelity-based channel distinguishability measure — the channel purified distance (Definition V.1) — which emerges as a natural fully quantum measure to analyse channel coding and channel simulation problems (versus trace-distance-based measures such as the diamond distance). Second, we determine when the tight triangle inequality for the purified distance [62, Proposition 3.16] holds (Lemma III.7). Third, from this refined triangle inequality, we derive a tight entropy inequality between the min-relative entropy and the hypothesis testing relative entropy (Lemma III.9). At the same time, we then find the expansion of the partially smoothed max-information for i.i.d. states in the moderate deviation regime.

Our paper is organized as follows. We introduce the relevant notation in Section II, followed by our analysis of quantum state splitting in Section III — where we also introduce the necessary mathematical moderate deviation tools. Quantum source coding is discussed in Section IV, quantum channel simulation in Section V, and quantum channel coding in Section VI. We conclude and discuss some open problems in Section VII. Various technical proofs are deferred to Appendices A–C.

Note added: During finalization of our work we became aware of the related concurrent work [43] by Li and Yao.

---

1 We resolve either the high-error or the low-error for each quantum information-theoretic task we consider.
II. NOTATION

A. Mathematics

The exponential and logarithm functions, denoted by exp and log respectively, are taken with respect to base 2 unless otherwise specified. The set of real numbers is denoted by $\mathbb{R}$ and the set of natural numbers by $\mathbb{N}$. We use the big-O and little-o notation which are defined as follows: For functions $f(n), g(n)$ on $\mathbb{N}$, we denote $f(n) = O(g(n))$ if there exist some positive constants $c$ and $n_0$ such that for all $n \geq n_0$, it holds that $|f(n)| \leq c|g(n)|$. We denote $f(n) = o(g(n))$ if $\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0$.

B. Quantum states

We associate quantum registers with Hilbert spaces and assume them to be finite-dimensional. We label quantum registers and the associated Hilbert spaces with capital letters, e.g., $A$, $B$, and denote their dimension by $|A|$, $|B|$, and so on. The notation $A \cong B$ states that $A$ and $B$ are isomorphic. The tensor product of spaces $A$ and $B$ is denoted by $AB = A \otimes B$. The set of linear operators on $A$ is denoted by $\mathcal{L}(A)$ and the set of positive semi-definite operators on $A$ is denoted by $\mathcal{P}(A)$, where for elements $\rho_A \in \mathcal{P}(A)$ we write $\rho_A \succeq 0$. The set of positive semi-definite operators with unit trace — called normalized quantum states or quantum states for short — is denoted by $\mathcal{S}(A)$. The set of positive semi-definite operators with unit trace — called normalized quantum states or quantum states for short — is denoted by $\mathcal{S}(A)$. If the rank of $\rho_A \in \mathcal{S}(A)$ is one, then the (sub-normalized) quantum state is called pure. An extension of $\rho_A \in \mathcal{S}(A)$ is $\rho_{AB} \in \mathcal{S}(AB)$ such that $\text{Tr}_B(\rho_{AB}) = \rho_A$, where $\text{Tr}_B$ denotes taking a trace over the quantum register $B$ (see [66, Definition 4.3.4] for a definition of the partial trace). If an extension $\rho_{AB}$ of $\rho_A \in \mathcal{S}(A)$ is pure, then it is called a purification.

When the registers associated with a state are clear from context or do not need to be explicitly stated, we drop them for simplicity of notation. For $\rho, \sigma \succeq 0$ with the support of $\rho$ contained in the support of $\sigma$, we write $\rho \ll \sigma$. Orthogonality for $\rho, \sigma \succeq 0$ is defined as $\rho \sigma = \sigma \rho = 0$ and denoted as $\rho \perp \sigma$. The identity operator on a $d$-dimensional quantum register is denoted by $I_d$. The generalized inverse [49] of $\rho \in \mathcal{S}(A)$ is denoted by $\rho^{-1}$ and is the inverse on the support of the quantum state.

C. Quantum operations

A quantum channel $\mathcal{N}_{A \to B}$ is a linear completely positive trace-preserving (CPTP) map taking $\mathcal{L}(A)$ to $\mathcal{L}(B)$. The identity quantum channel between isomorphic quantum registers $A$ and $B$ is denoted by $\mathcal{I}_{A \to B}$ and we abbreviate $\mathcal{I}_{A \to A} = \mathcal{I}_A$. We denote composition of two quantum channels $\mathcal{N}_{A \to B}$ and $\mathcal{M}_{B \to C}$ by $\mathcal{M}_{B \to C} \circ \mathcal{N}_{A \to B}$.

A positive operator valued measure (POVM) is a finite set $\{M_i\}_{i=1}^n$ of $d$-dimensional positive definite matrices $M_i \succeq 0$ such that $\sum_{i=1}^n M_i = I_d$ for some $d, n \in \mathbb{N}$.

D. Distance measures

The trace norm of a linear operator $M$ is given by $\|M\|_1 = \text{Tr} \left( \sqrt{M^\dagger M} \right)$, where $M^\dagger$ is the transpose conjugate of $M$. The purified distance between $\rho, \sigma \in \mathcal{S}(A)$ is given by $P(\rho, \sigma) = \|\rho - \sigma\|_1$. The purification distance between $\rho, \sigma \in \mathcal{S}(A)$ is given by $D(\rho, \sigma) = \|\rho - \sigma\|_1$. The fidelity for quantum states $\rho, \sigma$ is defined as $F(\rho, \sigma) = \text{Tr} (\sqrt{\sqrt{\rho} \sigma \sqrt{\rho}})$.
\(\sqrt{1 - F(\rho, \sigma)^2}\), where \(F(\rho, \sigma)\) is the generalized fidelity defined as [61]

\[
F(\rho, \sigma) = \|\sqrt{\rho \otimes (1 - \text{Tr}(\rho))} \sqrt{\sigma \otimes (1 - \text{Tr}(\sigma))}\|_1.
\]  

(2)

Note that if either \(\rho\) or \(\sigma\) have unit trace, then the generalized fidelity is equal to the standard fidelity \(F(\rho, \sigma) = \|\sqrt{\rho} \sqrt{\sigma}\|_1\). For \(\rho, \sigma \in S_\leq(A)\) with \(P(\rho, \sigma) \leq \varepsilon\) we write \(\rho \approx \varepsilon \sigma\). We define the \(\varepsilon\)-ball around \(\rho \in S_\leq(A)\) by \(B^\varepsilon(\rho) = \{\tilde{\rho} \in S_\leq(A) : \rho \approx \varepsilon \tilde{\rho} \leq \varepsilon\}\). For \(\rho, \sigma \in S(A)\) the purified distance is bounded by the trace distance as [28]

\[
\sqrt{2}\|\rho - \sigma\|_1 \geq P(\rho, \sigma) \geq \|\rho - \sigma\|_1.
\]  

(3)

E. Entropic quantities

For \(\rho \in S(A)\), the von Neumann entropy is defined as

\[
S(A)_\rho = S(\rho) = -\text{Tr}(\rho \log \rho).
\]  

(4)

For \(\rho \in S_\leq(A)\) and \(\sigma \in P(A)\) the quantum relative entropy is defined as

\[
D(\rho || \sigma) = \text{Tr}(\rho (\log \rho - \log \sigma))
\]  

(5)

when \(\rho \ll \sigma\) and \(\infty\) otherwise. The relative entropy variance is given by [42, 63]

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

(6)

when \(\rho \ll \sigma\) and \(\infty\) otherwise. Following the classical definition [41], we define the quantum varentropy of \(\rho \in S(A)\) to be

\[
V(A)_\rho = \text{Tr}(\rho (\log \rho)^2) - S(A)_\rho^2.
\]  

(7)

Furthermore, we define the channel based entropic quantities for quantum channels \(N_{A \rightarrow B}\) as

\[
C(N) = \max_{\sigma_{AR} \in S(AR)} \frac{1}{2} I(B : R)_{(N \otimes I) (\sigma)} \quad \text{and} \quad V_{\max}(B : R)_N = \max_{\sigma' \in \Pi(N)} V(B : R)_{(N \otimes I) (\sigma')},
\]  

(8)

where \(R\) is any purifying quantum register of the input of the channel and

\[
\Pi(N) = \arg \max_{\tau} I(B : R)_{(N \otimes I) (\tau)}
\]  

(9)

is called the set of capacity achieving inputs for \(N\).

The sandwiched Rényi relative entropies for \(\rho \in S(A)\) and \(\sigma \in P(A)\) are defined [46, 67] as follows: For \(\alpha \in (0, 1) \cup (1, \infty)\), if either \(\alpha < 1\) and \(\rho \not\approx \sigma\) or \(\alpha > 1\) and \(\rho \ll \sigma\), we have

\[
\tilde{D}_\alpha(\rho || \sigma) = \frac{1}{\alpha - 1} \log \text{Tr}(\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha).
\]  

(10)

Furthermore, for \(\alpha = \{0, 1, \infty\}\) the respective limits are employed. The max-relative entropy of \(\rho \in S_\leq(A)\) with respect to \(\sigma \in P(A)\) is given by [21, 40]

\[
D_{\max}(\rho || \sigma) = \min \{\lambda \in \mathbb{R} : 2^\lambda \sigma \succeq \rho\},
\]  

(11)
where $\rho \succeq \sigma$ means that $\rho - \sigma \succeq 0$. Note that we have $D_{\text{max}}(\rho \| \sigma) = \tilde{D}_{\infty}(\rho \| \sigma)$. The min-relative entropy is given by

$$D_{\text{min}}(\rho \| \sigma) = -\log \left\| \sqrt{\rho} \sqrt{\sigma} \right\|^2 = -\log F(\rho, \sigma)^2.$$  \hfill (12)

Note that we have $D_{\text{min}}(\rho \| \sigma) = \tilde{D}_{1/2}(\rho \| \sigma)$.

The quantum mutual information for $\rho_{AB} \in S_{\leq}(AB)$ is defined as

$$I(A : B)_\rho = D(\rho_{AB} \| \rho_A \otimes \rho_B)$$  \hfill (13)

and the mutual information variance as

$$V(A : B)_\rho = V(\rho_{AB} \| \rho_A \otimes \rho_B).$$  \hfill (14)

The max-information that $B$ has about $A$ for $\rho_{AB} \in S(AB)$ is [13]

$$I_{\text{max}}(A ; B)_\rho = \min_{\sigma_B \in S(B)} D_{\text{max}}(\rho_{AB} \| \rho_A \otimes \sigma_B).$$  \hfill (15)

The generalized sandwiched Rényi mutual information for $\rho_{AB} \in S(AB)$ and $\tau_A \in P(A)$ with $\rho_A \ll \tau_A$ is [29]

$$\tilde{I}_\alpha(\rho_{AB} \| \tau_A) = \min_{\sigma_B \in S(B)} \tilde{D}_\alpha(\rho_{AB} \| \tau_A \otimes \sigma_B).$$  \hfill (16)

\section*{F. Smoothed entropic quantities}

The quantum hypothesis testing relative entropy is defined for $\varepsilon \in [0, 1)$, $\rho \in S(A)$, and $\sigma \in P(A)$ as [38, 48, 64]

$$D^\varepsilon_h(\rho \| \sigma) = -\log \inf_{\Lambda \geq 1 - \varepsilon} \rho \leq A \leq \varepsilon \rho \Lambda.$$

The information spectrum relative entropy is defined for $\varepsilon \in (0, 1)$, $\rho \in S(A)$, and $\sigma \in P(A)$ as [22, 63]

$$D^\varepsilon_s(\rho \| \sigma) = \sup \left\{ \gamma \mid \text{Tr} \left( (\rho - 2\gamma \sigma) \{ \rho > 2\gamma \sigma \} \right) \geq 1 - \varepsilon \right\},$$  \hfill (18)

where $\{ \cdot \}^+$ is the projector on the positive part of the argument. The information spectrum entropy is defined as $H^\varepsilon(\rho) = -D^\varepsilon_s(\rho \| I)$ [22].

One defines further smooth entropy measures by extremizing over a set of states in a small neighbourhood [53]. The smooth max-relative entropy between $\rho_A \in S_{\leq}(A)$ and $\sigma_A \in S_{\leq}(A)$ is given as

$$D^\varepsilon_{\text{max}}(\rho_A \| \sigma_A) = \inf_{\bar{\rho} \in B^\varepsilon(\rho)} D_{\text{max}}(\bar{\rho}_A \| \sigma_A)$$  \hfill (19)

and the smooth min-relative entropy is given as [27]

$$D^\varepsilon_{\text{min}}(\rho_A \| \sigma_A) = \sup_{\bar{\rho} \in B^\varepsilon(\rho)} D_{\text{min}}(\bar{\rho}_A \| \sigma_A).$$  \hfill (20)

The smooth max-information that $B$ has about $A$ for $\rho_{AB} \in S(AB)$ is given by [13]

$$I^\varepsilon_{\text{max}}(A ; B)_\rho = \inf_{\bar{\rho} \in B^\varepsilon(\rho)} I_{\text{max}}(A ; B)|_{\bar{\rho}}.$$  \hfill (21)
We define the partially smoothed max-information as [5]

\[ I^\varepsilon_{\text{max}}(\hat{A};B)_\rho = \inf_{\rho_{AB} \in S(\rho_{AB})} I_{\text{max}}(A;B)_\rho. \] (22)

Note that for all the above quantities, we may replace the infimum (supremum) with the minimum (maximum) since the set of subnormalized quantum states is closed and bounded and the functions being extremized are continuous.

III. QUANTUM STATE SPLITTING

A. Task

Here, we introduce quantum state splitting first formulated in [2], with more refined versions appearing in [5, 6, 13].

**Definition III.1 (One-shot state splitting)** Let \( \rho_{AB} \in S(AB) \) and \( \varepsilon \in [0,1] \). A one-shot quantum state splitting protocol consists of:

1. Quantum registers \( Q, K, L \) and \( A_1 \cong B \)
2. A resource \( \omega_{KL} \in S(KL) \)
3. An encoding quantum channel \( E_{AA_1K \rightarrow AQ} \)
4. A decoding quantum channel \( D_{QL \rightarrow B} \).

A \( \{q,\varepsilon\}\)-one-shot quantum state splitting protocol of \( \rho_{AB} \) is such that \( q = \log |Q| \) with

\[ (D \circ E)(\rho_{AA_1R} \otimes \omega_{KL}) \approx_{\varepsilon} \rho_{ABR} \text{ for any extension } \rho_{AA_1R} \text{ of } \rho_{AA_1}. \] (23)

The minimal quantum communication cost of state splitting \( \rho_{AB} \) is defined as

\[ q^*_\varepsilon(\rho_{AB}) = \min \left\{ q \in \mathbb{N} : \exists \text{ a } \{q,\varepsilon\}\text{-one-shot quantum state splitting protocol of } \rho_{AB} \right\}. \] (24)

**Remark III.2** Instead of considering an arbitrary extension \( \rho_{ABR} \) in Definition III.1, one can choose any fixed purification of \( \rho_{AB} \) in the analysis of quantum state splitting. Such a choice can be made without loss of generality because any two purifications are related by an isometry on the reference quantum register and such isometries commute with the protocol. Since the purified distance is non-increasing under partial trace, a \( \{q,\varepsilon\}\)-one-shot quantum state splitting protocol of a fixed purification is also a \( \{q,\varepsilon\}\)-one-shot quantum state splitting protocol of all extensions of \( \rho_{AB} \).

An alternative but equivalent description of quantum state splitting is through a quantum channel \( T_{AA_1 \rightarrow AB}^\sigma \) such that

\[ T^\sigma = D \circ E \circ P^\sigma \] (25)

where \( P^\sigma \) is the preparation quantum channel of the resource state \( \sigma_{KL} \). We note that \( T^\sigma(\cdot) = (D \circ E)(\cdot \otimes \sigma_{KL}) \). The minimal quantum communication cost is then quantified as follows.

