Coherent Distributed Source Simulation as Multipartite Quantum State Splitting

Ian George[1] and Hao-Chung Cheng[2-6]

[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

[2]Department of Electrical Engineering National Taiwan University, Taipei 10617, Taiwan (R.O.C.)

[3]Department of Mathematics, National Taiwan University, Taipei 10617, Taiwan (R.O.C.)

[4]Center for Quantum Science and Engineering, National Taiwan University

[5]Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan (R.O.C.)

[6]Hon Hai Quantum Computing Research Center, Taipei, Taiwan

***Abstract*—Recently, Cheng et al. generalized quantum state splitting (QSS) to the multipartite setting for applications in quantum network information theory. Furthermore, Cheng and Gao recently noted that unipartite convex splitting reduces to soft covering for classical-quantum states. Soft covering is a common primitive in (classical, network) information theory that does not seem to need an extension to the multipartite setting, e.g. when establishing rates for many party distributed source simulation (DSS). This suggests a gap in our understanding of how these tasks are related. In this work we clarify this relation. We first identify "incoherent" QSS without entanglement assistance with DSS in the bipartite setting. Motivated by this, we identify "coherent" DSS as a special case of multipartite QSS. We then establish a one-shot rate region for multipartite QSS in terms of smooth max multipartite mutual information quantities, which in turn implies a one-shot rate region for coherent DSS. Lastly, we establish a strong asymptotic equipartition property for mutual information quantities to establish a strong converse rate region for multipartite QSS and coherent DSS.**

I. Introduction & Summary

The soft-covering lemma [1] and quantum state splitting [2] are two fundamental techniques for establishing achievable rates for (quantum) information processing tasks in terms of mutual information-induced quantities. In particular, soft-covering has been used to establish achievability rates for distributed source simulation (DSS), channel simulation, channel resolvability (or, relatedly, identification) [3]–[12] and quantum state splitting (QSS) has been used to establish rates for the transfer of quantum messages and channel simulation [2], [13]–[17]. These works have relied on increasingly optimal characterizations of the soft-covering [7], [10], [18], [19] and QSS [2], [20], [21]. These results have culminated in a multipartite version of QSS, which was used to establish a rate for quantum broadcast channel simulation [16] and, along with a related result for multipartite decoupling [22], was a major advancement in tools for establishing quantum multi-user rate regions.

Along with these refinements of soft-covering and QSS, Cheng and Gao also observed that when one considers a classical-quantum state, the error measure for convex splitting, which is used to prove QSS rates, reduces to the measure of error for soft-covering [16]. This formal insight calls into question if there is an operational (rather than formal) correspondence between QSS and some application of soft-covering. In this proceeding, we answer this question, which consists of two parts. In the first part, we establish the rate of distributed source simulation (DSS) as determining the optimal strategy for QSS when correlations with a reference system do not need to be preserved and there is no entanglement assistance. This establishes a correspondence between DSS and a modified version of QSS. This also motivates the notion of distributed source simulation where we require preserving correlations with the reference system, which we call 'coherent' DSS. We establish that for coherent DSS, entanglement assistance from the sender to each party is in general necessary to achieve arbitrarily small error. This allows us to identify 'coherent distributed source simulation' (coherent DSS) as a special case of *multipartite* QSS. Having made this identification, we establish a one-shot rate region for multipartite QSS in terms of smooth max multipartite mutual information and thus a rate region for coherent DSS as well. Finally, we establish a strong asymptotic equipartition property for our smooth multipartite max mutual information quantities. This establishes a strong converse rate region for multipartite QSS and coherent DSS.

a) Conventions: We follow standard notation [23]–[25]. The (subnormalized) density matrices on finite-dimensional joint Hilbert space $A \otimes B$ is denoted $\mathrm{D}(A \otimes B) := \{\rho_{AB} \in \mathrm{Pos}(A \otimes B) : \mathrm{Tr}[\rho_{AB}] = 1\}$, where we use subscripts to denote the relevant spaces. Normally we supress the Kronecker product in our notation, e.g. we may write $A_1^L B_1^2$ to denote $A_1 \otimes \ldots \otimes A_L \otimes B_1 \otimes B_2$. The subnormalized density matrices on AB are denoted $\mathrm{D}_{\leq}(AB)$. We denote the trace distance between two (normalized) states ρ, σ as $\Delta(\rho, \sigma)$ and the purified distance between two (subnormalized) states ρ, σ as $P(\rho, \sigma)$. We denote the space of completely positive, trace-preserving (CPTP) maps from endomorphisms on A to endomorphisms on B as $\mathrm{C}(A, B)$. The Hartley entropy of a register A of a state ρ_{AB} is $H_0(A)_\rho := \log(\mathrm{rank}(\rho_A))$. When we have $L \in \mathbb{N}$, spaces $\{A_i\}_{i\in[L]}$, integers $\{M_i\}_{i\in[L]} \subset \mathbb{N}$, and $S \subset [L]$, we define $M_S := \Pi_{\ell \in S} M_\ell$ and $A_S := \otimes_{\ell \in S} A_\ell$. We also denote any $A_1 - X - A_2$ (short) quantum Markov chains via a subscript, e.g. $\rho_{A_1 - X - A_2}$. Finally, throughout, we define any minimization over an empty set to be positive infinity.

II. Unassisted Incoherent Quantum State Splitting as Distributed Source Simulation

In this section we relate QSS modified to not require preservation of correlations with a reference system and without any entanglement assistance to DSS in the two party setting. We first define one-shot distributed source simulation as in [12] except, for consistency within this proceeding, our error measure is the trace distance rather than the L_1-distance.

Definition 1. *Let $\rho_{A_1^2} \in \mathrm{D}(A_1^2)$. Let $\varepsilon \in (0,1)$. The one-shot rate of distributed source simulation (ε-DSS) is the correlation of formation defined as*

$$\begin{aligned} &C_F^{\varepsilon}(A_1 : A_2)_{\rho} \\ &\quad = \min \left\{ H_0(X)_{\widetilde{\rho}} : \widetilde{\rho}_{A_1 - X - A_2} \,\&\, \Delta(\widetilde{\rho}_{A_1^2}, \rho_{A_1^2}) \leq \varepsilon \right\} . \end{aligned}$$

With this defined, we define our modified bipartite QSS protocol. It only differs from QSS in not considering the reference register in the error criterion, (1), and not providing entanglement assistance (see [17] for comparison).