---

\( ^2\) Since free entanglement can be included in the resource state \( \omega_{KL} \), we equivalently have that the classical communication \( 2q \).
Theorem III.3 (One-shot quantum state splitting [5, 6]) Let \( \rho_{AB} \in S(AB) \), \( \varepsilon \in (0, 1] \), and \( \delta \in (0, \varepsilon] \). Then, the minimal quantum communication cost of \( \varepsilon \)-error quantum state splitting of \( \rho_{AB} \) is bounded as

\[
\frac{1}{2} I_{\max}^\varepsilon(\hat{R}; B)_\rho \leq q_\varepsilon^*(\rho_{AB}) \leq \frac{1}{2} I_{\max}^{\varepsilon-\delta}(\hat{R}; B)_\rho + \log \frac{2}{\delta},
\]

where \( \rho_{AB} \) is any purification of \( \rho_{AB} \).

For completeness we include the proof for our exact setting in Appendix A. From Theorem III.3, one directly obtains the quantum communication cost of quantum state splitting of i.i.d. states in the low-error moderate deviation regime. Namely, by Lemma B.11, one can bound the partially smoothed max-information with the smoothed max-information, and then the asymptotic equipartition property for the smoothed max-information states that [13, Corollary B.22]

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} I_{\max}^\varepsilon(A : B)_{\rho^{\otimes n}} = I(A : B)_\rho.
\]

Hence, the minimal quantum communication cost \( q_\varepsilon^*(\rho_{AB}^{\otimes n}) \) of \( \varepsilon \)-error one-shot quantum state splitting of \( \rho_{AB}^{\otimes n} \) satisfies

\[
\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{q_\varepsilon^*(\rho_{AB}^{\otimes n})}{n} = \frac{1}{2} I(R : B)_\rho.
\]

B. Moderate deviation expansion

We investigate as a refinement of Eq. (28), namely, the minimal quantum communication cost of one-shot state splitting protocols for i.i.d. states in the low-error moderate deviation regime. We start with moderate sequences as defined in [18] and additionally define strictly moderate sequences.

Definition III.4 (Moderate sequences) A sequence of non-negative numbers \( \{a_n\} \) with \( n \in \mathbb{N} \) is a moderate sequence if \( \lim_{n \to \infty} a_n = 0 \) and \( \lim_{n \to \infty} na_n^2 = \infty \). A strictly moderate sequence is a moderate sequence that additionally satisfies \( \lim_{n \to \infty} \frac{\log n}{na_n^2} = 0 \).

In particular, the sequence \( \{n^\alpha\} \) for \( n \in \mathbb{N} \) and \(-1/2 < \alpha < 0\) is a strictly moderate sequence. For \( \varepsilon_n = e^{-a_n^2} \) for some moderate sequence \( \{a_n\} \), we investigate \( \varepsilon_n \)-one-shot quantum state splitting protocols of \( \rho_{AB}^{\otimes n} \) and investigate how \( q_\varepsilon^*(\rho_{AB}^{\otimes n}) \) behaves as a function of \( n \).

In the moderate deviation regime, the communication cost in our derivations will contain higher order terms that depend on \( a_n \). The lemma below shows that for moderate sequences, the error can be multiplied by a constant and still leave the communication cost unchanged up to \( o(a_n) \) terms. Similarly, if the error is multiplied by a poly\((n)\) factor for strictly moderate sequences, the communication cost remains unchanged up to \( o(a_n) \) terms.

Lemma III.5 For \( k > 0 \), moderate sequences \( \{a_n\} \), and \( \varepsilon_n = e^{-a_n^2} \), it holds that

\[
k\varepsilon_n = \exp \left( -n \left( a_n^2 - \frac{\log k}{n} \right) \right) = \exp(-nb_n^2)
\]

for some moderate sequence \( \{b_n\} \). In particular, for \( \eta > 0 \), there exists \( n \in \mathbb{N} \) sufficiently large such that \( b_n \leq a_n + \eta a_n \). If \( \{a_n\} \) is a strictly moderate sequence, then poly\((n)\)\(\varepsilon_n = \exp(-nb_n^2)\) for a moderate sequence \( \{b_n\} \) and for \( \eta' > 0 \), there exists sufficiently large \( n \in \mathbb{N} \) such that \( b_n \leq a_n + \eta' a_n \).
Proof. For the first case we have that for sufficiently large \( n \)

\[
b_n = \sqrt{a_n^2 - \frac{\log k}{n}} = a_n \left( \sqrt{1 - \frac{\log k}{na_n^2}} \right) \leq a_n \left( 1 - \frac{\log k}{2na_n^2} \right) \leq a_n + \eta a_n ,
\]

(30)

where the first inequality uses \( \sqrt{1 - x} \leq 1 - \frac{x}{2} \) for \( x \leq 1 \) and the second inequality uses the fact that \( \lim_{n \to \infty} na_n^2 = \infty \). The second case follows the same argument except that we replace the constant \( k \) with \( \text{poly}(n) \) and note that for strictly moderate sequences, \( \lim_{n \to \infty} \frac{\log n}{na_n^2} = 0 \) for any constant \( r \in \mathbb{R} \).

In the following, the main idea is to obtain an expansion for the partially smoothed max-information of i.i.d. states. This is done by bounding the partially smoothed max-information with the hypothesis testing relative entropy and then using the expansion of the hypothesis testing relative entropy in the moderate deviation regime. Our main result in this section is the following characterization.

Theorem III.6 (State splitting moderate deviation) The minimal quantum communication cost of \( \varepsilon_n \)-error state splitting of \( \rho_{AB}^{n} \) for \( \rho_{AB} \in S(AB) \) with \( \varepsilon_n = e^{-na_n^2} \) for a moderate sequence \( \{a_n\} \) has the asymptotic expansion

\[
\frac{1}{n} q_{\varepsilon_n}^*(\rho_{AB}^{n}) = \frac{1}{2} I(R : B)_{\rho} + a_n \sqrt{V(R : B)_{\rho}} + o(a_n) ,
\]

(31)

where \( \rho_{ABR} \) is any purification of \( \rho_{AB} \).

The proof is developed in the following subsections.

C. Technical tools

As the first step for the proof of Theorem III.6, we clarify when the tighter triangle inequality for the purified distance [62, Proposition 3.16] holds. The following lemmas may be of independent interest since this tighter triangle inequality yields tight bounds between the smoothed min-relative entropy and the smoothed max-relative entropy, as well as between the smoothed min-relative entropy and the hypothesis testing relative entropy.

Lemma III.7 (Tight triangle inequality purified distance) For \( \rho, \sigma, \tau \in S(A) \) we have

\[
P(\rho, \sigma)^2 + P(\sigma, \tau)^2 \leq 1 \quad \Rightarrow \quad P(\rho, \tau) \leq P(\rho, \sigma)F(\sigma, \tau) + P(\sigma, \tau)F(\rho, \sigma).
\]

(32)

Proof. The proof of [62, Proposition 3.16] shows that Eq. (32) holds when \( \sin^{-1}(P(\rho, \sigma)) + \sin^{-1}(P(\sigma, \tau)) \leq \pi/2 \). We show that this condition is equivalent to \( P(\rho, \sigma)^2 + P(\sigma, \tau)^2 \leq 1 \). By the monotonicity of the cosine function on \([0, \pi]\), we have

\[
\sin^{-1}(P(\rho, \sigma)) + \sin^{-1}(P(\sigma, \tau)) \leq \frac{\pi}{2}
\]

(33)

\[
\Leftrightarrow \cos(\sin^{-1}(P(\rho, \sigma)) + \sin^{-1}(P(\sigma, \tau))) \geq \cos\left(\frac{\pi}{2}\right)
\]

(34)

\[
\Leftrightarrow \cos(\sin^{-1}(P(\rho, \sigma))) \cos(\sin^{-1}(P(\sigma, \tau))) - \sin(\sin^{-1}(P(\rho, \sigma))) \sin(\sin^{-1}(P(\sigma, \tau))) \geq 0
\]

(35)

\[
\Leftrightarrow \sqrt{1 - P(\rho, \sigma)^2} \sqrt{1 - P(\sigma, \tau)^2} - P(\rho, \sigma)P(\sigma, \tau) \geq 0
\]

(36)

\[
\Leftrightarrow P(\rho, \sigma)^2 + P(\sigma, \tau)^2 \leq 1,
\]

(37)
where the third line follows by using the addition formula for cosine and the fourth line uses that \( \cos^2(\theta) = 1 - \sin^2(\theta) \).

The following lemma was already known for \( \rho_{AB}, \sigma_{AB} = I_A \otimes \rho_B \in S(AB) \) [61, Remark 5.6] but we get it for general quantum states.

**Lemma III.8** Let \( \rho \in S(A) \), \( \sigma \in P(A) \), and \( \varepsilon, \varepsilon' \in [0, 1] \) with \( \varepsilon^2 + \varepsilon'^2 \leq 1 \). Then, we have that

\[
D_{\min}^\varepsilon(\rho\|\sigma) \leq D_{\max}^{\varepsilon'}(\rho\|\sigma) - \log \left( 1 - (\varepsilon \sqrt{1 - \varepsilon'^2} + \varepsilon' \sqrt{1 - \varepsilon^2})^2 \right). \tag{38}
\]

**Proof.** Let \( \lambda = D_{\max}^{\varepsilon'}(\rho\|\sigma) = D_{\max}(\tilde{\rho}\|\sigma) \) for \( \tilde{\rho} \approx \varepsilon \rho \), it then holds that \( \tilde{\rho} \leq 2^\lambda \rho \). For some \( \tilde{\rho} \approx \varepsilon \rho \) we have that

\[
D_{\min}^\varepsilon(\rho\|\sigma) = D_{\min}(\tilde{\rho}\|\sigma) = -\log F(\tilde{\rho}, \sigma)^2 \tag{39}
\]
\[
\leq -\log F(\tilde{\rho}, 2^{-\lambda} \rho)^2 \tag{40}
\]
\[
= \lambda - \log F(\rho, \rho)^2 \tag{41}
\]
\[
= D_{\max}^{\varepsilon'}(\rho\|\sigma) - \log(1 - P(\tilde{\rho})^2) \tag{42}
\]
\[
\leq D_{\max}^{\varepsilon'}(\rho\|\sigma) - \log \left( 1 - (\varepsilon \sqrt{1 - \varepsilon'^2} + \varepsilon' \sqrt{1 - \varepsilon^2})^2 \right). \tag{43}
\]

where the last step is via the tighter triangle inequality for the purified distance (Lemma III.7) and a basic monotonicity argument (Lemma B.1).

Note that the constraint on the choices of \((\varepsilon, \varepsilon')\) is less stringent compared to [61, Remark 5.6], which is due to the tighter triangle inequality. Next, we have the following upper bond on the min-relative entropy in terms of the hypothesis testing relative entropy.

**Lemma III.9** Let \( \rho, \sigma \in P(A) \) and \( \varepsilon \in (0, k^{-1/2}] \) for \( k > 1 \). Then, we have that

\[
D_{\min}^\varepsilon(\rho\|\sigma) \leq D_{h}^{\varepsilon^2}(\rho\|\sigma) - \log f(\varepsilon) \tag{44}
\]

**Proof.** We choose \( \varepsilon' = \sqrt{1 - k\varepsilon^2} \) in the preceding Lemma III.8 to obtain

\[
D_{\min}^\varepsilon(\rho\|\sigma) \leq D_{\max}^{\sqrt{1 - k\varepsilon^2}}(\rho\|\sigma) - \log f(\varepsilon) \quad \text{for} \quad f(\varepsilon) = \log \left( 1 - (\varepsilon^2 \sqrt{k} + \sqrt{1 - k\varepsilon^2} \sqrt{1 - \varepsilon^2})^2 \right). \tag{45}
\]

By a connection of the smooth max-relative entropy to the hypothesis testing relative entropy (Lemma B.8), we have for \( \delta \in (0, 1) \)

\[
D_{\max}^{\sqrt{\delta}}(\rho\|\sigma) \leq D_{h}^{1 - \delta}(\rho\|\sigma) - \log(1 - \delta). \tag{46}
\]

Combining the two inequalities, we get

\[
D_{\min}^\varepsilon(\rho\|\sigma) \leq D_{\max}^{\sqrt{1 - k\varepsilon^2}}(\rho\|\sigma) - \log f(\varepsilon) \leq D_{h}^{1 - (\sqrt{1 - k\varepsilon^2})^2}(\rho\|\sigma) - \log f(\varepsilon) - \log k\varepsilon^2 \tag{47}
\]
\[
\leq D_{h}^{k\varepsilon^2}(\rho\|\sigma) - \log f(\varepsilon) - \log k\varepsilon^2. \tag{48}
\]

Finally, we recall the moderate deviation expansion for the hypothesis testing relative entropy of i.i.d. states.
Proposition III.10 (Hypothesis testing AEP [16, 18, 65]) For any moderate sequence \( \{a_n\} \), let \( \varepsilon_n = e^{-a_n^2} \). For \( \rho, \sigma \in \mathcal{S}(A) \) with \( \rho \ll \sigma \) and \( \eta_1, \eta_2 > 0 \), there exists \( n^* \in \mathbb{N} \) such that for \( n \geq n^* \) we have

\[
\frac{1}{n} D_h^{\varepsilon_n} (\rho^\otimes n \| \sigma^\otimes n) \leq D(\rho \| \sigma) - \sqrt{2V(\rho \| \sigma)} a_n + \eta_1 a_n. \\
\frac{1}{n} D_h^{1-\varepsilon_n} (\rho^\otimes n \| \sigma^\otimes n) \leq D(\rho \| \sigma) + \sqrt{2V(\rho \| \sigma)} a_n + \eta_2 a_n. 
\]

\[
\text{(49)} \quad \text{(50)}
\]

D. Proof of Theorem III.6

We now have the main technical results in hand to derive our moderate deviation analysis: An expansion for the partially smoothed max-information of i.i.d. states in the low-error moderate deviation regime.

Proposition III.11 (Partially smoothed max-information AEP) Let \( \varepsilon_n = e^{-a_n^2} \) for a moderate sequence \( \{a_n\} \) and let \( \rho_{BR} \in \mathcal{S}(BR) \). Then, for \( \eta > 0 \), there exists \( n^* \in \mathbb{N} \) such that for \( n \geq n^* \) we have

\[
I(B : R)_\rho + a_n \sqrt{4V(B : R)_\rho} - \eta a_n \leq \frac{1}{n} I_{\max}^\varepsilon (\tilde{R}^n; B^n)_{\rho^\otimes n} \leq I(B : R)_\rho + a_n \sqrt{4V(B : R)_\rho} + \eta a_n. 
\]

\[
\text{(51)}
\]

Proof. We start by showing the upper bound. For \( \eta > 0 \), there exists \( n_1^* \) such that for \( n \geq n_1^* \) we have

\[
I_{\max}^\varepsilon (\tilde{R}^n; B^n)_{\rho^\otimes n} \leq I_{\max} (R^n; B^n)_{\rho^\otimes n} + \log \frac{8 + (\varepsilon_n)^2}{(\varepsilon_n)^2} \\
\leq D_{\max} (\rho^\otimes n \| \tilde{\rho}_{B'}^n \otimes \tilde{\rho}_{BR}^n) + \log \frac{8 + (\varepsilon_n)^2}{(\varepsilon_n)^2} \\
\leq D_h (\rho_{BR}^\otimes_n \| \rho_B^\otimes_n \otimes \rho_R^\otimes_n) + \log \frac{8 + (\varepsilon_n)^2}{(\varepsilon_n)^2} \\
\leq n D(\rho_{BR} \| \rho_B \otimes \rho_R) + n \cdot a_n \sqrt{4V(\rho_{BR} \| \rho_B \otimes \rho_R)} + n \cdot \eta a_n, 
\]

where the first inequality follows from the bound on the partially smoothed max-information with the smoothed max-information (Lemma B.11), the second inequality is by choosing \( \rho_{BR}^\otimes_n \) instead of a minimization, the third inequality follows from the bounds between the max relative entropy and the hypothesis testing relative entropy (Lemma B.8) and the fourth inequality follows by the AEP for the hypothesis testing relative entropy (Proposition III.10), a property of moderate deviation sequences (Lemma III.5), the boundedness of the relative entropy variance (Lemma B.6), and noting that the term in the logarithm is \( n \cdot o(a_n) \).

To prove the lower bound, let \( \rho_{BBR}^\otimes \) be a purification of \( \rho_{BBR}^\otimes \) with \( R' \equiv BR \). For any \( \tilde{\rho}_{BR^n} \approx \varepsilon_n \rho_{BR}^\otimes_M \) there exists a purification \( \tilde{\rho}_{BR^n} = \varepsilon_n \rho_{BBR}^\otimes \) by Uhlmann’s theorem. We have

\[
I_{\max}^\varepsilon (\tilde{R}^n; B^n)_{\rho^\otimes n} = \inf_{\tilde{\rho}_{BR^n} \in \mathcal{B}(\rho_{BBR}^\otimes)} \inf_{\sigma_{BR^n}} D_{\max}(\tilde{\rho}_{BR^n} \| \tilde{\rho}_R^\otimes_n \otimes \sigma_{BR^n}) \\
\geq \inf_{\tilde{\rho} \in \mathcal{B}(\rho_{BBR}^\otimes)} \inf_{\sigma_{BR^n}} D_{\max}(\tilde{\rho}_{BR^n} \| \tilde{\rho}_R^\otimes_n \otimes \sigma_{BR^n}) \\
= \inf_{\tilde{\rho} \in \mathcal{B}(\rho_{BBR}^\otimes)} I_{\max}(\tilde{\rho}_{BR^n} \| \tilde{\rho}_R^\otimes_n) 
\]

\[
\text{(56)} \quad \text{(57)} \quad \text{(58)}
\]
where the first inequality follows from relaxing the constraint on $\tilde{\rho}_{BR}$, the second equality is from the definition of the sandwiched Rényi mutual information, the third equality follows from the duality of the sandwiched Rényi mutual information (Lemma III.9) and choosing a specific purification satisfying $\tilde{\rho}_{BR^n} \approx \rho_{BR^n}$. The second inequality follows by choosing a $\rho_{BR^n}^\otimes$ instead of a minimization over $\sigma_{BR}^n \in \mathcal{S}(R^n \otimes n)$, and the last inequality is by the definition of the smoothed min-relative entropy. We now choose $\eta > 0$ and $k > 1$ to obtain some $n_2$ such that for all $n \geq n_2$, we have

$$-D_{\min}^{\epsilon_n}(\rho_{RR^n}^\otimes (\rho_{R}^\otimes)^{-1} \otimes \rho_{R'}^\otimes) \geq -D_{\min}^{k\epsilon_n}(\rho_{RR^n}^\otimes (\rho_{R}^\otimes)^{-1} \otimes \rho_{R'}^\otimes) - n \cdot \frac{\eta a_n}{2}$$

$$\geq -nD(\rho_{RR^n}^\otimes (\rho_{R}^\otimes \otimes \rho_{R'}^\otimes) + n \cdot a_n \sqrt{4V(\rho_{RR^n}^\otimes (\rho_{R}^\otimes \otimes \rho_{R'}^\otimes)} - n \cdot \eta a_n$$

$$= nD(\rho_{BR^n}^\otimes (\rho_{R}^\otimes \otimes \rho_{R'}^\otimes) + n \cdot a_n \sqrt{4V(\rho_{BR^n}^\otimes (\rho_{R}^\otimes \otimes \rho_{R'}^\otimes)} - n \cdot \eta a_n,$$

where the first inequality follows due to Lemma III.9 since for sufficiently large $n$, we will have $\epsilon_n \in (0, k^{-1/2}]$ for any choice of $k > 1$ as well as the fact that $\log f(\epsilon_n) = n \cdot o(a_n)$, the second inequality is due to the AEP for the hypothesis testing relative entropy (Proposition III.10), a property of moderate deviation sequences (Lemma III.5), and the boundedness of the relative entropy variance (Lemma B.6). The final equality uses the duality of the relative entropy and relative entropy variance (Lemma B.9). By choosing $n^* = \max(n_1, n_2^*)$, the proposition follows.

The AEP for the partially smoothed max-information (Proposition III.11) is strengthened straightforwardly in the following corollary by noting that constant multiplicative factors on the error $\epsilon_n$ do not affect the moderate deviation analysis and multiplicative factors that are polynomial in $n$ do not affect the strict moderate deviation analysis. Both results hold due to a property of moderate deviation sequences (Lemma III.5) and the boundedness of the relative entropy variance (Lemma B.6).

**Corollary III.12** Proposition III.11 holds for $0 < \epsilon'_n \leq 1$ with $\epsilon'_n = \Theta(e^{-a_n^2})$ and a moderate sequence $\{a_n\}$. That is, for $\eta > 0$ there exists $n^* \in \mathbb{N}$ such that for $n \geq n^*$ we have

$$I(B : R)_\rho + a_n \sqrt{4V(B : R)_\rho - \eta a_n} \leq I_{\max}(R^n) \leq I(B : R)_\rho + a_n \sqrt{4V(B : R)_\rho + \eta a_n}.$$

Moreover, for a strictly moderate sequence $\{a_n\}$ and $\epsilon'_n = \text{poly}(n)e^{-a_n^2}$, for $\eta > 0$, there exists $n^* \in \mathbb{N}$ such that for $n \geq n^*$ we have

$$I(B : R)_\rho + a_n \sqrt{4V(B : R)_\rho - \eta a_n} \leq I_{\max}(R^n) \leq I(B : R)_\rho + a_n \sqrt{4V(B : R)_\rho + \eta a_n}.$$

**Proof of Theorem III.5.** The moderate deviation analysis of quantum state splitting (Theorem III.6) is now straightforward. For any purification $\rho_{ABR}$ of $\rho_{AB}$, we obtain the cost of one-shot quantum state splitting (Theorem III.3). By combining the AEP for the partially smoothed max-information (Proposition III.11) and noting that constant multiplicative factors do not change the communication cost (Corollary III.12), the claim follows.
E. Equivalence: Quantum state merging

Quantum state merging can be understood as a time reversed version of a quantum state splitting [2]. Alice and Bob start with a state $\rho_{ABR}$, where Alice holds the $A$ quantum register, Bob holds the $B$ quantum register and $R$ is inaccessible. The goal is to use pre-shared entanglement and one-way communication to send the $B$ quantum register to Alice. An equivalent statement to the cost of one-shot quantum state splitting (Theorem III.3) was shown for one-shot quantum state merging in [5, Theorem 6]. Hence, our moderate deviation analysis for quantum state splitting (Theorem III.6) also holds for quantum state merging.

IV. QUANTUM SOURCE CODING

Here, we investigate the task of quantum source coding in the moderate deviation regime. There are two main settings for source coding, namely the ensemble and purification settings based on whether one uses the fidelity or entanglement fidelity to measure the success of source coding. The ensemble setting is further divided into the blind and visible settings, where the state to be compressed is either known or unknown. More precise definitions of these settings can be found in the literature [31, 33, 34, 56]. Here, we consider quantum source coding in the blind purification scheme, which is equivalent to quantum state splitting with a trivial $A$ quantum register and no entanglement-assistance.

Definition IV.1 (Quantum source coding) Let $\rho_B \in \mathcal{S}(B)$ and $\varepsilon \in [0,1]$. A one-shot quantum source coding protocol consists of:

1. Quantum registers $Q$ and $A_1 \cong B$
2. An encoding quantum channel $\mathcal{E}_{A_1 \rightarrow Q}$
3. A decoding quantum channel $\mathcal{D}_{Q \rightarrow B}$.

A $\{q,\varepsilon\}$-one-shot quantum source coding protocol of $\rho_B$ is such that $q = \log |Q|$ with
\[
(D \circ \mathcal{E})(\rho_{A_1 R}) \approx_{\varepsilon} \rho_{BR} \text{ for any extension } \rho_{A_1 R} \text{ of } \rho_{A_1}.
\]

The minimal quantum communication cost of source coding of $\rho_B$ is defined as
\[
q^*_\varepsilon(\rho_B) = \min \left\{ q \in \mathbb{N} : \exists \text{ a } \{q,\varepsilon\}\text{-one-shot quantum source coding protocol of } \rho_B \right\}.
\]

We start with the following lemma about the mutual information variance for pure states.

Lemma IV.2 (Pure state mutual information variance) For $\rho_{AB} \in \mathcal{S}(AB)$ pure, the mutual information variance and the conditional entropy variance $V(\rho_{AB}||I_A \otimes \rho_B)$ take the form
\[
V(A:B)_{\rho} = 4V(A)_{\rho}, \quad V(\rho_{AB}||I_A \otimes \rho_B) = V(A)_{\rho},
\]
respectively.

Proof. From the definition of the relative entropy variance, we have
\[
V(\rho_{AB}||\rho_A \otimes \rho_B) = \text{Tr} (\rho_{AB}(\log \rho_{AB})^2) + \text{Tr} (\rho_A(\log \rho_A)^2) + \text{Tr} (\rho_B(\log \rho_B)^2) + 2\text{Tr} (\rho_{AB} \log \rho_A \otimes \log \rho_B)
\]
\begin{align}
- 2\text{Tr} (\rho_{AB} \log \rho_{AB} \log \rho_A) - 2\text{Tr} (\rho_{AB} \log \rho_{AB} \log \rho_B) - D(\rho_{AB} \| \rho_A \otimes \rho_B)^2 \\
= \text{Tr} (\rho_A (\log \rho_A)^2) + \text{Tr} (\rho_B (\log \rho_B)^2) + 2\text{Tr} (\rho_{AB} \log \rho_A \otimes \log \rho_B) - (S(A)_\rho + S(B)_\rho)^2 \\
= 4V(A)_\rho,
\end{align}

where in the second equality we have removed various terms which are zero due to \(\rho_{AB}\) being pure, and in the third equality we have used that the marginals \(\rho_A\) and \(\rho_B\) share the same eigenvalues and that \(\text{Tr} (\rho_{AB} \log \rho_A \otimes \log \rho_B) = \text{Tr} (\rho_A (\log \rho_A)^2)\) for pure \(\rho_{AB}\). The last claim can be verified by expanding the left hand side in the Schmidt basis of \(\rho_{AB}\). A similar argument shows the second equality in the lemma. ■

We now consider quantum source coding in the moderate deviation regime. Note that quantum source coding in the large deviation setting was explored in [31, 33] using the information spectrum method. Remark 16 of [47] also touches on this connection between source coding and information spectrum. Classical source coding in the moderate deviation regime was investigated in [37]. Some of our results below may also be proved using a different route, namely the moderate deviation expansion of the information spectrum entropy [24, Section 3.7]. The minimal communication cost of quantum source coding in the low-error moderate deviation regime is characterized as follows.

**Theorem IV.3 (Source coding moderate deviation)** The minimal quantum communication cost of one-shot \(\varepsilon_n\)-error source coding of \(\rho_B^{\otimes n}\) for \(\rho_B \in S(B)\) with \(\varepsilon_n = e^{-na_n^2}\) for a moderate sequence \(\{a_n\}\) is given by

\[
\frac{1}{n} q_{\varepsilon_n}^* (\rho_B^{\otimes n}) = S(B)_\rho + 2a_n \sqrt{V(B)_\rho} + o(a_n).
\]

**Proof.** The converse statement follows as a corollary of our moderate deviation analysis of state splitting (Theorem III.6) and properties of the mutual information variance for pure states (Lemma IV.2). While a related achievability statement also follows from those results, this additionally assumes the presence of entanglement-assistance between the encoder and decoder which is not present in Definition IV.1. Thus, we use a different proof technique that does not require entanglement-assistance. Namely, we employ the bounds shown in [1, 22], where the one-shot quantum source coding error is quantified in terms of the fidelity. We modify their result for our setting in purified distance. For \(\varepsilon_n = e^{-na_n^2}\) for a moderate sequence \(\{a_n\}\), we have

\[
\frac{1}{n} q_{\varepsilon_n}^* (\rho_B^{\otimes n}) \leq \frac{1}{n} H_s^2 (\rho_B^{\otimes n}) = -\frac{1}{n} D_{\rho_B^{\otimes n}}^\varepsilon (\rho_B^{\otimes n} || I_B^n) \\
\leq -\frac{1}{n} D_{\rho_B^{\otimes n}}^{\varepsilon^2/4} (\rho_B^{\otimes n} || I_B^n) - \frac{1}{n} \log \varepsilon^2 \frac{1}{4} \\
= -D(\rho_B || I_B) + a_n \sqrt{4V(\rho_B || I_B)} + o(a_n) \\
= D(\rho_{BR} || I_B \otimes \rho_R) + a_n \sqrt{4V(\rho_{BR} || I_B \otimes \rho_R)} + o(a_n) \\
= S(B)_\rho + 2a_n \sqrt{V(B)_\rho} + o(a_n),
\]

where the first inequality is due to the achievability bound in [22, Theorem 5.5 (ii)], the first equality is from the definition of the information spectrum entropy, the second inequality is from the bound between the hypothesis testing relative entropy and the information spectrum relative entropy (Lemma B.7), the second equality is due to the AEP for the hypothesis testing relative entropy (Proposition III.10) and because multiplicative factors on the error do not change the communication cost in the moderate deviation regime (Corollary III.12), and the third equality is
through the duality of the relative entropy (Lemma B.9) and the final equality is due to properties of the mutual information variance for pure states (Lemma IV.2).

We note that the converse proof allows for entanglement-assistance. Hence, it follows that entanglement-assistance does not reduce the quantum communication cost of low-error moderate deviation quantum source coding.

In the large error regime, we set $\varepsilon_n = 1 - e^{-na_n^2}$ for a moderate sequence $\{a_n\}$. We find the following achievability result.

**Proposition IV.4 (Source coding high-error moderate deviation)** The minimal quantum communication cost of one-shot $\varepsilon_n$-error source coding of $\rho_B^{\otimes n}$ for $\rho_B \in S(B)$ with $\varepsilon_n = 1 - e^{-na_n^2}$ for a moderate sequence $\{a_n\}$ is bounded as

$$\frac{1}{n} q^*_\varepsilon(\rho_B^{\otimes n}) \leq S(B)_{\rho} - a_n \sqrt{V(B)_{\rho}} + o(a_n).$$

**Proof.** We start with the achievability result. We have

$$\frac{1}{n} q^*_\varepsilon(\rho_B^{\otimes n}) \leq \frac{1}{n} H_s^{1/2}(\rho_B^{\otimes n}) = \frac{1}{n} H_s^{1/2}(\rho_B^{\otimes n})$$

$$\leq -\frac{1}{n} D_s(\rho_B^{\otimes n} \| \rho_B) - \frac{1}{n} \log \delta$$

$$= -D(\rho_B || I_B) - a_n \sqrt{V(\rho_B || I_B)} + o(a_n)$$

$$= S(B)_{\rho} - a_n \sqrt{V(B)_{\rho}} + o(a_n)$$

where the first inequality is due to the achievability result [1, Eq.(31)], the first equality substitutes $\varepsilon_n = 1 - e^{-na_n^2}$, the second equality is from the definition of the information spectrum entropy, the second inequality is from the bound between the hypothesis testing relative entropy and the information spectrum relative entropy (Lemma B.7), and the third inequality is due to the monotonicity of the hypothesis testing relative entropy in the $\varepsilon$-parameter. The third equality is due to the AEP for the hypothesis testing relative entropy (Proposition III.10), because multiplicative factors on the error do not change the communication cost in the moderate deviation regime (Corollary III.12), and the choice of $\delta = \sqrt{e^{-na_n^2}}$. The fourth equality is through the duality of the relative entropy (Lemma B.9), and the final equality is due to the properties of the mutual information variance for pure states (Lemma IV.2).