Definition 2. *Let $\rho_{A_1^2 R}$ be a purification of $\rho_{A_1^2} \in \mathrm{D}(A_1^2)$. Consider a sender Alice and two receivers, Bobs i for $i \in \{1,2\}$, who implement the following protocol:*

1) *Alice holds quantum registers A_1^2 and R is an inaccessible reference system.*
2) *Alice applies a local operation to generate messages $\vec{r} = (r_1, r_2)$, where her local registers update to A'.*
3) *Alice sends message r_i to Bob i via noiseless one-way classical communication.*
4) *Upon receiving their message r_i, Bob i applies a local operation resulting in a joint state $\widehat{\rho}_{RA'B_1^2}$.*

A $(\vec{r}, \varepsilon)$-2 party incoherent quantum state splitting protocol without entanglement-assistance (IQSS) for $\rho_{A_1^2 R}$ satisfies

$$\frac{1}{2} \| \widehat{\rho}_{B_1^2} - \rho_{B_1^2} \|_1 \leq \varepsilon \ , \tag{1}$$

where $\rho_{B_1^2 R} := (id_{A_1^2 \to B_1^2} \otimes id_R)(\rho_{A_1^2 R})$. Moreover, we define the sum rate as

$$\mathbf{C}_F^{\varepsilon}(A_1 : A_2)_{\rho} := \{ \log(M_1^2) : \frac{1}{2} \| \widehat{\rho}_{B_1^2} - \rho_{B_1^2} \|_1 \leq \varepsilon \} \ .$$

We now reduce the rate region to being determined by one-shot DSS.

Proposition 3. *For any bipartite state $\rho_{A_1^2} \in \mathrm{D}(A_1^2)$, the sum rate $(\vec{r}, \varepsilon)$-2 party IQSS is*

$$\mathbf{C}_F^{\varepsilon}(A_1 : A_2)_{\rho} = 2 C_F^{\varepsilon}(A_1 : A_2)_{\rho} \ ,$$

and thus optimized by the one-shot DSS scheme.

Proof. First, assume a strategy exists for IQSS for the given target state $\rho_{A_1^2}$ and tolerated error $\varepsilon \in (0,1)$. Without loss of generality denote it

$$\widehat{\rho}_{B_1^2} = \sum_{r_1^2} p(r_1^2) \widehat{\rho}_{B_1}^{r_1} \otimes \widehat{\rho}_{B_2}^{r_2} \ , \tag{2}$$

where we have used that the local operations are solely conditioned on the received messages by Definition 2.

We now show (2) is always a (short) quantum Markov chain. Define the preparation channels $\mathcal{E}_1 : r_1 \mapsto \widehat{\rho}_{B_1}^{r_1}$, $\mathcal{E}_2 : r_1 \mapsto \sum_{r_2} p(r_2|r_1) \widehat{\rho}_{B_2}^{r_2}$, where the latter is a preparation channel by the normalization of conditional probability. By definition of these preparation channels,

$$\begin{aligned} &(\mathcal{E}_1 \otimes \mathcal{E}_2)(\sum_{r_1} p(r_1) |r_1\rangle\langle r_1|_{M_1} \otimes |r_1\rangle\langle r_1|_{M_2}) \\ &= \sum_{r_1} \left[p(r_1) \widehat{\rho}_{B_1}^{r_1} \otimes \sum_{r_2} p(r_2|r_1) \widehat{\rho}_{B_2}^{r_2} \right] \\ &= \sum_{r_1^2} p(r_1) p(r_2|r_1) \widehat{\rho}_{B_1}^{r_1} \otimes \widehat{\rho}_{B_2}^{r_2} \\ &= \widehat{\rho}_{B_1^2} \ , \end{aligned}$$

where the last equality is using the chain rule for conditional probabilities. Thus, we have $B_1 - M_1 - B_2$. By symmetry of this argument, we may also establish $B_1 - M_2 - B_2$. Thus the optimal strategy is for both parties to use the smallest Markov chain that approximates $\rho_{A_1^2}$ to the tolerated error. That is,

$$\begin{aligned} \mathbf{C}_F^{\varepsilon}(A_1 : A_2)_{\rho} &= 2 \cdot \min_{\widehat{\rho}_{A_1 - X - A_2} : \Delta(\rho_{A_1^2}, \widehat{\rho}_{A_1^2}) \leq \varepsilon} H_0(X)_{\widehat{\rho}} \\ &= 2 C_F^{\varepsilon}(A_1 : A_2)_{\rho} \ . \end{aligned}$$

This proves the optimal strategy is the optimal ε−DSS strategy. We note if there does not exist a feasible strategy as we had assumed, then our convention implies both rates are infinite and thus the result still holds. ☐

III. Coherent Distributed Source Simulation as Multipartite Quantum State Splitting

While Proposition 3 does not seem to generalize to the multipartite scenario,[1] it motivates the question of what changes when one considers distributed source simulation (DSS) where we require preserving the correlations with the reference system. We begin by noting such a task requires entanglement assistance in general.

Proposition 4. *Consider the task in Definition 2, but with the error criterion replaced with the new criterion*

$$\frac{1}{2} \| \widehat{\rho}_{B_1^2 R} - \rho_{B_1^2 R} \|_1 \leq \varepsilon \ , \tag{3}$$

where $\rho_{B_1^2 R} := (id_{A_1^2 \to B_1^2} \otimes id_R)(\rho_{A_1^2 R})$. Then the task cannot be performed to arbitrary error $\varepsilon \in (0,1)$ unless $\rho_{A_1^2}$ is a product state.

The result follows immediately from the fact the resulting joint state $\rho_{B_1^2 R}$ is necessarily fully separable but $\rho_{A_1^2 R}$ is entangled between the three registers unless the original state was product.

As we have established that preserving correlations between the reference system and the target to distribute is impossible without assistance, we now define a version of the task which has entanglement assistance.

[1] This is because the construction in the proof of Proposition 3 does not work for more parties.

Definition 5. *Let $\rho_{A_1^L R}$ be a purification of $\rho_{A_1^L} \in \mathrm{D}(A_1^L)$. Consider sender Alice and L receiver, Bobs 1 through L, who implement the following protocol:*

1) *Alice holds quantum registers A_1^L and R is an inaccessible reference system.*
2) *For each $i \in [L]$, an entangled state resource $|\tau\rangle_{\overline{A}_i\overline{B}_i}$ is shared between Alice (holding the $\overline{A}_i$ register) and the receiver Bob i.*
3) *Alice applies a local operation on the A_1^L registers and her portion of the shared entanglement to obtain $\vec{r} = (r_1, \cdots, r_L)$ (her local registers update to A').*
4) *Alice sends r_i to Bob i via noiseless one-way classical communication.*
5) *Upon receiving the messages, each Bob i applies a local operation to $\overline{B}_i$ conditioned on their r_i, resulting in a joint state $\widehat{\rho}_{RA'B_1^L}$.*

A $(\vec{r}, \varepsilon)$-L party coherent distributed source simulation (QDSS) protocol for $\rho_{A_1^L}$ using $\tau_{\overline{A}_1^L\overline{B}_1^L} := \otimes_{i\in[L]}\tau_{\overline{A}_i\overline{B}_i}$ satisfies

$$P_{\mathsf{QDSS}}(\rho_{A_1^L}\|\tau_{A_1^L}) := P(\widehat{\rho}_{B_1^L R}, \rho_{B_1^L R}) \leq \varepsilon \ , \tag{4}$$

where $\rho_{B_1^L R} := (id_{A_1^L\to B_1^L} \otimes id_R)(\rho_{A_1^L} R)$. We define the rate region of QDSS as

$$\left\{ \bigtimes_{\emptyset\neq S\subset[L]} \log(M_S) : P(\widehat{\rho}_{B_1^L R}, \rho_{B_1^L R}) \leq \varepsilon \right\} \ .$$

The above definition is justified as in standard DSS no two receivers share a correlation or channel between them and the sender only sends classical messages. Moreover, the only difference between QDSS and multipartite QSS is that in Definition 5 Alice does not have an extra register A that she wishes to keep (See [17] for formal comparison). Given its importance, we state this observation as a proposition.