Based on the following discussion, we conjecture that the achievability statement in Proposition IV.4 is also tight. For $\varepsilon \in (0, 1)$ and $\rho_B \in S(B)$, it was shown in the second-order analysis of quantum source coding that [1]

$$\frac{1}{n} q^*_\varepsilon(\rho_B^{\otimes n}) = S(B)_{\rho} + \sqrt{\frac{V(B)_{\rho}}{n}} \phi^{-1}(\sqrt{1 - \varepsilon^2}) + O\left(\frac{\log n}{n}\right),$$

where $\phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We may substitute $\varepsilon = e^{-na_n^2}$ or $\varepsilon = 1 - e^{-na_n^2}$ and use the approximation $\phi^{-1}(\delta) \approx -\sqrt{-2 \log \delta}$ for small $\delta$ and that $\phi^{-1}(\delta) = -\phi^{-1}(1 - \delta)$ for $\delta \in [0, 1]$. With this approximation and ignoring higher order error terms, we recover the low-error regime result in Theorem IV.3, as well as the high-error regime achievability result in Proposition IV.4.
V. QUANTUM CHANNEL SIMULATION

Quantum state splitting is the underlying primitive to understand the task of quantum channel simulation. In this setting, Alice and Bob have access to pre-shared resource states and noiseless one-way communication which they will use to simulate i.i.d. copies of a noisy quantum channel. This problem is also known as the Quantum Reverse Shannon Theorem in the literature [10, 13]. We start with the measure we need to distinguish between two quantum channels. It is based on the channel fidelity $F(E,F) = \max_{\rho} F((E \otimes I_R)(\rho), (F \otimes I_R)(\rho))$, but we choose to work with the purified distance instead of the fidelity.

Definition V.1 (Channel purified distance) For any pair of quantum channels $E_{A\rightarrow B}$ and $F_{A\rightarrow B}$, the channel purified distance is defined as

$$P(E, F) = \sup_{\rho \in S_{\leq} (A \otimes R)} P((E_{A\rightarrow B} \otimes I_R)(\rho), (F_{A\rightarrow B} \otimes I_R)(\rho))$$  \hspace{1cm} (88)

where the supremum is over quantum registers $R$.

We now comment on a few properties that allow us to simplify Definition V.1. We can choose $|R| = |A|$ without loss of generality, as shown in Lemma B.2. The supremum over $\rho$ can then be replaced by a maximum since the set of subnormalized quantum states is closed and bounded for fixed $R$ and the purified distance is a continuous function (since continuity of the one-norm implies continuity of the fidelity). By Lemma B.3, we have that normalized states will achieve the maximum. Next, since the purified distance is non-increasing under partial trace, we may choose $\rho_{AR}$ to be pure. Finally, we remark that the channel purified distance is a metric since it inherits this property from the purified distance.

One-shot quantum channel simulation is as follows.

Definition V.2 (One-shot quantum channel simulation) Let $N_{A\rightarrow B}$ be a quantum channel, $\varepsilon \in [0,1]$, and take a quantum register $Q$ with $\log |Q| = q$. A $\{q, \varepsilon\}$-one-shot quantum channel simulation of $N_{A\rightarrow B}$ consists of

1. A resource $\sigma_{KL} \in S(KL)$
2. An encoding quantum channel $E_{AK\rightarrow Q}$
3. A decoding quantum channel $D_{QL\rightarrow B}$

such that we have for the composite quantum channel

$$N'_{A\rightarrow B} = D \circ E \circ P^\sigma = (D \circ E)(\cdot \otimes \sigma_{KL}) \text{ that } P(N, N') \leq \varepsilon.$$  \hspace{1cm} (89)

The minimal quantum communication cost of simulating $N_{A\rightarrow B}$ is defined as

$$q^*_\varepsilon(N) = \min \left\{ q : \exists a \{q, \varepsilon\}-one-shot quantum channel simulation of N \right\}. \hspace{1cm} (90)$$

When the channel being simulated is of the form $N^\otimes n$, we can make use of de Finetti reductions and symmetrization to achieve a one-shot quantum channel simulation protocol in terms of a one-shot quantum state splitting protocol. We can then employ our quantum state splitting results to characterize the minimal communication cost of quantum channel simulation in the low-error strictly moderate deviation setting. Our main result in this section is as follows.

---

3 Since free entanglement can be included in the resource state $\sigma_{KL}$, we equivalently have that the classical communication required is $2q$. 

Theorem V.3 (Channel simulation moderate deviation) For any quantum channel $N_{A\rightarrow B}$, the minimal quantum communication cost for $N^\otimes n$ with error $\varepsilon_n = e^{-n\alpha_n^*}$ for a strictly moderate sequence $\{a_n\}$ is

$$\frac{1}{n} q_n^* (N^\otimes n) = C(N) + a_n \sqrt{V_{\max}(B : R)} + o(a_n), \quad (91)$$

where $V_{\max}(B : R)$ was defined in Eq. (8).

For the rest of this section, we go through the technical tools for the proof of Theorem V.3. The following lemma about symmetrizing quantum protocols is standard in the literature (see, e.g., [10, 13]). However, we have not found a formal version, which is why we state it here and give a proof in Appendix B.

Lemma V.4 (Symmetrized protocol for state splitting) Let $\pi$ be a permutation of a set of $n$ elements. Let $\rho_{A^n B^n} \in S(A^n B^n)$ such that for any permutation quantum channel $\pi_{A^n} \otimes \pi_{B^n}: A^n B^n \rightarrow A^n B^n$ where $\pi_{A^n}(\pi_{B^n})$ applies the permutation $\pi$ to the $A^n(B^n)$ registers, we have

$$\pi_{A^n} \otimes \pi_{B^n}(\rho_{A^n B^n}) = \rho_{A^n B^n}. \quad (92)$$

Furthermore, let there be a $\{q, \varepsilon\}$-one-shot quantum state splitting protocol of $\rho_{A^n B^n}$ given in terms of the quantum channel $T_{A^n A^n_1 : A^n B^n}$ and some resource state $\sigma$. Then, there exists a one-shot quantum state splitting protocol given in terms of a quantum channel $T_{A^n A^n_1 : A^n B^n}$ such that

$$T_\omega \circ \pi_{A^n} \otimes \pi_{A^n_1} = \pi_{A^n} \otimes \pi_{B^n} \circ T_\omega \quad (93)$$

for any permutation $\pi$ on $n$ quantum registers, $\omega = \sigma \otimes \sigma'$, and $\sigma'$ is a state containing shared randomness of dimension $n!$. The quantum channel $T_{A^n A^n_1 : A^n B^n}$ satisfies

$$P(\rho_{A^n B^n R}, (T_\omega \otimes I_R)((\rho_{A^n A^n_1 R})) \leq P(\rho_{A^n B^n R}, (T_\omega \otimes I_R)((\rho_{A^n A^n_1 R})) \quad (94)$$

for any extension $\rho_{A^n B^n R}$ of $\rho_{A^n B^n}$.

We now state the achievability and converse statements for one-shot channel simulation of $n$-fold i.i.d. channels. Both are based on well-known arguments, but adapted to our exact setting.

Proposition V.5 (Channel simulation achievability) Let $N_{A\rightarrow B}$ be a quantum channel and $\varepsilon \in (0, 1)$, and $n \in \mathbb{N}$. Then, we have

$$q_n^* (N^\otimes n) \leq \max_{\phi_{AR}} \frac{1}{2} I_{\max}(\hat{R}^n; B^n)(\phi_\otimes (\phi) \otimes n) + g(\varepsilon, |A|, n), \quad (95)$$

where $\phi_{AR} \in S(AR)$ pure with $R \cong A$, and we have the fudge term

$$g(\varepsilon, |A|, n) = \frac{1}{2} \log \left( \frac{2}{(\varepsilon^2/2)^2} + 2 \right) + 2(|A|^2 - 1) \log(n + 1) + \frac{1}{2} \log \left( \frac{2}{(\varepsilon^2/24)^2} + 2 \right) \quad (96)$$

$$+ \frac{1}{2} \log \left( \frac{8 + (|A|^2)^2}{(|A|^2)^2} \right) + \log \frac{4}{\varepsilon^2} \quad \text{for} \quad \varepsilon = \frac{\varepsilon'}{\sqrt{2}} \left( n + 1 \right)^{1 - |A|^2}. \quad (97)$$

The mostly standard proof is included in Appendix B. The converse is as follows.
Proposition V.6 (Channel simulation converse) Let $\mathcal{N}_{A\rightarrow B}$ be a quantum channel and $\varepsilon \in (0,1)$, and $n \in \mathbb{N}$. Then, we have

$$q^*_n(\mathcal{N}^\otimes n) \geq \max_{\sigma_{AR}} \frac{1}{2} I^\varepsilon_{\max}(\hat{R}^n; B^n)_{((\mathcal{N} \otimes I)(\sigma))^\otimes n}$$

(98)

with $\phi_{AR} \in \mathcal{S}(AR)$ is pure and $R \cong A$.

Proof. Let $\mathcal{T}$ be a $\{q, \varepsilon\}$-one-shot quantum channel simulation of $\mathcal{N}^\otimes n$. Hence, $P(\mathcal{T}, \mathcal{N}^\otimes n) \leq \varepsilon$ and for any pure i.i.d. state $\phi_{AR}^\otimes n$ we have $P((\mathcal{T} \otimes I^n)(\phi_{AR}^\otimes n), (\mathcal{N}^\otimes n \otimes I^n)(\phi_{AR}^\otimes n)) \leq \varepsilon$. That is, $\mathcal{T}$ is a $\{q, \varepsilon\}$-one-shot quantum state splitting protocol for $\mathcal{N}_{A\rightarrow B}^\otimes n$. Given the cost of one-shot quantum state splitting (Theorem III.3) and that the smooth-max information is non-increasing under partial trace, we have that the minimal quantum communication cost for quantum channel simulation is lower bounded by $\frac{1}{2} I^\varepsilon_{\max}(\hat{R}^n; B^n)_{((\mathcal{N} \otimes I)(\sigma))^\otimes n}$. As this holds for any $\phi_{AR} \in \mathcal{S}(AR)$ pure, the claim follows. 

We now have all the ingredients to prove Theorem V.3 about quantum channel simulation in the moderate deviation regime.

Proof of Theorem V.3: Recall that the minimal quantum communication cost of a simulation of $\mathcal{N}^\otimes n$ with error $\varepsilon_n = e^{-na_n^2}$ for a strictly moderate sequence $\{a_n\}$ is denoted as $q^*_n(\mathcal{N}^\otimes n)$. We now combine the achievability result for one-shot channel simulation (Proposition V.5) with the AEP for the partially smoothed max-information (Proposition III.11), and note that multiplicative factors on the error do not change the communication cost in the moderate deviation regime (Corollary III.12). Consequently, we find for $\eta > 0$ that there exists an $n_1^* \in \mathbb{N}$ such that for $n \geq n_1^*$ we have

$$\frac{1}{n} q^*_n(\mathcal{N}^\otimes n) \leq \max_{\sigma} \left\{ \frac{1}{2} I(B : R)_{(\mathcal{N} \otimes I)(\sigma)} + a_n \sqrt{V(B : R)_{(\mathcal{N} \otimes I)(\sigma)}} + \eta a_n \right\}.$$  

(99)

Since the set of quantum states is a compact metric space, we can apply Lemma B.15 taken from [50] to obtain

$$\frac{1}{n} q^*_n(\mathcal{N}^\otimes n) \leq C(\mathcal{N}) + \max_{\sigma' \in \Pi(\mathcal{N})} a_n \sqrt{V(B : R)_{(\mathcal{N} \otimes I)(\sigma')}} + o(a_n).$$  

(100)

Next, we combine the converse result for one-shot quantum channel simulation for i.i.d. channels (Proposition V.6) with the AEP for the partially smoothed max-information (Proposition III.11). For $\eta > 0$, there exists $n_2^* \in \mathbb{N}$ such that for $n \geq n_2^*$, we have

$$\frac{1}{n} q^*_n(\mathcal{N}^\otimes n) \geq \max_{\sigma_{AR}} \left\{ \frac{1}{2} I(B : R)_{(\mathcal{N} \otimes I)(\sigma)} + a_n \sqrt{V(B : R)_{(\mathcal{N} \otimes I)(\sigma)}} - \eta a_n \right\}$$

$$\geq C(\mathcal{N}) + \max_{\sigma' \in \Pi(\mathcal{N})} a_n \sqrt{V(B : R)_{(\mathcal{N} \otimes I)(\sigma')}} + o(a_n).$$

(101)

(102)

VI. ENTANGLEMENT-ASSISTED QUANTUM CHANNEL CODING

Here, we characterize the cost of one-shot entanglement-assisted channel coding in the high-error moderate deviation regime. We consider entanglement-assisted quantum communication in this section. However, the results of this section apply to entanglement-assisted classical communication.
as well. This is because superdense coding and teleportation establish and equivalence between these two tasks when Alice and Bob have access to shared entanglement.

Our analysis of this task relies on the close connection between high-error channel coding and low-error channel simulation. In the classical setting, this connection was explored in [8]. A similar observation has been done for the relation between source coding and the random number generation in [32].

**Definition VI.1 (One-shot entanglement-assisted channel coding)** Consider a quantum channel \( N_{A' \to B'} , \varepsilon \in [0,1] \), and quantum registers \( A \cong B \) with \( r = \log |A| \). A \( \{r, \varepsilon\}\)-one-shot quantum channel coding protocol for \( N \) consists of

1. Registers \( K, L \) and a resource \( \sigma_{KL} \in S(KL) \)
2. An encoding quantum channel \( E_{AK} \to A' \)
3. A decoding quantum channel \( D_{B'L} \to B \)

giving for the composite channel

\[
\tilde{N}_{A \to B} = D \circ N \circ \mathcal{E} \circ P^\sigma = D \circ N \circ \mathcal{E}(\cdot \otimes \sigma) \quad \text{that} \quad P\left(\tilde{N}_{A \to B}, \mathcal{I}_{A \to B}\right) \leq \varepsilon .
\]

The maximum coding rate of \( N_{A' \to B'} \) is defined as

\[
r^*_\varepsilon(N) = \max \{ r : \exists \ a \{r, \varepsilon\}-one-shot \ quantum \ channel \ coding \ protocol \ for \ N \} .
\]

Our result in this section is the following characterization.\(^4\)

**Theorem VI.2 (Channel coding moderate deviation rate)** Let \( N_{A \to B} \) be a quantum channel, \( n \in \mathbb{N} \), and \( \varepsilon_n = e^{-na^2_n} \) for a strictly moderate sequence \( \{a_n\} \). Then, we have

\[
\frac{1}{n} r^*_{1-\varepsilon_n}(N^\otimes n) = C(N) + \frac{a_n}{\sqrt{2}} \sqrt{V_{\max}(B : R)_N} + o(a_n).
\]

Our proof of Theorem VI.2 proceeds by first upper bounding the rate of high-error channel coding with the cost of low-error channel simulation. We start by giving some notation for this section, following the definitions from the work [22] on entanglement-assisted quantum channel coding for sending classical information. For any quantum channel \( N \), let us define \( P_{\text{succ}}(N, r) \) to be the maximum probability of correctly transmitting a uniformly random message made of \( r \) classical bits, where the maximization is over all encoders and decoders. We also define \( P_{\text{fail}}(N, r) = 1 - P_{\text{succ}}(N, r) \). Finally, let \( \log M^*_{\text{ea}}(N, \varepsilon) \) be the maximum number of bits of information that can be transmitted through the channel \( N \) using an entanglement-assisted protocol with \( P_{\text{fail}}(N, \log M^*_{\text{ea}}(N, \varepsilon)) \) bounded by \( \varepsilon \in (0, 1) \).

**Lemma VI.3 (Coding converse via simulation)** Let \( N_{A' \to B'} \) be a quantum channel, \( n \in \mathbb{N} \), and \( \varepsilon \in (0,1] \). Then, we have

\[
r^*_{1-\varepsilon}(N^\otimes n) \leq q^*_{\varepsilon}(N^\otimes n) + O\left(\log \frac{1}{\varepsilon}\right).
\]

where \( q^*_{\varepsilon}(N^\otimes n) \) is the minimal quantum communication cost of simulating \( N^\otimes n \) as given in Definition V.2.

\(^4\) A complementary result for classical communication over quantum channels is given in [18].
Proof. Consider the identity channel \( I_{A \rightarrow B} \) with \( A \cong B \) and log \( |A| = q_\phi^\star(\mathcal{N} \otimes^n) \). This channel can be used for the communication required for a quantum channel simulation of \( \mathcal{N} \otimes^n \) by Definition VI.2. Let the simulated channel be called \( T_{A^n \rightarrow B^n} \), for which we have \( P(T, \mathcal{N} \otimes^n) \leq \sqrt{\varepsilon} \).