Proposition 6. *A $(\vec{r}, \varepsilon)$-m party coherent DSS protocol is a $(\vec{r}, \varepsilon)$-m party state splitting protocol where Alice's A register is trivial.*

a) Coherent DSS Cannot Be Identified with Unipartite Convex Splitting/QSS: Proposition 6 identifies QDSS with *multipartite* DSS. This means that one really needs multipartite convex splitting to achieve QDSS even in the 2-party case. This may be surprising as it is known that *unipartite* convex splitting reduces to soft-covering when one of the registers is classical [16]. However, note that to apply soft-covering to establish multipartite DSS one uses that classical information may be cloned to in some sense ignore the network structure [12], [26]. With quantum states one would not expect this to be the case as teleporting a state to one party is not the same as teleporting a state to a different party.

A. Main Theorems and Multipartite Max Mutual Information Quantities

Our main result is a one-shot rate region for multipartite QSS and QDSS, which relies on the following families of smooth multipartite mutual information quantities, which make use of the smoothing ball $\mathcal{B}^\varepsilon(\rho_A) := \{\widetilde{\rho} \in \mathrm{D}_\leq(A) : P(\widetilde{\rho}, \rho) \leq \varepsilon\}$.

Definition 7. *Let $\rho_{A_1^L B} \in \mathrm{D}_\leq(A_1^L B)$ and $\tau_{A_1^L} \in \mathrm{D}_\leq(A_1^L)$. We define the modified smooth max multipartite mutual information as*

$$\mathcal{I}^\varepsilon_{\max}(\,: A_1^L : B)_{\rho|\tau} = \min_{\sigma\in\mathrm{D}(B)} D^\varepsilon_{\max}(\rho\| \otimes_i \tau_{A_i} \otimes \sigma_B) \ ,$$

and the 'down arrow' smooth max multipartite mutual information as

$$\begin{aligned} &I^{\downarrow,\varepsilon}_{\max}(\,: A_1^L : B)_\rho \\ &:= \min_{\widetilde{\rho}\in\mathcal{B}^\varepsilon(\rho)} \min_{\mathcal{Q}} D_{\max}(\widetilde{\rho}\|\tau_{A_1} \otimes \tau_{A_2} \ldots \otimes \tau_{A_L} \otimes \sigma_B) \ , \end{aligned}$$

where $\mathcal{Q} := \{\bigtimes_{i\in[L]} \tau_{A_i} \times \sigma_B \in \bigtimes_{i\in[L]} \mathrm{D}(A_i) \times \mathrm{D}(B)\}$.

We note that we use the symbol $\downarrow$ as a superscript as in the case there is only one A register, this reduces to $I^\downarrow_{\max}$ in the notation of [12]. We also remark that whenever we consider the marginal $\rho_{A_S B}$ where $A_S = \otimes_{i\in S} A_i$, we will simplify the mutual information quantity $\mathbb{I}$ we are considering evaluating over to the notation $\mathbb{I}(\,: A_S : B)$.

With the multipartite mutual information quantities and simplified notation defined, we state our main theorem.

Theorem 8. *For any $L \in \mathbb{N}$, the one-shot rate region for multipartite quantum state splitting of a target state $\rho_{AA_1^L} \in \mathrm{D}(AA_1^L)$ with error $\delta \in (0,1)$ and entanglement assistance induced by the state $\tau_{A_1^L} \in \mathrm{D}(A_1^L)$, satisfies for all $\emptyset \neq S \subset [L]$,*

$$\begin{aligned} &\mathcal{I}^{\varepsilon_S}_{\max}(: A_S : R)_{\rho_{A_S R}|\tau_{A_S}} + u(\delta) \\ &\qquad \geq \log(M_S) \\ &\qquad\qquad \geq I^{\downarrow,\delta}_{\max}(: C_{i_1} : \cdots C_{i_{|S|}} : R)_\rho \ . \end{aligned}$$

where $\rho_{A_1^L R}$ is an arbitrary purification of $\rho_{A_1^L}$, $\{\delta_S\}_{\emptyset\neq S\subset[L]} \subset (0,1)$ such that $\sum_{\emptyset\neq S\subset[L]} \delta_S \leq \delta$, $\varepsilon_S \in (0, 2^{1-|S|}\delta_S)$ for all $\emptyset \neq S \subset [L]$, and $u(\delta) := 2\log(\frac{1}{\kappa_S})$ where $\kappa_S := 2^{1-|S|}\delta_S - 2\varepsilon_S > 0$.

Using Proposition 6, we may conclude the following.

Corollary 9. *For any $L \in \mathbb{N}$ and target state $\rho_{A_1^L} \in \mathrm{D}(A_1^L)$, Theorem 8 provides inner and outer bounds on the rate region for $(\vec{r}, \varepsilon)$-L party coherent distributed source simulation.*

Moreover, we are able to establish a strong asymptotic equipartition property (AEP) for the mutual information quantities in Theorem 8 by generalizing results in [12], [27].

Theorem 10. *Consider any mutual information quantity $\mathbb{I}^\varepsilon_{\max}(\,: A_1^L : B)$ satisfying*

$$\begin{aligned} I^{\downarrow,\varepsilon}_{\max}(: A_1^L : B)_\rho - f_\varepsilon(n) \leq &\mathbb{I}^\varepsilon_{\max}(\,: A_1^L : B)_\rho \\ \leq &I_{\max}(\,: A_1^L : B)_\rho + g_\varepsilon(n) \ , \end{aligned}$$

where for every $\varepsilon \in (0,1)$, $f_\varepsilon(\cdot), g_\varepsilon(\cdot) \in o(n)$. Then,

$$\begin{aligned} \lim_{n\to\infty} \left[\frac{1}{n}\mathbb{I}^\varepsilon_{\max}(\,: A_1^L : B)_{\rho^{\otimes n}}\right] =&I(\,: A_1^L : B)_\rho \\ =:&D(\rho_{A_1^L B}\| \otimes_{i\in[L]} \rho_{A_i} \otimes \rho_B) \ . \end{aligned}$$

The main technical idea in establishing this is using the dominance property of $D_{\max}$ [25] to remove the need for a 'dual projector' in the proofs of [27] so that they extend to the multipartite case. Combining Theorems 8 and 10, one obtains a strong rate region for multipartite QSS and coherent DSS.

Theorem 11. *For any $\delta \in (0,1)$ and state $\rho_{A_1^L}$, the asymptotic rate region of multipartite quantum state splitting is the polytope*

$$\bigtimes_{\emptyset \neq S \subset [L]} [I(:A_S:R)_\rho), +\infty) \ ,$$

where $\rho_{A_1^L R}$ is an arbitrary purification of $\rho_{A_1^L}$.

Theorem 8 is established by combining Corollary 14 and Lemma 17. The achievability is established in three steps. The first step is to convert the error exponent statement for multipartite convex splitting in terms of Rényi multipartite mutual information quantities from [17] into an error exponent in terms of our $\mathcal{I}^{\varepsilon}_{\max}(:A_S:B)_{\rho|\tau}$ using roughly the same method as [22] uses to convert their Rényi entropy error exponents into smooth entropy error exponents. The second step is to establish error exponents for multipartite QSS from multipartite convex splitting. The last step is to convert our error exponent form into a small deviations achievability result. The converse follows the same proof method as the converse for channel simulation given in [17], [28], which relies on a few properties that we establish for our relevant quantity. The rest of the proceeding explains these proofs in as much detail as the space allows, and the rest is in the appendix for the reviewers.

B. Achievability Proof

The main technical step is to establish error exponents for multipartite convex splitting in terms of smoothed quantities. We do this by modifying the proof of [17, Theorem 3.12] to be more similar to that of the proof of [22, Theorem 3.2].

Lemma 12. *Consider $\rho_{A_1^L E}$ and $\tau_{A_1^L}$. Let $\{M_i\}_{i\in[L]} \subset \mathbb{N}$. Define*

$$\omega_{A_1^{M_1}\dots A_L^{M_L} E} := (M_L)^{-1} \sum_{i\in[L], m_i \in M_i} \rho_{A_{1,m_1}\dots A_{L,m_l} E} \cdot \left(\otimes_{\ell\in[L], \overline{m}_\ell \in M_\ell \setminus \{m_\ell\}} \tau_{A_\ell, \overline{m}_\ell}\right) \ ,$$

where $A_\ell^{M_\ell} := A_{\ell,1}\dots A_{\ell,M_\ell} \cong A_\ell^{\otimes M_\ell}$ for each $\ell \in [L]$. Consider smoothing parameters $\{\varepsilon_S\}_{\emptyset\neq S\subset[L]} \in [0,1]$, then we have the following error exponents for convex splitting:

$$\Delta_{M_S}(\omega, \otimes_{i\in[L]} \tau_{A_i} \otimes \rho_R) \leq \sum_{\emptyset \neq S \subset [L]} 2^{|S|} \cdot \left(2\varepsilon_S + 2^{-E^{\varepsilon_S}_{\log M_S}(\rho_{A_S R}\|\tau_{A_S})}\right) \ , \tag{5}$$

where the error exponent function is defined as

$$E^{\varepsilon_S}_r(\rho_{A_S}) := \sup_{p\in[1,2]} \frac{p-1}{p}(r - \mathcal{I}^{\varepsilon_S}_p(:A_1^L:E)_{\rho_{A_S}|\tau_{A_S}}) \ ,$$

and $I^{\varepsilon_S}_p(:A_1^L:R)$ is the same as the definition $I^{\downarrow,\varepsilon_S}_{\max}$, but with the max divergence replaced with a Sandwiched Rényi divergence with parameter p [25].

Proof. As explained in [17, Top of Page 18], by triangle inequality, we are interested in bounding the term

$$\|\Theta_{[L]} \circ T_S(\frac{\rho_{\mathbf{A}_S E}}{\tau_{A_S} \otimes E})\|_{L_1(\tau_{\mathbf{A}_S}\otimes\sigma_E)}$$

for each S where the norm is Kosaki's weight L_p-norm with respect to the density operator $\tau_{\mathbf{A}_S} \otimes \sigma_E$) (See [16], [17] for background) and $\tau_{\mathbf{A}_S} := \otimes_{i\in S} \tau_{A_i}^{\otimes M_i}$ and $\rho_{\mathbf{A}_S}$ is on the same space. The choice of $\tau_{A_1^L}$ must be consistent across every S. σ_E may be varied for each S, and we will make use of this.

Imagine we choose any $p \in [1,2]$. Now let $\widetilde{\rho}_{A_S}, \sigma_E$ be the optimizer for the modified smooth Rényi mutual information $I^{\varepsilon_S}_p(:A_S:E)_{\rho_{A_S}|\tau_{A_S}}$. Note this means $\frac{1}{2}\|\widetilde{\rho}_{A_S} - \rho_{A_S}\|_1 \leq \varepsilon_S$ since purified distance upper bounds trace distance. Now,

$$\begin{aligned}
&\|\Theta_{[L]} \circ T_S(\frac{\rho_{A_S E}}{\tau_{A_S}\otimes\sigma_E} - \frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E})\|_{L_1(\tau_{\mathbf{A}_S}\otimes\sigma_E)} \\
\leq& \|\Theta_{[L]} \circ T_S\| \cdot \|\frac{\rho_{A_S E}}{\tau_{A_S}\otimes\sigma_E} - \frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E}\|_{L_1(\tau_{\mathbf{A}_S}\otimes\sigma_E)} \\
\leq& 2^{|S|} \cdot \|\frac{\rho_{A_S E}}{\tau_{A_S}\otimes\sigma_E} - \frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E}\|_{L_1(\tau_{A_S}\otimes\sigma_E)} \\
=& 2^{|S|} \cdot \|\rho_{A_S E} - \widetilde{\rho}_{A_S E}\|_1 \\
\leq& 2^{|S|} \cdot 2\varepsilon_S \ ,
\end{aligned}$$

where the second inequality is [17, Lemmas 3.7 and 3.9], the equality is by definition of Kosaki's weighted L_p-norm for the weight $\gamma = 1/2$, and the final inequality is by our smoothing ball bound. Therefore, by triangle inequality, the inequality we just established, and the map norm,

$$\begin{aligned}
&\|\Theta_{[L]} \circ T_S(\frac{\rho_{A_S E}}{\tau_{A_S}\otimes\sigma_E})\|_{L_1(\tau_{\mathbf{A}_S}\otimes\sigma_E)} \\
\leq& 2^{|S|} 2\varepsilon_S + \|\Theta_{[L]} \circ T_S(\frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E})\|_{L_1(\tau_{\mathbf{A}_S}\otimes\sigma_E)} \\
=& 2^{|S|}\left(2\varepsilon_S + \|\frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E}\|_{L_1(\tau_{\mathbf{A}_S}\otimes\sigma_E)}\right) \ .
\end{aligned}$$

By complex interpolation provided in [17, Lemma 3.10], we can replace $\|\frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E}\|_{L_1(\tau_{A_S}\otimes\sigma_E)}$ with $M_{[L]}^{\frac{1-p}{p}} \|\frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E}\|_{L_p(\tau_{A_S}\otimes\sigma_E)}$ for any choice of $p \in [1,2]$.