Now, consider an entanglement-assisted quantum channel coding protocol for \( \mathcal{N} \otimes^n \) with maximum coding rate \( r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n) \). There exist quantum channels \( D, E \) and a preparation quantum channel \( \mathcal{P}^\sigma \) such that for \( \mathcal{N}^n = D \circ \mathcal{N} \otimes^n \circ E \circ \mathcal{P}^\sigma \), we have \( P(\mathcal{N}^n, I_{A \rightarrow B}) \leq 1 - \varepsilon \) for \( A \cong B \) and log \( |A| = r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n) \).

Next, let us now define \( \hat{T} = D \circ T \circ E \circ \mathcal{P}^\sigma \). Using the tighter triangle inequality for the purified distance (Lemma III.7) and that the purified distance is non-increasing under quantum channels, we have

\[
P(\hat{T}, I_{A \rightarrow B}) \leq (1 - \varepsilon)\sqrt{1 - \varepsilon} + \sqrt{\varepsilon} \sqrt{1 - (1 - \varepsilon)^2} \leq 1 - \frac{3}{2} \varepsilon + O(\varepsilon^{3/2}),
\]

(107)

Since we have \( P(\hat{T}, I_{A \rightarrow B}) \leq 1 - \frac{3}{2} \varepsilon + O(\varepsilon^{3/2}) \), we can use \( \hat{T} \) to send \( r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n) \) bits with \( P_{\text{fail}}(\hat{T}, r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n)) \leq 1 - \frac{3}{2} \varepsilon + O(\varepsilon^{3/2}) \). The strong converse of the capacity of the quantum identity channel (Lemma B.16) states that if the rate exceeds the capacity, the failure probability of correctly decoding approaches 1 exponentially, that is

\[
P_{\text{fail}}(\hat{T}, r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n)) \geq 1 - 2^{-r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n) - q_\sqrt{\varepsilon}(\mathcal{N} \otimes^n)}
\]

(108)

\[
\implies 1 - \frac{3}{2} \varepsilon + O(\varepsilon^{3/2}) \geq 1 - 2^{-r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n) - q_\sqrt{\varepsilon}(\mathcal{N} \otimes^n)}
\]

(109)

\[
\implies r_{1-\varepsilon}^\star(\mathcal{N} \otimes^n) \leq q_\sqrt{\varepsilon}(\mathcal{N} \otimes^n) + O\left(\log \frac{1}{\varepsilon}\right).
\]

(110)

We now have the converse statement for quantum channel coding.

Proposition VI.4 (Channel coding moderate deviation converse) Let \( \mathcal{N}_{A \rightarrow B} \) be a quantum channel, \( n \in \mathbb{N} \) and \( \varepsilon_n = e^{-na_2^2} \) for a strictly moderate sequence \( \{a_n\} \). Then, we have that

\[
\frac{1}{n} r_{1-\varepsilon_n}^\star(\mathcal{N} \otimes^n) \leq C(\mathcal{N}) + \frac{a_n}{\sqrt{2}} \sqrt{\lambda_{\max}^\star(B : R)^n} + o(a_n).
\]

(111)

Proof. We substitute \( \varepsilon = e^{-na_2^2} \) for a strictly moderate sequence \( \{a_n\} \) in the coding converse via simulation (Lemma VI.3) and the result then follows by our results for channel simulation in the moderate deviation regime (Theorem V.3).

Note that existing techniques such as the meta-converse of [44] yield the same bound for covariant quantum channels (as shown in Appendix C), but our converse statement as above extends that result to all channels.

Next, we show the achievability in Theorem VI.2. The main idea in the following is that if one can send classical messages using a quantum channel and entanglement-assistance, then one can also achieve entanglement-assisted quantum channel coding in the sense of Definition VI.1 through a teleportation protocol.

Lemma VI.5 (Teleportation protocol for channel coding) Consider a quantum channel \( \mathcal{N}_{A' \rightarrow B'} \), \( \varepsilon \in [0, 1] \), and quantum registers \( A \cong B \) with log \( |A| = \frac{1}{2} \log M_{\text{eh}}^\star(\mathcal{N}, \varepsilon) \). There exist

1. registers \( K, L \) and shared entanglement through a resource \( \sigma_{KL} \in S(KL) \)

2. an encoding quantum channel \( E_{AK \rightarrow A'} \)


3. a decoding quantum channel $\mathcal{D}_{B'\rightarrow B}$
such that for the quantum channel

$$\tilde{N}_{A\rightarrow B} = \mathcal{D} \circ \mathcal{N} \circ \mathcal{E} \circ \mathcal{P}^\sigma = \mathcal{D} \circ \mathcal{N} \circ \mathcal{E}(\cdot \otimes \sigma) \quad \text{we have} \quad P\left(\tilde{N}_{A\rightarrow B}, \mathcal{I}_{A\rightarrow B}\right) \leq \sqrt{\varepsilon}. \quad (112)$$

**Proof.** Let $X$ be a classical register with $|X| = \log M_{ea}^*(\mathcal{N}, \varepsilon)$. Let $\omega_{K'L'}$ be a resource state used in the entanglement-assisted classical communication protocol for $\mathcal{N}_{A'\rightarrow B'}$. By the definition of $\log M_{ea}^*(\mathcal{N}, \varepsilon)$, there exists an encoder $\mathcal{E}'_{XK'\rightarrow A'}$ and a decoder $\mathcal{D}'_{L'B'\rightarrow Y}$, where $Y \cong X$ such that we can construct the composite channel $\mathcal{D}' \circ \mathcal{N} \circ \mathcal{E}'(\cdot \otimes \omega)$. If the state of register $X$ is $x \in \mathcal{X}$ where $x$ chosen uniformly at random from alphabet $\mathcal{X}$, it holds that $y = \mathcal{D}' \circ \mathcal{N} \circ \mathcal{E}'(x \otimes \omega)$ satisfies $\Pr(y = x) = P_{\text{succ}}(\mathcal{N}, \log M_{ea}^*(\mathcal{N}, \varepsilon)) \geq 1 - \varepsilon$.

Let $|\phi\rangle \langle \phi|_{KL} \in \mathcal{S}(KL)$ be a $\frac{1}{2} \log M_{ea}^*(\mathcal{N}, \varepsilon)$-dimensional maximally entangled state. There exists an encoder $\mathcal{E}_{AK \rightarrow X}$ and a decoder $\mathcal{D}_{LY \rightarrow B}$ that is used to teleport the state in register $A$ to register $B$ [9].

The channel $\tilde{N}_{A\rightarrow B}$ now combines the above elements. It uses the resource state $\sigma_{KL} = \omega_{K'L'} \otimes |\phi\rangle \langle \phi|_{KL}$, the encoder $\mathcal{E}_{AK \rightarrow A'} = \mathcal{E}'_{XK'\rightarrow A'} \circ \mathcal{E}_{AK \rightarrow X}$ and the decoder $\mathcal{D}_{B'\rightarrow B} = \mathcal{D}_{LY \rightarrow B} \circ \mathcal{D}'_{L'B'\rightarrow Y}$. For a fixed quantum state, the classical message that needs to be transmitted in the teleportation protocol is uniformly random and hence the teleportation of any state also occurs with success probability $p = P_{\text{succ}}(\mathcal{N}, \log M_{ea}^*(\mathcal{N}, \varepsilon)) \geq 1 - \varepsilon$. We now apply the above protocol to the $A$ register of any pure state $|\psi\rangle \langle \psi|_{AR}$, where $R \cong A$. The output state is $p|\psi\rangle \langle \psi|_{AR} + (1-p)|\psi^\perp\rangle \langle \psi^\perp|_{AR}$. We have

$$F(p|\psi\rangle \langle \psi|_{AR} + (1-p)|\psi^\perp\rangle \langle \psi^\perp|_{AR}, |\psi\rangle \langle \psi|_{AR}) \geq F((1-\varepsilon)|\psi\rangle \langle \psi|_{AR} + \varepsilon|\psi^\perp\rangle \langle \psi^\perp|_{AR}, |\psi\rangle \langle \psi|_{AR}) \geq \sqrt{1-\varepsilon}. \quad (113)$$

In terms of purified distance, we have

$$P((p|\psi\rangle \langle \psi|_{AR} + (1-p)|\psi^\perp\rangle \langle \psi^\perp|_{AR}, |\psi\rangle \langle \psi|_{AR}) \leq \sqrt{\varepsilon} \implies P\left(\tilde{N}_{A\rightarrow B}, \mathcal{I}_{A\rightarrow B}\right) \leq \sqrt{\varepsilon}, \quad (115)$$

where the last implication follows since we can apply the protocol to any pure quantum state $|\psi\rangle \langle \psi|_{AR}$. ■

Entanglement-assisted classical communication over a quantum channel is a well-studied topic. An achievability result for $\log M_{ea}^*(\mathcal{N}^{\otimes n}, \varepsilon)$ is as follows.

**Lemma VI.6 (Entanglement-assisted classical coding achievability [23, Eq. 4.64])** Let $\mathcal{N}_{A\rightarrow B}$ be a quantum channel, $\varepsilon \in (0,1)$, $g(n, \mu) = 2^{-\frac{\varepsilon}{2}(\mu-|A| \log(n+1)) \over n}$, and $0 < 2\delta < \varepsilon - g(n, \mu)$.

Then, we have that

$$\log M_{ea}^*(\mathcal{N}^{\otimes n}, \varepsilon) \geq D_k^{\varepsilon-2\delta - g(n, \mu)} \left(\left(\mathcal{N} \otimes \mathcal{I}_R\right)(\psi_{AR})\right)^{\otimes n} \| (\mathcal{N}(\rho_A))^{\otimes n} \otimes \rho_R^{\otimes n} \right)^{\otimes n} \right) - f(\varepsilon, \delta) - \log \gamma_{n, \mu}, \quad (116)$$

where $f(\varepsilon, \delta) = \log \frac{1-\varepsilon}{\delta^2}$, $\gamma_{n, \mu} = (n+1)^{|A|2^\mu}$, and $\psi_{AR} \in \mathcal{S}(AR)$ pure.

We now have the achievability in Theorem VI.2.

**Proposition VI.7 (Channel coding moderate deviation achievability)** Let $\mathcal{N}_{A\rightarrow B}$ be a quantum channel, $n \in \mathbb{N}$, and $\varepsilon_n = e^{-na_n^2}$ for a strictly moderate sequence $\{a_n\}$. Then, we have that

$$\frac{1}{n} r_{1-\varepsilon}^*(\mathcal{N}^{\otimes n}) \geq C(\mathcal{N}) + \frac{a_n}{\sqrt{2}} \sqrt{V_{\max}(B : R)_{\mathcal{N}}} + o(a_n). \quad (117)$$
Proof. For some strictly moderate sequence \{a_n\}, let us choose \(\mu = \frac{2}{\ln^2 a_n^2 + |A| \log \frac{(n+1)}{n}}\), \(\delta = e^{-na_n^2}\), and \(\varepsilon = \sqrt{1 - e^{-na_n^2}}\). From this, we obtain \(g(n, \mu) = e^{-na_n^2}\), \(\log \gamma_{n, \mu} = 2|A| \log (n+1) + \frac{2}{\ln^2 a_n^2} = n \cdot o(a_n)\) and \(f(\varepsilon, \delta) = n \cdot o(a_n)\). For some sufficiently large \(n\), we then satisfy \(0 < 2\delta < \varepsilon^2 - g(n, \mu)\). Consequently, by Lemma VI.6 on the achievability of entanglement-assisted quantum channel coding for classical information transmission, we have for sufficiently large \(n\) that

\[
\log M_{\psi_n}(N^{\otimes n}, \varepsilon^2) \geq D_h^{e^{-2\delta} - g(n, \mu)} \left( (\langle N \otimes I_R \rangle (\psi_{AR}))^{\otimes n} \| (N (\rho A))^{\otimes n} \otimes \rho_R^{\otimes n} \right) - f(\varepsilon, \delta) - \log \gamma_{n, \mu}
\]

\[
= D_h^{1 - 3e^{-na_n^2}} \left( (\langle N \otimes I_R \rangle (\psi_{AR}))^{\otimes n} \| (N (\rho A))^{\otimes n} \otimes \rho_R^{\otimes n} \right) + n \cdot o(a_n)
\]

\[
\geq I(B : R)_{(N \otimes I)\psi} + \sqrt{2V(B : R)_{(N \otimes I)\psi}} + n \cdot o(a_n),
\]

where the last inequality holds due to [18, Theorem 1]. The inequality above holds for any pure quantum state \(\psi_{AR}\). We may choose \(\psi_{AR} = \arg \max_{\psi \in \Pi(N)} V(B : R)_{(N \otimes I)\psi}\), where \(\Pi(N)\) is the set of capacity achieving channel inputs defined in Eq. (9). By using the teleportation protocol for channel coding (Lemma VI.5), the claim follows.

Combining the converse (Proposition VI.4) and the achievability (Proposition VI.7), we obtain the moderate deviation analysis for channel coding (Theorem VI.2).

VII. CONCLUSION

In summary, we have resolved the quantum communication cost of several quantum information processing tasks in the moderate deviation regime. The task of quantum state splitting emerged as the fundamental primitive and is closely connected with quantum source coding. Our quantum state splitting results in the low-error moderate deviation regime are then used to characterize quantum channel simulation in the low-error strictly moderate deviation regime. In turn, quantum channel simulation is used to obtain a tight bound for entanglement-assisted quantum channel coding rates in the high-error strictly moderate deviation regime. This extends a result that was previously only available for a restricted class of channels (as reviewed in Appendix C). It would be interesting to explore if the findings in [29] based on sandwiched Rényi entropies could lead to similar moderate deviation results.

Several interesting problems remain open. Our proof technique for obtaining the AEP for the partially smoothed max-information (Proposition III.11) does not work for the large error moderate deviation regime, when the error goes as \(1 - e^{-na_n^2}\) for a moderate sequence \(\{a_n\}\). The missing technical ingredient appears to be a tighter bound on the partially smoothed max-information with the smoothed max-information (Lemma B.11). Hence, the problem of characterizing quantum state splitting in the high-error regime is open. In the case of quantum source coding, one can obtain a converse in the high-error regime using our quantum state splitting converse but this does not match the achievability statement in Proposition IV.4 that we conjecture to be tight. As such, that problem also remains open. Finally, the polynomial multiplicative factor on the error in the de Finetti reduction (Proposition B.14) is another obstacle that prevents us from characterizing channel simulation in the high-error strictly moderate deviation regime. It would be interesting to find an achievability proof for channel simulation without using state splitting and de Finetti reductions. This would then also yield a tighter converse bound for channel coding in the low-error strictly moderate deviation regime.

Lastly, we remark that our quantum state splitting result does not resolve the second-order regime, where the error is a constant. This is due to a gap in the achievability and converse
results in the AEP for the partially smoothed max-information (Proposition III.11). The technical ingredient required again appears to be a tighter relationship between the partially smoothed max-information and the smoothed max-information. In the moderate deviation analysis, this gap closes due to the fact that constant (poly($n$)) multiplicative factors on the error do not affect the cost of the (strict) moderate deviation analysis as explained in Lemma III.5. However, that no longer holds for the second-order regime.

VIII. ACKNOWLEDGEMENTS

This research is supported by the National Research Foundation, Prime Minister’s Office, Singapore and the Ministry of Education, Singapore under the Research Centres of Excellence programme. MT is also supported in part by NUS startup grants (R-263-000-E32-133 and R-263-000-E32-731).