Now note by definition of Kosaki's weighted L_p-norm for the choice of $\gamma = 1/2$, one can verify [16, Eqs. 113-115]

$$\left\|\frac{\widetilde{\rho}_{A_S E}}{\tau_{A_S}\otimes\sigma_E}\right\|_{L_p(\widetilde{\rho}_{A_S E})} = \exp\left(\frac{p-1}{p} D_p(\widetilde{\rho}_{A_S E}\|\tau_{A_S}\otimes\sigma_E)\right) \ .$$

Finally, note that for every S we chose $\widetilde{\rho}_{A_S E}$ to optimize $\mathcal{I}^{\varepsilon}_p(:A_S:E)_{\rho|\tau}$. Noting there is actually a factor of a half in front of each norm, re-organizing everything in the exponents, and summing over the $\emptyset \neq S \subset [L]$ completes the proof. $\square$

We can now convert the previous lemma to establish an error exponent result for multipartite QSS, where error is measured in purified distance and we use the same error exponents as in the previous lemma. This makes use of Uhlmann's theorem and is similar to proofs in [16], [17].

Lemma 13. *Given a pure state $\rho_{AA_1^L R}$, there exists a $(\vec{r},\varepsilon)$-L-receiver quantum state splitting protocol using entangled resources $|\tau\rangle^{\otimes[M_i]}_{A_i'B_i}$ for $i\in[L]$ such that*

$$P_{\mathsf{Split}}(\rho\|\tau)\leq\sum_{\emptyset\neq S\subset[L]}2^{|S|-1}\cdot\left(2\varepsilon_S+2^{-E^{\varepsilon_S}_{\log M_S}(\rho_{A_S E}\|\tau_{A_S})}\right)\ ,$$

where

$$P_{\mathsf{Split}}(\rho_{AA_1^L R}\|\tau_{A_1^L}):=P(\widehat{\rho}_{RAB_1^L},\rho_{RAB_1B_2})\leq\varepsilon\ , \tag{6}$$

and $\widehat{\rho}_{AB_1^L R}$ is the output of the protocol where A is a register Alice holds onto and $B_i\cong A_i$ for all $i\in[L]$.

Finally, we can convert the error exponents into a small deviations achievable rate region.

Corollary 14. *Let $\delta\in(0,1)$ and a $\delta_S\in(0,1)$ for each $\emptyset\neq S\subset[L]$ such that $\sum_{\emptyset\neq S\subset[L]}\delta_S=\delta$. Let $\varepsilon_S\in(0,2^{1-|S|}\delta_S)$ for each S. Then there exists a $\delta-$approximate state splitting scheme in purified distance satisfying $\forall\emptyset\neq S\subset[L]$,*

$$\log(M_S)\leq\mathcal{I}^{\varepsilon_S}_{\max}(:A_S:E)_{\rho_{A_S E}|\tau_{A_S}}+2\log\left(\frac{1}{\kappa_S}\right)\ ,$$

where $\kappa_S:=2^{1-|S|}\delta_S-2\varepsilon_S>0$.

Proof. First note that we can always relax the error function from $I^{\delta_S}_p$ to $I^{\delta_S}_{\max}$ as the Sandwiched divergence monotonically increases in parameter p. In this case, we can always set $p=2$ as that always gives the loosest bound. Thus, we have replaced each error exponent term with the new the error exponent function

$$E^{\varepsilon_S}_{\max,r}(\rho_{A_S E}\|\tau_{A_S})=\frac{1}{2}(r-I^{\varepsilon_S}_{\max}(:A_S:E)_{\rho|\tau})\ .$$

Now we fix a total $\delta\in(0,1)$ and a $\delta_S\in(0,1)$ for each $\emptyset\neq S\subset[L]$ such that $\sum_{\emptyset\neq S\subset[L]}\delta_S=\delta$. Let $\varepsilon_S\in(0,2^{1-|S|)}\delta)$ for each S and define $\kappa_S:=2^{1-|S|}\delta_S-2\varepsilon_S>0$. Now all we need is each term in the error exponent to be less than the corresponding δ_S, which we can solve for:

$$\begin{aligned}
&\delta_S\geq2^{|S|-1}\cdot\left(2\varepsilon_S+2^{-E^{\varepsilon_S}_{\max,\log M_S}(\rho_{A_S E}\|\tau_{A_S})}\right)\\
\Leftrightarrow&2^{1-|S|}\delta_S-2\varepsilon_S\geq2^{-E^{\varepsilon_S}_{\max,\log M_S}(\rho_{A_S E}\|\tau_{A_S})}\\
\Leftrightarrow&\log(\kappa_S)\geq\frac{1}{2}\left(\log(M_S)-\mathcal{I}^{\varepsilon_S}_{\max}(:A_S:E)_{\rho_{A_S E}|\tau_{A_S}}\right)\\
\Leftrightarrow&\mathcal{I}^{\varepsilon_S}_{\max}(:A_S:E)_{\rho_{A_S E}|\tau_{A_S}}+2\log(\frac{1}{\kappa_S})\geq\log(M_S)\ ,
\end{aligned}$$

where the second equivalence is the definition of κ_S. ☐

C. Converse Proof

The converse follows from the structure of the entanglement assistance, monotonicity under local maps for $I^{\downarrow,\varepsilon}_{\max}(:A_1^L:B)_\rho$, and the following chain rule.

Proposition 15 (Classical-Specific Chain Rule)**.** *Let $\varepsilon\in[0,1)$. Consider $\rho_{A_1^L X_1^L B}$, where the register sizes may vary across index i. Then*

$$\begin{aligned}
&I^{\downarrow,\varepsilon}_{\max}(A_1X_1:A_2X_2:\cdots:A_LX_L:B)_\rho\\
&\qquad\leq I^{\downarrow,\varepsilon}_{\max}(:A_1^L:B)_\rho+\sum_i\log|X_i|
\end{aligned}$$

This chain rule may be proven by building a feasible point of the optimization for $I^{\downarrow,\varepsilon}_{\max}(A_1X_1:A_2X_2:\cdots:A_LX_L:B)_\rho$ using the optimizer of $I^{\downarrow,\varepsilon}_{\max}(:A_1^L:B)_\rho$.

General multipartite mutual information quantities induced by divergences satisfying data processing are monotonic under local maps via a straightforward proof. We state the specific case we will use.

Proposition 16. *For $\rho\in\mathrm{D}_{\leq}(A_1^LB)$, $\varepsilon\in[0,1)$, and CPTP maps $\{\mathcal{E}_i\in\mathrm{C}(A_i,A_i')\}_{i\in[L]}$, $\mathcal{F}\in\mathrm{C}(B,B')$,*

$$I^{\downarrow,\varepsilon}_{\max}(:A_1^L:B)_\rho\geq I^{\downarrow,\varepsilon}_{\max}(:{A'}_1^{L}:B')_{\otimes_i\mathcal{E}_i\otimes\mathcal{F}(\rho)}$$

We now establish our converse.

Lemma 17 (Converse Rate Region)**.** *Let $\delta\in(0,1)$ be the error of the state splitting in purified distance. Then,*

$$\log(M_S)\geq I^{\downarrow,\delta}_{\max}(:C_{i_1}:\cdots C_{i_{|S|}}:R)_\rho\quad\forall\emptyset\neq S\subset[L]\ .$$

Proof. This follows the converse for broadcast channel simulation in [17], which follows [28]. Let $\mathcal{E}$ be Alice's encoder and $\mathcal{D}_i$ be party i's local decoder. Define the final state as $\widehat{\rho}$. Note that by definition of state splitting, the original entangled resource is assumed to be such that each party *individually* shares entangled states with each of the encoders just as in Definition 5. This means the registers $\overline{B}_i$ and $\overline{B}_{i'}$ are independent for $i\neq i'$ and each of these registers is also independent of the R register. Thus, any mutual information quantity between the $\overline{B}_i$ registers and R is zero. Also note that by partial trace $\tau_{\overline{B}_1^L R}=(\mathcal{E}(\rho\otimes\tau))_{\overline{B}_1^L}=\otimes_{i\in[L]}\tau_{\overline{B}_i}\otimes\rho_R$. That is, this marginal does not change under the action of the encoder. In this case, for any $\emptyset\neq S\subset[L]$,

$$\begin{aligned}
&\sum_{i\in[S]}\log(M_i)\\
=&\sum_{i\in[S]}\log(M_i)+I^{\downarrow}_{\max}(:\overline{B}_{i_1}:\cdots\overline{B}_{i_{|S|}}:R)_\tau\\
\geq&\sum_{i\in[S]}\log(M_i)+I^{\downarrow}_{\max}(:\overline{B}_{i_1}:\cdots\overline{B}_{i_{|S|}}:R)_{\mathcal{E}(\rho\otimes\tau)}\\
\geq&I^{\downarrow}_{\max}(:C_{i_1}:\cdots C_{i_{|S|}}:R)_{\otimes_i\mathcal{D}_i\circ\mathcal{E}(\rho\otimes\tau)}\\
=&I^{\downarrow}_{\max}(:C_{i_1}:\cdots C_{i_{|S|}}:R)_\sigma\\
\geq&I^{\downarrow,\delta}_{\max}(:C_{i_1}:\cdots C_{i_{|S|}}:R)_\rho,
\end{aligned}$$

where the first inequality is Proposition 15, the second is the local data-processing property (Proposition 16), and the third is noting we require the output state be δ close to the target state ρ in purified distance, so we may minimize over that smoothing ball to get a lower bound. ☐

Acknowledgment

I.G. is supported by NSF Grant No. 2112890. H.-C. Cheng is supported by Grants No. NSTC 112-2636-E-002-009, No. NSTC 112-2119-M-007-006, No. NSTC 112-2119-M-001-006, No. NSTC 112-2124-M-002-003, No. NTU-112V1904-4, NTU-CC-113L891605, and NTU-113L900702.

References

[1] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[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*, vol. 465, no. 2108, pp. 2537–2563, 2009.

[3] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 569–579, 2002.

[4] I. Devetak and A. Winter, "Classical data compression with quantum side information," *Physical Review A*, vol. 68, no. 4, p. 042301, 2003.

[5] N. Cai, A. Winter, and R. W. Yeung, "Quantum privacy and quantum wiretap channels," *problems of information transmission*, vol. 40, pp. 318–336, 2004.

[6] A. Winter, "Secret, public and quantum correlation cost of triples of random variables," in *Proceedings. International Symposium on Information Theory, 2005. ISIT 2005.* IEEE, 2005, pp. 2270–2274.

[7] M. Hayashi, *Quantum Information: An Introduction*. Springer, 2006.

[8] P. Cuff, "Communication requirements for generating correlated random variables," in *2008 IEEE International Symposium on Information Theory*, 2008, pp. 1393–1397.

[9] Z. Luo and I. Devetak, "Channel simulation with quantum side information," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1331–1342, 2009.

[10] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[11] 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*, vol. 60, no. 5, pp. 2926–2959, 2014.

[12] I. George, M.-H. Hsieh, and E. Chitambar, "One-shot distributed source simulation: As quantum as it can get," 2023.

[13] M. Berta, M. Christandl, and R. Renner, "The quantum reverse shannon theorem based on one-shot information theory," *Communications in Mathematical Physics*, vol. 306, pp. 579–615, 2011.

[14] A. Anshu, V. K. Devabathini, and R. Jain, "Quantum communication using coherent rejection sampling," *Physical review letters*, vol. 119, no. 12, p. 120506, 2017.

[15] A. Anshu, M. Berta, R. Jain, and M. Tomamichel, "Partially smoothed information measures," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 5022–5036, 2020.

[16] H.-C. Cheng and L. Gao, "Tight one-shot analysis for convex splitting with applications in quantum information theory," *arXiv preprint arXiv:2304.12055*, 2023.

[17] H.-C. Cheng, L. Gao, and M. Berta, "Quantum broadcast channel simulation via multipartite convex splitting," 2023.

[18] H.-C. Cheng and L. Gao, "Error exponent and strong converse for quantum soft covering," *IEEE Transactions on Information Theory*, 2023.

[19] Y.-C. Shen, L. Gao, and H.-C. Cheng, "Optimal second-order rates for quantum soft covering and privacy amplification," *IEEE Transactions on Information Theory*, pp. 1–1, 2024.

[20] F. Leditzky, M. M. Wilde, and N. Datta, "Strong converse theorems using rényi entropies," *Journal of Mathematical Physics*, vol. 57, no. 8, 2016.

[21] N. Ramakrishnan, M. Tomamichel, and M. Berta, "Moderate deviation expansion for fully quantum tasks," *IEEE Transactions on Information Theory*, 2023.

[22] P. Colomer Saus and A. Winter, "Decoupling by local random unitaries without simultaneous smoothing, and applications to multi-user quantum information tasks," *arXiv e-prints*, pp. arXiv–2304, 2023.

[23] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.

[24] M. M. Wilde, *Quantum information theory*. Cambridge university press, 2013.

[25] M. Tomamichel, *Quantum information processing with finite resources: mathematical foundations*. Springer, 2015, vol. 5.

[26] W. Liu, G. Xu, and B. Chen, "The common information of n dependent random variables," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2010, pp. 836–843.

[27] N. Ciganović, N. J. Beaudry, and R. Renner, "Smooth max-information as one-shot generalization for mutual information," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1573–1581, 2013.