[1] D. Abdelhadi and J. M. Renes. “On the Second-Order Asymptotics of the Partially Smoothed Conditional Min-Entropy & Application to Quantum Compression”. IEEE Journal on Selected Areas in Information Theory 1(2): 416–423 2020.
[2] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. “The mother of all protocols: restructuring quantum information’s family tree”. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 465(2108): 2537–2563 2009.
[3] Y. Altug and A. B. Wagner. “Moderate Deviations in Channel Coding”. IEEE Transactions on Information Theory 60(8): 4417–4426 2014.
[4] A. Anshu, M. Berta, R. Jain, and M. Tomamichel. “A minimax approach to one-shot entropy inequalities”. Journal of Mathematical Physics 60(12): 122201 2019.
[5] A. Anshu, M. Berta, R. Jain, and M. Tomamichel. “Partially Smoothed Information Measures”. IEEE Transactions on Information Theory 66(8): 5022–5036 2020.
[6] A. Anshu, V. K. Devabathini, and R. Jain. “Quantum Communication Using Coherent Rejection Sampling”. Physical Review Letters 119(12): 120506 2017.
[7] S. Arimoto. “On the converse to the coding theorem for discrete memoryless channels (Corresp.)”. IEEE Transactions on Information Theory 19(3): 357–359 1973.
[8] C. Bennett, P. Shor, J. Smolin, and A. Thapliyal. “Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem”. IEEE Transactions on Information Theory 48(10): 2637–2655 2002.
[9] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. “Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels”. Physical Review Letters 70(13): 1895 1993.
[10] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter. “The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels”. IEEE Transactions on Information Theory 60(5): 2926–2959 2014.
[11] C. H. Bennett and S. J. Wiesner. “Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states”. Physical Review Letters 69(20): 2881 1992.
[12] M. Berta. Quantum side information: Uncertainty relations, extractors, channel simulations. ETH Zurich 2013.
[13] M. Berta, M. Christandl, and R. Renner. “The Quantum Reverse Shannon Theorem Based on One-Shot Information Theory”. Communications in Mathematical Physics 306(3): 579–615 2011.
[14] H.-C. Cheng and L. Gao. “Error Exponent and Strong Converse for Quantum Soft Covering”. arXiv preprint arXiv:2202.10995 , 2022.
[15] H.-C. Cheng, E. P. Hanson, N. Datta, and M.-H. Hsieh. “Non-Asymptotic Classical Data Compression With Quantum Side Information”. IEEE Transactions on Information Theory 67(2): 902–930 2021.
[16] H.-C. Cheng and M.-H. Hsieh. “Moderate Deviation Analysis for Classical-Quantum Channels and Quantum Hypothesis Testing”. IEEE Transactions on Information Theory 64(2): 1385–1403 2018.
Appendix A: Proofs one-shot quantum state splitting

Here, we prove Theorem III.3. We start with the convex-split lemma and define an extended state as follows: For any finite non-empty set \( \Sigma \), define the quantum register \( A_\Sigma = \bigotimes_{s_i \in \Sigma} A_{s_i} \), where all \( A_{s_i} \) are isomorphic. An extended state is of the form \( \rho_{A_\Sigma} = \rho_{A_{s_1}} \otimes \rho_{A_{s_2}} \otimes \ldots \otimes \rho_{A_{|\Sigma|}} \in \mathcal{S}_\prec(A_\Sigma) \).
Lemma A.1 (Convex-split lemma [6]) For $I = \{1, 2, \ldots, n\}$ consider the quantum register $B_I$ and a quantum register $R$. Then, for $i \in I$, let $\rho_{B_iR} \in S(B_iR)$ and $\sigma_{B_i} \in S(B_i)$ and define the extended state $\sigma_{B_{I\setminus i}}$ and $\tau_{B_iR} = \frac{1}{n} \sum_{i=1}^{n} \rho_{B_iR} \otimes \sigma_{B_{I\setminus i}}$. For $\delta > 0$ and $n$ with $\log n \geq D_{\max}(\rho_{B_iR}||\sigma_{B_i} \otimes \rho_R) + \log \frac{1}{\delta}$, it holds that
\[
F(\tau_{B_iR}, \sigma_{B_i} \otimes \rho_R) \geq \sqrt{(1 - \delta)}.
\]

Next, we show the achievability bound for the cost of one-shot quantum state splitting. The following proposition closely follows the ideas in [6], with the sole difference that we employ the partially smoothed max-information instead of the smoothed max-information.

Proposition A.2 (One-shot quantum state splitting achievability) Consider a one-shot quantum state splitting protocol for $\rho_{AB} \in S(AB)$ with purification $\rho_{ABR} \in S(ABR)$. For $\varepsilon \in (0, 1]$ with $\delta \in (0, \varepsilon]$ it holds that
\[
q_\varepsilon^*(\rho_{AB}) \leq \frac{1}{2} I_{\max}^{\varepsilon-\delta}(\hat{R}; B)_\rho + \log \frac{2}{\delta}.
\]

Proof. Let $I_{\max}^{\varepsilon-\delta}(\hat{R}; B)_{\rho_{BR}} = D_{\max}(\rho_{BR}'||\sigma_{B} \otimes \rho_R')$, where $\rho_{BR}' \in B^{\varepsilon-\delta}(\rho_{BR})$, $\rho_R' = \rho_R$, and $\sigma_B \in S(B)$. For $i \in I = [n]$, let $L_i \cong B \cong A_1$ and let $|\sigma_{K_iL_i}\rangle$ be any purification of $\sigma_{L_i}$. Let the resource state be $|\sigma_{(K_L)i}\rangle \otimes |\phi_{K_i' L_i'}\rangle^\otimes_n$, where $|\phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is the maximally entangled qubit state. Let this resource state be shared between two parties, Alice who holds $K_i \in I$ and $K^m$, and Bob who holds $L_i \in I$ and $L^m$. Excluding the maximally entangled states, the joint state at the start of the protocol is
\[
|\psi\rangle := |\rho_{AA_iR}\rangle \otimes |\sigma_{(K_L)i}\rangle \approx_{\varepsilon-\delta} |\rho'_{AA_iR}\rangle \otimes |\sigma_{(K_L)i}\rangle := |\omega\rangle,
\]
where $|\rho'_{AA_iR}\rangle$ is the purification of $\rho'_{A_iR}$ on Alice’s quantum register $A$ that achieves the minimal purified distance from the purification $|\rho_{AA_iR}\rangle$ of $\rho_{AR}$. Uhlmann’s theorem guarantees that such a purification exists. The reduced state of $|\omega\rangle$ after tracing over Alice’s quantum registers $A, A_i$ and $K_i$ is $\rho_R' \otimes \sigma_{L_i}$. By the convex-split lemma (Lemma A.1), we have that for $\delta \in (0, 1]$ and $n \in \mathbb{N}$ satisfying
\[
\log n \geq D_{\max}(\rho_{BR}'||\sigma_{B} \otimes \rho_R') + \log \frac{1}{\delta^2},
\]
it holds that
\[
F\left(\rho_R' \otimes \sigma_{L_i}, \frac{1}{n} \sum_{i=1}^{n} \rho_{RL_i} \otimes \sigma_{L_i\setminus i}\right) \geq \sqrt{1 - \delta^2}.
\]
A purification of $\frac{1}{n} \sum_{i=1}^{n} \rho_{RL_i} \otimes \sigma_{L_i\setminus i}$ is
\[
|\tau\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |i\rangle_X \otimes |\rho_{ARL_i}\rangle \otimes |\sigma_{(K_L)i\setminus i}\rangle,
\]
where Alice holds the classical register $X$. We now construct a $(q, \varepsilon)$-one-shot state splitting protocol of $\rho_{AB}$ assuming (A4) holds.
1. \( \mathcal{E} \) operation: There exists an isometry \( U \) on Alice’s quantum registers such that \( F(U(|\omega\rangle), |\tau\rangle) \geq \frac{1}{\sqrt{1+\delta^2}} \) and hence \( P(U(|\omega\rangle), |\tau\rangle) \leq \delta \). Moreover, since \( \rho'_R = \rho_R \), there exists an isometry \( V \) on \( AA_1 \) such that \( V|\psi\rangle = |\omega\rangle \). Both these isometries exist due to Uhlmann’s theorem. Alice therefore performs the operation \( UV \) on \( |\psi\rangle \) and discards all the \( K'_{A1} \) quantum registers.

2. Alice and Bob use super-dense coding \([11]\) to send the classical register \( X \) to Bob. The state \( |\phi_{K'X}\rangle^\otimes n \) is used for super-dense coding and this has a quantum communication cost of \( \frac{1}{2} \log n \).

3. \( \mathcal{D} \) operation: Depending on the state \( |i\rangle \) of \( X \), Bob applies controlled swap operations \( L_i \leftrightarrow L_1 \) to obtain \( |\rho'_{ARL}i\rangle \) which we relabel as \( |\rho'_{ABR}\rangle \) and discards all \( L_i \) quantum registers where \( i \neq 1 \).

   Let us denote the protocol above by \( \mathcal{T}_\sigma \). Then, since \( UV(|\psi\rangle) \approx_\delta |\tau\rangle \) and \( \mathcal{D}(|\tau\rangle) \approx_{\varepsilon-\delta} \rho_{ABR} \), we can use the triangle inequality to conclude that \( \mathcal{T}_\sigma(\rho_{AA1}) \approx_{\varepsilon} \rho_{AB} \). This \( \varepsilon \)-error one-shot quantum state splitting protocol then has a communication cost of \( \frac{1}{2} \log n \) where \( n \in \mathbb{N} \). Since we have

\[
\exp \left( D_{\max}(\rho'_{BR}||\sigma_B \otimes \rho'_R) + \log \frac{4}{\delta^2} \right) \geq \left[ \exp \left( D_{\max}(\rho'_{BR}||\sigma_B \otimes \rho'_R) + \log \frac{1}{\delta^2} \right) \right], \tag{A7}
\]

we have that \( q^*_{\varepsilon}(\rho_{AB}) \leq D_{\max}(\rho'_{BR}||\sigma_B \otimes \rho'_R) + \log \frac{4}{\delta^2} \), which proves the claim. \( \blacksquare \)

Next, we show the converse for one-shot quantum state splitting.

**Proposition A.3 (One-shot quantum state splitting converse \([5, 13]\))** Consider \( \rho_{AB} \in S(AB) \) with purification \( \rho_{ABR} \) and let \( \varepsilon \in (0, 1] \). The minimal quantum communication cost of one-shot state splitting of \( \rho_{AB} \) satisfies

\[
q^*_{\varepsilon}(\rho_{AB}) \geq \frac{1}{2} I^*_{\max}(\hat{R}; B)_{\rho} . \tag{A8}
\]

**Proof.** We take \( A_1 \cong B \), initial resource state \( \sigma_{KL} \) and assume the communication cost is \( q = \log |Q| \). Let \( \omega_{AQLR} = (\mathcal{E} \otimes \mathcal{I}_{RL})(\rho_{\sigma_{AA1}R} \otimes \sigma_{KL}) \) and \( \tau_{ABR} = (\mathcal{D} \otimes \mathcal{I}_{AR})(\omega_{AQLR}) \approx_{\varepsilon} \rho_{ABR} \). Note that \( \tau_R = \rho_R \) and thus

\[
I^*_{\max}(\hat{R}; B)_{\rho} \leq I_{\max}(R; B)_\tau \leq I_{\max}(R; QL)_\omega \leq I_{\max}(R; L)_\omega + 2 \log |Q| \leq I_{\max}(R; L)_{\rho \otimes \sigma} + 2 \log |Q| , \tag{A9}
\]

where we have used the data processing inequality on the max-information and the non-lockability of the max-information (Lemma B.10). Since \( I_{\max}(R; L)_{\rho \otimes \sigma} = 0 \) as Bob is not correlated with the reference at the start of the protocol, the result follows. \( \blacksquare \)

**Appendix B: Technical lemmas and missing proofs**

1. **Distance measures**

**Lemma B.1** For \( \varepsilon, \varepsilon' \in [0, 1] \) with \( \varepsilon^2 + \varepsilon'^2 \leq 1 \), the function \( \varepsilon \sqrt{1 - \varepsilon'^2} + \varepsilon' \sqrt{1 - \varepsilon^2} \) is monotonically increasing in both \( \varepsilon \) and \( \varepsilon' \).

**Proof.** From the proof of tight triangle inequality for the purified distance (Lemma III.7), we have that \( \sin^{-1}(\varepsilon) + \sin^{-1}(\varepsilon') \leq \pi/2 \). Since the sine function is monotone increasing in \([0, \pi/2]\) and \( \varepsilon \sqrt{1 - \varepsilon^2} + \varepsilon' \sqrt{1 - \varepsilon'^2} = \sin(\sin^{-1}(\varepsilon) + \sin^{-1}(\varepsilon')) \), the result follows. \( \blacksquare \)
Lemma B.2 For any pair of quantum channels $\mathcal{E}_{A\rightarrow B}$ and $\mathcal{F}_{A\rightarrow B}$ and $\phi_{AR} \in \mathcal{S}(A \otimes R)$ pure, there exists $\psi_{AR'} \in \mathcal{S}(A \otimes R')$ pure such that

$$P((\mathcal{E} \otimes \mathcal{I}_R)(\phi), (\mathcal{F} \otimes \mathcal{I}_R)(\phi)) = P((\mathcal{E} \otimes \mathcal{I}_{R'})(\psi), (\mathcal{F} \otimes \mathcal{I}_{R'})(\psi)) \quad \text{with} \quad |R'| = |A|. \quad (B1)$$

Proof. If $|R| \leq |A|$, we may choose some isometry $U_{R\rightarrow R'}$ such that $\psi_{AR'} = (I_A \otimes U_{R\rightarrow R'})(\phi_{AR})(I_A \otimes U_{R\rightarrow R'})^\dagger$. Similarly, if we have that $|R| > |A|$, then we may choose $\psi_{AR'}$ to be the Schmidt decomposition of the state $\phi_{AR}$ with $|R'| = |A|$. Now there exists an isometry $V_{R'\rightarrow R}$ such that $(I_A \otimes V_{R'\rightarrow R})\psi_{AR'} = \phi_{AR}$. The purified distance is invariant under isometries acting on both arguments and the isometries here commute with the channels. This gives the desired result. ■

Lemma B.3 For $\rho, \sigma \in \mathcal{S}_\leq(A)$, let $\text{Tr}(\lambda \rho) = \text{Tr}(\lambda \sigma) = 1$ for some $\lambda \geq 1$. Then, we have

$$P(\rho, \sigma) \leq P(\lambda \rho, \lambda \sigma) \leq \sqrt{2\lambda P(\rho, \sigma)}. \quad (B2)$$

Proof. We use the generalized fidelity for sub-normalized states. We have

$$F(\lambda \rho, \lambda \sigma) = \lambda \|\rho \sqrt{\sigma}\|_1 = \lambda \left(\bar{F}(\rho, \sigma) - \left(1 - \frac{1}{\lambda}\right)\right) = 1 - \lambda(1 - \bar{F}(\rho, \sigma)). \quad (B3)$$

Rewriting, we have

$$\bar{F}(\rho, \sigma) = \frac{1}{\lambda} F(\lambda \rho, \lambda \sigma) + 1 - \frac{1}{\lambda}. \quad (B4)$$

The right hand side is a convex combination of $F(\lambda \rho, \lambda \sigma)$ and 1 and we can hence lower bound it by $F(\lambda \rho, \lambda \sigma)$. Switching to purified distance, we have the first inequality. We also have

$$P^2(\lambda \rho, \lambda \sigma) = 1 - (1 - \lambda(1 - \bar{F}(\rho, \sigma)))^2 = 2\lambda(1 - \bar{F}(\rho, \sigma)) - \lambda^2(1 - \bar{F}(\rho, \sigma))^2 \quad (B5)$$

$$\leq 2\lambda(1 - \bar{F}(\rho, \sigma)) \quad (B6)$$

$$\leq 2\lambda(1 - \bar{F}^2(\rho, \sigma)) \quad (B7)$$

$$= 2\lambda P^2(\rho, \sigma). \quad (B8)$$

Lemma B.4 For classical-quantum states $\rho = \frac{1}{d} \sum_{i=1}^d |i\rangle \langle i| \otimes \rho_i$ and $\sigma = \frac{1}{d} \sum_{i=1}^d |i\rangle \langle i| \otimes \sigma_i$, where $\{|i\rangle\}$ form an orthonormal basis, it holds that

$$F(\rho, \sigma) = \frac{1}{d} \sum_{i=1}^d F(\rho_i, \sigma_i). \quad (B9)$$

Proof. This follows from the fact that for classical-quantum states $\tau = \sum_i p(i) |i\rangle \langle i| \otimes \tau_i$ and $\omega = \sum_i q(i) |i\rangle \langle i| \otimes \omega_i$ where $p(i), q(i)$ are probability vectors and $\tau_i$ and $\omega_i$ are quantum states, one has $F(\tau, \omega) = \sum_i \sqrt{p(i)q(i)} F(\tau_i, \omega_i)$. ■

Lemma B.5 (Quasi-convexity of purified distance [62, Eq. (3.54)]) For $\lambda \in [0, 1]$ $i \in \{1, 2\}$ and $\rho_i, \tau_i \in \mathcal{S}_\leq(A)$, it holds that

$$P(\lambda \rho_1 + (1 - \lambda)\rho_2, \lambda \tau_1 + (1 - \lambda)\tau_2) \leq \max_i P(\rho_i, \tau_i). \quad (B10)$$
2. Entropic quantities

Lemma B.6 (Mutual information variance [26, Corollary III.5]) For $\rho_{AB} \in S(AB)$, we have that

$$V(\rho_{AB}\|\rho_A \otimes \rho_B) \leq 4 \log^2(2d_A+1). \quad (B11)$$

Lemma B.7 ([22, Proposition 4.7]) Let $0 < \varepsilon < 1, \delta, \eta > 0$, and $\rho, \sigma \in \mathcal{P}(A)$ with $\text{Tr}(\rho) \leq 1$. Then, we have

$$D^{\varepsilon-\delta}(\rho\|\sigma) + \log \delta \leq D^{\varepsilon}_{h}(\rho\|\sigma). \quad (B12)$$

Lemma B.8 ([4, Theorem 4]) Let $\rho \in S(A), \sigma \in \mathcal{P}(A), \varepsilon \in (0,1)$, and $\delta \in (0,1-\varepsilon^2)$. Then, we have that

$$D^{1-\varepsilon}_{h}(\rho\|\sigma) \geq D^{\max}_{\sigma,P}(\rho\|\sigma) - \log \frac{1}{1-\varepsilon} \geq D^{1-\varepsilon-\delta}_{h}(\rho\|\sigma) - \log \frac{4}{\delta^2}. \quad (B13)$$

Lemma B.9 (Duality of Rényi relative entropies [35, Lemma 6]) Let $\rho_{AB} \in S(AB)$ and $\tau_A \geq 0$ such that $\tau_A \gg \rho_A$. Then, for any purification $\rho_{ABC}$ of $\rho_{AB}$, we have

$$\bar{I}_\alpha(\rho_{AB}\|\tau_A) = -\bar{I}_\beta(\rho_{AC}\|\tau_A^{-1}) \quad \text{for} \quad \alpha, \beta \in [1/2, \infty), \quad \alpha^{-1} + \beta^{-1} = 2 \quad (B14)$$

$$D_\alpha(\rho_{AB}\|\tau_A \otimes \rho_B) = -D_\beta(\rho_{AC}\|\tau_A^{-1} \otimes \rho_C) \quad \text{for} \quad \alpha, \beta \in [0,2], \quad \alpha + \beta = 2 \quad (B15)$$

$$V(\rho_{AC}\|\tau_A^{-1} \otimes \rho_C) = V(\rho_{AB}\|\tau_A \otimes \rho_B), \quad (B16)$$

where the inverse is taken on the support of $\tau_A$.

Lemma B.10 (Non-lockability of max-information [12, Lemma A.12]) Let $\rho_{ABC} \in S(ABC)$ and $\varepsilon \geq 0$. Then, we have that

$$I^{\varepsilon}_{\max}(A;BC)_\rho \leq I^{\varepsilon}_{\max}(A;B)_\rho + 2 \log |C|. \quad (B17)$$

Lemma B.11 (Partially smoothed max-information bound [5, Theorem 2]) Let $\rho_{AB} \in S(AB)$ and $0 \leq 2\varepsilon + \delta \leq 1$ with $\delta > 0$. Then, we have

$$I^{2\varepsilon+\delta}_{\max}(A;B)_\rho \leq I^{\varepsilon,P}_{\max}(A;B)_\rho + \log \frac{8+\delta^2}{\delta^2}. \quad (B18)$$

3. De Finetti and more

Definition B.12 (De Finetti state) For $\sigma \in S(A)$ and $\mu(.)$ a probability measure on $S(A)$, $\zeta_{H^n} = \int \sigma^\otimes n \mu(\sigma) \in S(A^n)$ is called a de Finetti state.

Lemma B.13 (Post-selection technique [17]) For a $\rho_{A^nR^n} \in S(A^nR^n)$ with support on the symmetric subspace $\text{Sym}^n (A \otimes R)$ with $R \cong A$, there exists a completely positive trace-non-increasing map $\mathcal{G}_{R \rightarrow C}$ such that

$$\rho_{A^nR^n} = g_{n,|A|}(I_{A^nR^n} \otimes \mathcal{G})(\zeta_{A^nR^n}), \quad (B19)$$

where $g_{n,d} = \left( \frac{n + d^2 - 1}{n} \right) \leq (n+1)d^2-1$, $\zeta_{AR} = \int \sigma\otimes_{AR} d (\sigma_{AR})$ with $\sigma_{AR} = |\sigma\rangle\langle\sigma|_{AR} \in S(AR)$, $A \cong R$ and $d(.)$ is the measure on the normalized pure states on $AR$ induced by the Haar measure on the unitary group acting on $AR$ normalized to $\int d(.) = 1$. Hence, $\zeta_{A^nR^n} \in S(A^nR^n)$ is the purification of $\zeta_{A^nR^n}$ and we can assume without loss of generality that $|R'| \leq (n+1)|A|^2-1$. 

The following proposition is as [17, Theorem 1], but we changed the distance measure from the diamond distance to the channel purified distance.

**Proposition B.14 (De Finetti reduction)** Let $\mathcal{E}_{A^n \rightarrow B^n}$, $\mathcal{F}_{A^n \rightarrow B^n}$ be quantum channels and $\varepsilon > 0$. If $\mathcal{E}$ and $\mathcal{F}$ are permutation covariant, i.e., for any permutation $\pi$ on $n$ quantum registers, one has $\mathcal{E} \circ \pi_{A^n} = \pi_{B^n} \circ \mathcal{E}$ and $\mathcal{F} \circ \pi_{A^n} = \pi_{B^n} \circ \mathcal{F}$, where $\pi_{A^n}$ ($\pi_{B^n}$) denotes the channel that applies the permutation $\pi$ to the $A^n$ ($B^n$) registers. Then

$$P(\mathcal{E}, \mathcal{F}) \leq \sqrt{2(n+1)} \frac{|A|^2 - 1}{2} P((\mathcal{E} \otimes I_{R^nR'})((\mathcal{F} \otimes I_{R^nR'})(\zeta_{A^nR^nR'}), (\mathcal{F} \otimes I_{R^nR'})(\zeta_{A^nR^nR'})), \tag{B20}$$

where $\zeta_{A^nR^nR'}$ is the purification of the de Finetti state defined in Lemma B.13.

**Proof.** We wish to show that

$$P((\mathcal{E} \otimes I_{R^n})(\rho_{A^nR^n}), (\mathcal{F} \otimes I_{R^n})(\rho_{A^nR^n})) \leq \sqrt{2g_{n,|A|}} P((\mathcal{E} \otimes I_{R^nR'})(\zeta_{A^nR^nR'}), (\mathcal{F} \otimes I_{R^nR'})(\zeta_{A^nR^nR'})), \tag{B22}$$

where $R \cong A$. First we may assume that $\rho_{A^nR^n}$ has support on $\text{Sym}^n(A \otimes R)$. To see this, define the quantum state

$$\tilde{\rho}_{A^nR^nR'} = \frac{1}{n!} \sum_\pi (\pi_{A^n} \otimes I_{R^n})(\rho_{A^nR^n}) \otimes |\pi\rangle \langle |\pi|', \tag{B23}$$

where $\pi$ is a permutation operation on the $n$ quantum registers, the summation is over all permutations on $A^n$ and $\{|\pi|\}$ is a basis of $R'$. We now have

$$F((\mathcal{E} \otimes I_{R^n})(\rho_{A^nR^n}), (\mathcal{F} \otimes I_{R^n})(\rho_{A^nR^n})) \leq \frac{1}{n!} \sum_\pi F((\pi_{B^n} \circ \mathcal{E} \otimes I_{R^n})(\rho_{A^nR^n}), (\pi_{B^n} \circ \mathcal{F} \otimes I_{R^n})(\rho_{A^nR^n})) \tag{B25}$$

$$= \frac{1}{n!} \sum_\pi F((\mathcal{E} \otimes I_{R^n})(\pi_{A^n} \otimes I_{R^n})(\rho_{A^nR^n}), (\mathcal{F} \otimes I_{R^n})(\pi_{A^n} \otimes I_{R^n})(\rho_{A^nR^n})) \tag{B26}$$

$$= F((\mathcal{E} \otimes I_{R^nR'})(\tilde{\rho}_{A^nR^nR'}), (\mathcal{F} \otimes I_{R^nR'})(\tilde{\rho}_{A^nR^nR'})) \tag{B27}$$

where the first equality holds due to the invariance of the fidelity under isometries, the second equality holds due to the permutation covariance of $\mathcal{E}$ and $\mathcal{F}$, the third equality is due to Lemma B.4 on the fidelity of classical-quantum states and the linearity of the channels $\mathcal{E}$ and $\mathcal{F}$.

The reduced state $\tilde{\rho}_{A^n}$ is permutation invariant and therefore has a permutation invariant purification $\tilde{\rho}_{A^nA^n'}$ with $A' \cong A$ [53, Lemma 4.2.2]. Since all purifications are equivalent up to isometries on the purifying quantum register, there exists a quantum channel $\mathcal{M}_{A^n \rightarrow R^nR'}$ such that $\tilde{\rho}_{A^nR^nR'} = (I_{A^n} \otimes \mathcal{M})(\tilde{\rho}_{A^nA^n'})$. Switching to the purified distance, we have

$$P((\mathcal{E} \otimes I_{R^n})(\rho_{A^nR^n}), (\mathcal{F} \otimes I_{R^n})(\rho_{A^nR^n})) \leq P((\mathcal{E} \otimes I_{A^n})(\tilde{\rho}_{A^nA^n'}), (\mathcal{F} \otimes I_{A^n})(\tilde{\rho}_{A^nA^n'})) \tag{B28}$$

$$\leq P((\mathcal{E} \otimes I_{A^n})(\tilde{\rho}_{A^nA^n'}), (\mathcal{F} \otimes I_{A^n})(\tilde{\rho}_{A^nA^n'})) \tag{B29}$$

where the inequality holds because the purified distance is non-increasing under quantum channels. Now, we have

$$P((\mathcal{E} \otimes I_{A^n})(\tilde{\rho}_{A^nA^n'}), (\mathcal{F} \otimes I_{A^n})(\tilde{\rho}_{A^nA^n'})) \tag{B31}$$
\[
\begin{align*}
&= P \left( (E^n \otimes I_{A^n}) \circ g_{n,A}(I_{A^n A^n} \otimes \mathcal{G})(\zeta_{A^n A^n R^n}), (F^n \otimes I_{A^n}) \circ g_{n,A}(I_{A^n A^n} \otimes \mathcal{G})(\zeta_{A^n A^n R^n}) \right) \\
&\leq \sqrt{2g_{n,A}} P \left( (E^n \otimes I_{A^n}) \circ (I_{A^n A^n} \otimes \mathcal{G})(\zeta_{A^n A^n R^n}), (F^n \otimes I_{A^n}) \circ (I_{A^n A^n} \otimes \mathcal{G})(\zeta_{A^n A^n R^n}) \right) \\
&\leq \sqrt{2g_{n,A}} P \left( (E^n \otimes I_{A^n}) (\zeta_{A^n A^n R^n}), (F^n \otimes I_{A^n}) (\zeta_{A^n A^n R^n}) \right),
\end{align*}
\]

where the first equality follows from Lemma B.13, the first inequality is due to Lemma B.3 and the second inequality is because the purified distance cannot increase under completely positive trace non-increasing maps.

**Proof of Lemma V.4:** We construct \( \tilde{T}^\omega \) from \( T^\sigma \) as follows. The resource state is \( \omega = \sigma_{KL} \otimes \sigma'_{K'L'} \), where
\[
\sigma'_{K'L'} = \frac{1}{n!} \sum_{\pi} |\pi\pi\rangle \langle \pi\pi|_{K'L'}.
\]
The encoders and decoders are modified from \( E \) and \( D \) respectively to \( E' = \frac{1}{n!} \sum_{\pi} E \circ (\pi_{A^n} \otimes \pi_{B^n}) \otimes \langle \pi| \cdot |\pi\rangle_{K'} \) and \( D' = \frac{1}{n!} \sum_{\pi} (\pi_{A^n}^{-1} \otimes \pi_{B^n}^{-1}) \circ D \otimes \langle \pi| \cdot |\pi\rangle_{L'} \). The following properties are clear:

1. \( \tilde{T}^\omega = \frac{1}{n!} \sum_{\pi} (\pi_{A^n}^{-1} \otimes \pi_{B^n}^{-1}) \circ T^\sigma \circ (\pi_{A^n} \otimes \pi_{A^n}') \)

2. \( \tilde{T}^\omega = D' \circ E' \circ \mathcal{T}^\omega \) satisfies \( \tilde{T}^\omega \circ (\tilde{\pi}_{A^n} \otimes \tilde{\pi}_{A^n}') = (\tilde{\pi}_{A^n} \otimes \tilde{\pi}_{B^n}) \circ \tilde{T}^\omega \)

3. The communication cost \( q \) associated with \( T^\sigma \) is the same as that of \( \tilde{T}^\omega \).

Since \( \rho_{A^n B^n} \) is permutation invariant, it has a permutation invariant purification \( \rho_{A^n B^n R^n} \) i.e.,
\[
(\pi_{A^n} \otimes \pi_{B^n} \otimes \pi_{R^n}) \rho_{A^n B^n R^n} = \rho_{A^n B^n R^n} \quad \text{for any permutation } \pi \quad [53, \text{Lemma 4.2.2}].
\]
We can, without loss of generality, consider this purification due to Remark III.2. We have that
\[
P \left( \rho_{A^n B^n R^n}, (\tilde{T}^\omega \otimes I_{R^n})(\rho_{A^n A^n R^n}) \right) \\
= P \left( \rho_{A^n B^n R^n}, \left( \frac{1}{n!} \sum_{\pi} (\pi_{A^n}^{-1} \otimes \pi_{B^n}^{-1}) \circ T^\sigma \circ (\pi_{A^n} \otimes \pi_{A^n}') \otimes I_{R^n} \right)(\rho_{A^n A^n R^n}) \right) \\
= P \left( \rho_{A^n B^n R^n}, \left( \frac{1}{n!} \sum_{\pi} (\pi_{A^n}^{-1} \otimes \pi_{B^n}^{-1}) \circ T^\sigma \circ \pi_{R^n}^{-1} \right)(\rho_{A^n A^n R^n}) \right) \\
= P \left( \rho_{A^n B^n R^n}, \left( \frac{1}{n!} \sum_{\pi} (\pi_{A^n}^{-1} \otimes \pi_{B^n}^{-1} \otimes \pi_{R^n}^{-1}) \circ T^\sigma \circ I_{R^n} \right)(\rho_{A^n A^n R^n}) \right) \\
\leq \max_{\pi} P \left( \rho_{A^n B^n R^n}, \left( (\pi_{A^n}^{-1} \otimes \pi_{B^n}^{-1} \otimes \pi_{R^n}^{-1}) \circ T^\sigma \circ I_{R^n} \right)(\rho_{A^n A^n R^n}) \right) \\
= \max_{\pi} P \left( (\pi_{A^n} \otimes \pi_{B^n} \otimes \pi_{R^n})(\rho_{A^n B^n R^n}), (T^\sigma \otimes I_{R^n})(\rho_{A^n A^n R^n}) \right) \\
= P \left( \rho_{A^n B^n R^n}, (T^\sigma \otimes I_{R^n})(\rho_{A^n A^n R^n}) \right),
\]
where the second equality uses the fact that the \( \rho \) is permutation invariant, the third equality is because the state splitting protocol commutes with operations on the \( R^n \) quantum registers, the first inequality uses the quasi-convexity of the purified distance (Lemma B.5), the fourth equality uses the invariance of the purified distance under isometries and the last equality uses that \( \rho \) is permutation invariant.
4. Miscellaneous

Lemma B.15 ([50, Lemma 48]) Let $D$ be a compact metric space and $\{a_n\}$ be a moderate sequence. For any continuous functions $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ we have

$$\max_{x \in D} \left\{ f(x) + a_n g(x) \right\} = f^* + a_n g^* + o(a_n)$$

(B42)

for $f^* = \max_{x \in D} f(x)$ and $g^* = \sup_{x : f(x) = f^*} g(x)$.