[28] M. X. Cao, N. Ramakrishnan, M. Berta, and M. Tomamichel, "Channel simulation: Finite blocklengths and broadcast channels," *arXiv preprint arXiv:2212.11666*, 2022.

[29] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.

[30] A. McKinlay and M. Tomamichel, "Decomposition rules for quantum Rényi mutual information with an application to information exclusion relations," *Journal of Mathematical Physics*, vol. 61, no. 7, 2020.

[31] A. McKinlay, "Rényi divergence inequalities via interpolation, with applications to generalised entropic uncertainty relations," *arXiv preprint arXiv:2106.10415*, 2021.

[32] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, "Leftover hashing against quantum side information," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5524–5535, 2011.

[33] M. Tomamichel, "A framework for non-asymptotic quantum information theory," *arXiv preprint arXiv:1203.2142*, 2012.

APPENDIX

Here we provide the extra proofs for an interested reviewer. We first provide the proof that entanglement assistance is necessary for coherent disitributed source simulation (Proposition 4). We then provide omitted proofs for establishing Theorem 8. The rest of the appendix is proving Theorem 10.

Proof of Proposition 4. We prove the bipartite case for clarity, but a multipartite extension follows the same proof. The result follows immediately from the fact the resulting joint state $\rho_{B_1^2 R}$ is necessarily fully separable but $\rho_{A_1^2 R}$ is entangled between the three registers in general. To formally show this, note that Alice performs a local CPTP map $\mathcal{E}_{A_1^2 \to A'^2_1 M_1^2}$, resulting in the state

$$\rho_{RA'^2_1 M_1^2} = \sum_{m_1,m_2} p(m_1,m_2)|m_1\rangle\langle m_1| \otimes |m_2\rangle\langle m_2| \otimes \rho^{(m_1,m_2)}_{RA'^2_1} \ .$$

After the messages are transmitted, the Bobs apply local maps $\mathcal{D}_{M_i \to B_i}$, resulting in the state

$$\rho_{RA'^2_1 B_1^2} = \sum_{m_1,m_2} p(m_1,m_2)\rho^{m_1}_{B_1} \otimes \rho^{m_2}_{B_2} \otimes \rho^{(m_1,m_2)}_{RA'^2_1} \ .$$

By definition, this state is fully separable across the subsystems B_1, B_2, and R. As the set of fully separable states is closed,[2] if $\rho_{RA_1^2}$ was entangled between any two registers, one cannot approach the target state to arbitrary precision under the restrictions of the protocol. The only case where the original state is fully separable is when $\rho_{A_1^2}$ is a product state, i.e. $\rho_{A_1^2} = \otimes_{i\in\{1,2\}}|\psi_i\rangle\langle\psi_i|_{A_i}$. This completes the proof. □

A. Proofs for Theorem 8

Proof of Lemma 13. We prove the bipartite case, the multipartite argument is an immediate extension. Let $M_1 = 2^{r_1}$, $M_2 = 2^{r_2}$. Let $\tau_{A'_i} = \rho_{A_i}$ for $i \in [2]$. This means $|\tau\rangle_{A'_i B'_i}$ is a purification of ρ_{A_i}. Let Alice and receiver i (Bob i) share M_i copies of $|\tau\rangle_{A'_i B'_i}$, denoted $\otimes_{m\in[M_i]}|\tau\rangle_{A'_{i,m}B'_{i,m}}$, where Alice has the $A'_{i,m}$ systems and Bob has the $B'_{i,m}$ systems. Thus, the initial global state is

$$|\overline{\omega}\rangle = |\rho\rangle_{AA'_1A'_2R} \otimes_{m\in[M_1]} |\tau\rangle_{A'_{1,m}B'_{1,m}} \otimes_{m\in[M_2]} |\tau\rangle_{A'_{2,m}B'_{2,m}} \ . \tag{7}$$

Now the goal is to perform a process such that one ends with the final pure state

$$\begin{aligned} |\omega\rangle := \frac{1}{\sqrt{M_{[2]}}} &\sum_{\substack{m_1\in M_1\\ m_2\in M_2}} |m_1\rangle_{M_1}|m_2\rangle_{M_2}|\rho\rangle_{AB_{1,m_1}B_{2,m_2}R} \\ &\cdot |0\rangle_{A'_{1,m_1}A'_{2,m_2}} \otimes_{\overline{m_1}\in[M_1]\setminus\{m_1\}} |\tau\rangle_{A'_{1,\overline{m}_1}B_{1,\overline{m}_1}} \\ &\otimes_{\overline{m_2}\in[M_2]\setminus\{m_2\}} |\tau\rangle_{A'_{2,\overline{m}_2}B_{2,\overline{m}_2}} \ , \end{aligned} \tag{8}$$

where $M_{[2]} := \pi_{\ell\in[2]}M_\ell$, Alice holds registers $M_1, M_2, A, A'_{1,[M_1]} := A'_{1,1} \otimes \dots A'_{1,M_1}$, and $A'_{2,[M_2]} := A_{2,1} \otimes \dots A'_{2,M_2}$, and Bob holds registers $B_{1,[M_1]} := B_{1,1} \otimes \dots B_{1,M_1}$, and $B_{2,[M_2]} := B_{2,1} \otimes \dots B_{2,M_2}$. If this is the case, then Alice can measure registers M_1, M_2 and send outcome m_i to receiver i (Bob i) using r_i bits. Then the two receivers can simply keep registers B_{i,m_i} so that the global state is $|\rho\rangle_{AB_{1,m_1}B_{2,m_2}R} \cong |\rho\rangle_{AB_1B_2R}$.

The rest of the proof is showing there is a process that approximates $|\omega\rangle$ (8) to an error given in the theorem statement. First note that a marginal of $|\overline{\omega}\rangle$ (7) is

$$\overline{\omega}_{A'_{1,[M_1]}A'_{2,[M_2]}R} = \otimes_{m_1\in[M_1]}\tau_{A'_{1,m_1}} \otimes_{m_2\in[M_2]} \tau_{A'_{2,m_2}} \otimes \rho_R \ . \tag{9}$$