Lemma B.16 (Strong converse for quantum identity channel) Consider a quantum identity channel $\mathcal{I}_{A \rightarrow B}$ where $A \cong B$. The average success probability of correctly transmitting $2^R$ classical messages over $\mathcal{I}$ with entanglement-assistance is bounded as

$$P_{\text{succ}}(\mathcal{I}_{A \rightarrow B}, R) \leq 2^{-(R - 2 \log |A|)}.$$  

(B43)

Proof. Let Alice send a uniformly random message $m$ from the set $M$ where $|M| = 2^R$ with average success probability $P_{\text{succ}}$ using the channel $\mathcal{I}_{A \rightarrow B}$. To do so, Alice and Bob share an entangled resource $\sigma_{KL}$. For each message $m \in M$, the encoder can be chosen without loss of generality as $\mathcal{E}^m_{K \rightarrow A}$ taken from the set $\{\mathcal{E}^m_{K \rightarrow A}\}$. Let $\rho_A = \frac{1}{|M|} \sum_{m \in M} \mathcal{E}^m_{K \rightarrow A}(\sigma_K)$. By the meta-converse argument of [44] (see also [29, Proposition 1]), there exists a two-element POVM $\{T_{RB}, I - T_{RB}\}$ such that for any purification $\psi_{RA}$ of $\rho_A$ and any state $\omega_A$, we have

$$P_{\text{succ}} = \text{Tr} (T_{RB} \mathcal{I}_{A \rightarrow B}(\psi_{RA}))$$

(B44)

$$\frac{1}{|M|} = \text{Tr} (T_{RB} \mathcal{I}_{A \rightarrow B}(\psi_R \otimes \omega_A))$$

(B45)

Finally, define the state $\rho_p = p|0\rangle\langle 0| + (1 - p)|1\rangle\langle 1|$ for any $p \in [0, 1]$. We now have the following chain of inequalities

$$\log \frac{P_{\text{succ}}}{|M|} \leq D_{\max} \left( \rho_{P_{\text{succ}}} \| \rho_{\frac{1}{|M|}} \right)$$

(B46)

$$\leq D_{\max} \left( \psi_{RA} \| \psi_R \otimes \frac{I_A}{|A|} \right)$$

(B47)

$$\leq 2 \log |A|,$$

(B48)

where the first inequality is due to the definition of the max-relative entropy, the second inequality is due to the monotonicity of the max-relative entropy under the test $\{T_{RB}, I - T_{RB}\}$ and the choice of $\omega_A = \frac{I_A}{|A|}$, and the final inequality is due to [12, Lemma B.6].

Proof of Proposition V.5: Let $\zeta_{A^n}$ be the de Finetti state as given in Lemma B.13. Let $U_{A^n}^{\otimes n}$ be the Stinespring dilation of the $n$-fold tensor product of the channel $\mathcal{N}_{A \rightarrow A_1}$ and denote $U_{A^n}^{\otimes n} = \Phi_{A^n}^{E^n}$. Since $\zeta_{A^n}$ is permutation invariant, we have that $\Phi_{A^n}^{E^n}$ is also permutation invariant. For any $\epsilon' \in (0, 1]$ and $\delta \in (0, \epsilon')$, let $T_n$ be $\{q_n, \epsilon'\}$-one-shot state splitting protocol of $\Phi_{A^n}^{E^n}$. By Lemma V.4, we can assume that $T_n$ is permutation covariant. Let $T'_n = \text{Tr}_{E^n} \circ T_n \circ U_{E^n}^{\otimes n}$. We make the following observations about $T'_n$:

1. $T'_n$ is permutation covariant.

2. For any purification $\zeta_{A^n}$ of $\zeta_{A^n}$, we have $P \left( (\mathcal{N}_{A \rightarrow B}^{\otimes n} \otimes I_R) \left( \zeta_{A^n R} \right), (T'_n \otimes I_R) \left( \zeta_{A^n R} \right) \right) \leq \epsilon'$.
By choosing $\epsilon' \leq \frac{\epsilon}{\sqrt{2(n+1)^2}}$, we obtain that $P\left(\mathcal{N}_{A\rightarrow B}^{\otimes n}, T_n'\right) \leq \epsilon$ through the de Finetti reduction (Proposition B.14). $T_n'$ is therefore a $\{q_n, \epsilon\}$-one-shot simulation protocol of $\mathcal{N}_{A\rightarrow B}^{\otimes n}$.

It now remains to pick a purification of $\zeta_{A^n}$ to calculate $q_n$. By [13, Corollary D.6], we may choose the extension $\zeta_{A^nR^n} = \sum_{i \in I} p_i (\omega_{AR}^i)^{\otimes n}$ where $I = \{1, 2, \ldots, (n + 1)^{|A| - 2}\}$ and $\omega_{AR}^i$ are pure states. Since this state is permutation invariant, we may choose a permutation invariant purification $\zeta_{A^nR^nR'}$ with $|R'| \leq (n + 1)^{|A| - 1}$. By the one-shot characterization of quantum state splitting (Theorem III.3) and by choosing $\delta = \frac{\epsilon'}{2}$, we have

$$q^*_e(\mathcal{N}_{A\rightarrow B}^{\otimes n}) \leq q^*_e(\Phi_{B^nE^n})$$

$$\leq \frac{1}{2} I_{\max}^\prime (R^n; B^n) + \log \frac{4}{\epsilon'}$$

$$\leq \frac{1}{2} I_{\max}^\prime (R^n; B^n) + \frac{1}{2} \log \frac{8 + (\frac{\epsilon'}{2})^2}{\epsilon'^2} + \log \frac{4}{\epsilon'}$$

$$\leq \frac{1}{2} I_{\max}^\prime (B^n; R^n) + \frac{1}{2} \log \left(\frac{2}{\epsilon'^2/24} + 2\right) + \frac{1}{2} \log \frac{8 + (\frac{\epsilon'}{2})^2}{\epsilon'^{22}} + \log \frac{4}{\epsilon'},$$

where the second inequality holds due to the cost of one-shot state splitting (Theorem III.3), the third inequality follows due to the bound on the partially smoothed max-information using the smoothed max-information (Lemma B.11) and the fourth inequality follows due to the approximate symmetry of the smooth max-information [12, Lemma 2.11]). We now have

$$\frac{1}{2} I_{\max}^\prime (B^n; R^n) \leq \frac{1}{2} I_{\max}^\prime (B^n; R^n) + \log |R'|$$

$$\leq \max_{i \in I} \frac{1}{2} I_{\max}^\prime (B^n; R^n) ((\mathcal{N}\otimes I) (\omega^i))^{\otimes n} + \frac{1}{2} \log |I| + \log |R'|$$

$$\leq \max_{\tau_{AR}} \frac{1}{2} I_{\max}^\prime (B^n; R^n) ((\mathcal{N}\otimes I) (\tau))^{\otimes n} + 2(|A|^2 - 1) \log (n + 1),$$

where the first inequality holds due to the bound on the increase of the max-information due to an additional quantum register (Lemma B.9 of [13]), the second inequality follows by the quasi-convexity of the smooth max-entropy (Lemma B.18 of [13]) and the third inequality holds due to the bound on the dimension of the purification of the de Finetti state in the de Finetti reduction (Proposition B.14) and the bound on $|I|$. We use again the approximate symmetry of the smoothed max-information and bound the smoothed-max information by the partially smoothed max-information to obtain

$$\max_{\tau_{AR}} \frac{1}{2} I_{\max}^\prime (B^n; R^n) ((\mathcal{N}\otimes I) (\tau))^{\otimes n} \leq \max_{\tau_{AR}} \frac{1}{2} I_{\max}^\prime (B^n; R^n) ((\mathcal{N}\otimes I) (\tau))^{\otimes n} + \frac{1}{2} \log \left(\frac{2}{\epsilon'^2/24} + 2\right).$$

#### Appendix C: Meta-converse channel coding

Here, we compare the results from Section VI with results obtained by a meta-converse argument for a certain class of quantum channels. Our meta-converse below adapts the argument in [44] to the channel purified distance.
Proposition C.1 (Meta-converse channel coding) Let $\mathcal{N}_{A'\rightarrow B'}$ be a quantum channel, registers $A, B$ and $R$ all be isomorphic, and $\varepsilon \in (0, 1]$. Suppose there exists some $\tilde{\mathcal{N}}_{A\rightarrow B}$ per Definition VI.1 such that $P(\tilde{\mathcal{N}}, I_R) \leq \varepsilon$. Then, we have that

$$r^*_\varepsilon(\mathcal{N}) \leq \min_{\rho_{A'R}} \frac{1}{2} D_h^2 \left( (\mathcal{N} \otimes I_R) (\rho_{A'R}) \parallel \sigma_{B'R} \otimes \rho_R \right).$$

(C1)

Proof. We are given that $P(\tilde{\mathcal{N}}, I_R) \leq \varepsilon$. Let $|\phi\rangle = \frac{1}{\sqrt{|R|}} \sum_{i=1}^{|R|} |i\rangle |i\rangle$ be the $|R|$-dimensional maximally entangled state on the appropriate quantum registers. It holds that

$$P \left( \left( \tilde{\mathcal{N}} \otimes I_R \right) (|\phi\rangle\langle\phi|_{AR}), |\phi\rangle\langle\phi|_{BR} \right) \leq \varepsilon$$

(C2)

$$\Rightarrow F \left( \left( \tilde{\mathcal{N}} \otimes I_R \right) (|\phi\rangle\langle\phi|_{AR}), |\phi\rangle\langle\phi|_{BR} \right) \geq \sqrt{1 - \varepsilon^2}$$

(C3)

$$\Rightarrow \text{Tr} \left( |\phi\rangle\langle\phi|_{BR} \left( \tilde{\mathcal{N}} \otimes I_R \right) (|\phi\rangle\langle\phi|_{AR}) \right) \geq 1 - \varepsilon^2.$$  

(C4)

We also have for $\sigma_B \in S(B)$ that

$$\text{Tr} \left( |\phi\rangle\langle\phi|_{BR} \left( \sigma_B \otimes \frac{I_R}{|R|^2} \right) \right) = \frac{1}{|R|^2}.$$  

(C5)

Let $\mathcal{M}_{A'\rightarrow B'}$ be a channel that outputs a fixed state $\sigma_{B'}$ for any input and $\tilde{\mathcal{M}}_{A\rightarrow B}$ be defined with the encoding operation, decoding operation and resource state chosen to be the same as those of $\tilde{\mathcal{N}}$. It then holds that

$$D_h^2 \left( \left( \tilde{\mathcal{N}} \otimes I_R \right) (|\phi\rangle\langle\phi|), \left( \tilde{\mathcal{M}} \otimes I_R \right) (|\phi\rangle\langle\phi|) \right) = \max_{0 \leq T \leq \varepsilon^2} - \log \text{Tr} \left( T \left( \tilde{\mathcal{M}} \otimes I_R \right) (|\phi\rangle\langle\phi|) \right)$$

(C6)

$$\geq - \log \text{Tr} \left( |\phi\rangle\langle\phi| \left( \tilde{\mathcal{M}} \otimes I_R \right) (|\phi\rangle\langle\phi|) \right)$$

(C7)

$$= 2 \log |R|.$$  

(C8)

Let the set of all channels $\mathcal{M}_{A'\rightarrow B'}$ with constant output be denoted by $\mathcal{M}_{\text{const}}$. Then we have

$$2 \log |R| \leq \min_{\mathcal{M} \in \mathcal{M}_{\text{const}}} D_h^2 \left( \left( \tilde{\mathcal{N}} \otimes I_R \right) (|\phi\rangle\langle\phi|), \left( \tilde{\mathcal{M}} \otimes I_R \right) (|\phi\rangle\langle\phi|) \right)$$

(C9)

$$\leq \max_{\omega_{AR}} \min_{\mathcal{M} \in \mathcal{M}_{\text{const}}} D_h^2 \left( \left( \tilde{\mathcal{N}} \otimes I_R \right) (\omega_{AR}), \left( \tilde{\mathcal{M}} \otimes I_R \right) (\omega_{AR}) \right)$$

(C10)

$$\leq \max_{\rho_{A'R}} \min_{\mathcal{M} \in \mathcal{M}_{\text{const}}} D_h^2 \left( \left( \mathcal{N} \otimes I_R \right) (\rho_{A'R}), \left( \mathcal{M} \otimes I_R \right) (\rho_{A'R}) \right)$$

(C11)

$$\leq \max_{\rho_{A'R}} \min_{\sigma_{B'R}} D_h^2 \left( \left( \mathcal{N} \otimes I_R \right) (\rho_{A'R}) \parallel \sigma_{B'R} \otimes \rho_R \right),$$

(C12)

where the third inequality follows by data processing inequalities due to removing the encoding and decoding operations and the last inequality holds since $\mathcal{M} \in \mathcal{M}_{\text{const}}$. \hfill \blacksquare

For quantum channels that are covariant with respect to unitaries on the input quantum register (see [39] or [23] for a more precise definition), the maximization over input states is achieved by the $n$-fold tensor product of maximally entangled states [23, Eq (4.81)]. For such covariant channels, we have the following

$$r^*_\varepsilon(A^\otimes n) \leq \min_{\sigma_{B^n}} \frac{1}{2} D_h^2 \left( (A \otimes I_R) (|\phi\rangle\langle\phi|_{A'R})^{\otimes n}, \sigma_{B^n} \otimes \left( \frac{I_R}{d} \right)^{\otimes n} \right).$$

(C13)
\[
\leq \frac{1}{2} D_h^2 \left( \left( (\mathcal{N} \otimes I_R) (|\phi\rangle \langle \phi|_{A^{'}}) \right) \otimes_n \left( \mathcal{N}(\rho_{A'}) \right) \otimes_n \left( \frac{I_R}{d} \right) \right). \tag{C14}
\]

For the particular case where the error is chosen to be \(1 - \varepsilon_n = 1 - e^{-na_n^2}\) and using the AEP for the hypothesis testing relative entropy (Proposition III.10), we get that for \(\eta > 0\), there exists \(n^* \in \mathbb{N}\) such that for \(n \geq n^*\) we have

\[
\frac{1}{n} r_{1-\varepsilon}(\mathcal{N}^{\otimes n}) \leq \frac{1}{2n} D_h \left( 1 - 2\varepsilon_n + \varepsilon_n^2 \right) \left( \left( (\mathcal{N} \otimes I_R) (|\phi\rangle \langle \phi|_{A^{'}}) \right) \otimes_n \left( \mathcal{N}(\rho_{A'}) \right) \otimes_n \left( \frac{I_R}{d} \right) \right) \tag{C15}
\]
\[
\leq \frac{1}{2} \left( I(\mathcal{N} \otimes I_R)(|\phi\rangle \langle \phi|) + \sqrt{2V(\mathcal{N} \otimes I_R)(|\phi\rangle \langle \phi|)} a_n + \eta a_n \right) \tag{C16}
\]
\[
\leq C(\mathcal{N}) + \frac{a_n}{\sqrt{2}} \sqrt{V_{\max}(\mathcal{N} \otimes I_R)} + o(a_n), \tag{C17}
\]

which matches the converse statement for channel coding in the moderate deviation regime (Proposition VI.4).