Lemma 12 tells us that we can approximate the above marginal by the density matrix

$$\begin{aligned} &\omega_{A_{1,M_1}'A'_{2,[M_2]}R} \\ =&\frac{1}{M_{[2]}} \sum_{(m_1,m_2)\in[M_1]\times[M_2]} \rho_{A_{1,m_1}'A'_{2,m_2}R} \\ &\otimes_{\overline{m_1}\in[M_1]\setminus\{m_1\}} \tau_{A'_{1,\overline{m}_1}} \otimes_{\overline{m_2}\in[M_2]\setminus\{m_2\}} \tau_{A'_{2,\overline{m}_2}} \end{aligned} \tag{10}$$

up to trace distance error ε' that is upper bounded by the right hand side of (5) via the replacements $A \to A'_1$, $B \to A'_2$, $E \to R$. Note that $\omega_{A_{1,M_1}'A'_{2,[M_2]}R}$ is a reduced state of ideal final state (8). As already noted, we have $|\rho\rangle_{AB_{1,m_1}B_{2,m_2}R} \cong |\rho\rangle_{AB_1B_2R}$, which implies $\rho_{AB_1B_2R} \cong \rho_{AB_1B_2R}$. Combining these points, $\omega_{AA'_1A'_2R}$ is the marginal of $|\omega\rangle$ such that $\omega_{AB_1B_2} \cong \overline{\omega}_{AA'_1A'_2}$. It follows by Uhlmann's theorem, there is an isometry $V : AA'_1A'_2A'_{1,[M_1]}A'_{2,[M_2]} \to M_1M_2AA'_{1,[M_1]}A'_{2,[M_2]}$ such that $\frac{1}{2}\|V|\overline{\omega}\rangle\langle\overline{\omega}|V^* - |\omega\rangle\langle\omega|\|_1 \leq \sqrt{2\varepsilon'}$. Replacing ε' with the error exponent in Lemma 12 gets the trace distance bound. Alternatively, ε' bounds the purified distance directly, which is what is stated. □

Proof of Proposition 15. We prove the two party case as it's notationally simpler and the method immediately generalizes. By definition, we have

$$\begin{aligned} &\exp(I^{\downarrow,\varepsilon}_{\max}(A_1X_1 : A_2X_2 : B)_\rho) \\ =& \min_{\substack{\widetilde{\rho}\in\mathcal{B}^\varepsilon(\rho)\\ \{\tau_{A_iX_i}\in D(A_iX_i)\}_{i\in[2]}\\ \sigma_B\in\mathrm{D}(B)}} \left\{\lambda : \lambda\tau_{A_1X_1} \otimes \tau_{A_1^2X_1^2} \otimes \sigma_B \succeq \widetilde{\rho}_{A_1^2X_1^2B}\right\} \\ =& \min_{\substack{\widetilde{\rho}\in\mathcal{B}^\varepsilon(\rho)\\ \{q_{X_i}\in\mathcal{P}(\mathcal{X}_i)\}_{i\in[2]}\\ \{\tau_{A_i}\in D(A_i)\}_{i\in[2]}\\ \sigma_B\in\mathrm{D}(B)}} \Big\{\lambda : \lambda q_{X_1}(x_1)q_{X_2}(x_2)\tau^{x_1}_{A_1} \otimes \tau^{x_2}_{A_2} \otimes \sigma_B \\ &\qquad \succeq \widetilde{p}_{X_1^2}(x_1^2)\widetilde{\rho}^{x_1^2}_{A_1^2B} \ \forall x \in \mathcal{X}\Big\} \ , \end{aligned} \tag{11}$$

where we have used the block diagonality and we have labeled the distributions' spaces for clarity.

Let $\log(\mu) = I^{\downarrow,\varepsilon}_{\max}(A_1 : A_2)_\rho$. That is there exists $\widehat{\rho}_{A_1^2} \in \mathcal{B}^\varepsilon(\rho_{A_1^2})$, $\{\sigma_{A_i} \in D(A_i)\}_{i\in[2]}$, $\sigma_B \in \mathrm{D}(B)$ such that $\mu\sigma_{A_1} \otimes \sigma_{A_2} \otimes \sigma_B \succeq \widehat{\rho}_{A_1^2B}$. Now note that by a standard property of purified distance [25], there exists an extension of $\widetilde{\rho}_{A_1^2}$, $\widehat{\rho}_{A_1^2X_1^2B}$, such that $P(\widehat{\rho}_{A_1^2X_1^2B}, \rho_{A_1^2X_1^2B}) \leq \varepsilon$. Moreover, by

[2]This follows from a direct extension of the proof of bipartite separable states being closed in [29].

DPI of purified distance, this $\widehat{\rho}$ has the same QCQ structure as ρ. Then we may write

$$\mu\sigma_{A_1} \otimes \sigma_{A_2} \otimes \sigma_B \succeq \widehat{\rho}_{A_1^2 B} = \sum_{x_1^2 \in \mathcal{X}_1 \times \mathcal{X}_2} \widehat{p}_{X_1^2}(x_1^2)\widehat{\rho}^{x_1^2}_{A_1^2 B} \ , \tag{12}$$

where we just used the structure of our CQ extension. Next note that for any distributions $\{\widehat{q}_{X_1} \in \mathcal{P}(\mathcal{X}_i)\}_{i\in[2]}$, we have

$$\begin{aligned}
&\mu\sigma_{A_1} \otimes \sigma_{A_2} \otimes \sigma_B \\
=&\mu\left(\sum_{x_1\in\mathcal{X}_1} q_{X_1}(x_1)\right)\sigma_{A_1} \otimes \left(\sum_{x_1\in\mathcal{X}_2} q_{X_2}(x_2)\right)\sigma_{A_2} \otimes \sigma_B \\
\preceq&\frac{\mu}{\widehat{q}_{\min,1}\cdot\widehat{q}_{\min,2}} \sum_{x_1^2\in\mathcal{X}^2} q_{X_1}(x_1)q_{X_2}(x_2)\sigma_{A_1}\otimes\sigma_{A_2}\otimes\sigma_B \\
=:&\frac{\mu}{\widehat{q}_{\min,1}\cdot\widehat{q}_{\min,2}} q_{X_1}(x_1)q_{X_2}(x_2) \otimes \widehat{\tau}^{x_1}_{A_1} \otimes \widehat{\tau}^{x_2}_{A_2} \otimes \sigma_B \ ,
\end{aligned} \tag{13}$$

where the first line is by normalization, the second is using $q_{X_1}(x_i)/\widehat{q}_{\min,1} \geq 1$ where $\widehat{q}_{\min,1} := \min_x q_{X_1}(x_1)$ and similarly indexed for 2, and the third is just defining $\widehat{\tau}^{x_1}_{A_1} := \sigma_{A_1}$ for all x_1 and again similarly or the second space.

Now combining the relations (12) and (13) defines a feasible solution $\left(\widehat{q}_{X_1},\ \widehat{q}_{X_2},\ \{\widehat{\tau}^{x_1}_{A_1}\}_{x_1\in\mathcal{X}_1},\ \{\widehat{\tau}^{x_2}_{A_2}\}_{x_2\in\mathcal{X}_2}, \sigma_B,\ \widehat{\rho}_{A_1^2 X_1^2 B}\right)$ for (11). Using that $I^{\downarrow,\varepsilon}_{\max}$ is a minimization problem,

$$\begin{aligned}
&I^{\downarrow,\varepsilon}_{\max}(:A_1^L X_1^L : B)_\rho \\
\leq&\log\left(\frac{\mu}{\widehat{q}_{\min,1}\cdot\widehat{q}_{\min,2}}\right) \\
\leq&I^{\downarrow,\varepsilon}_{\max}(:A_1^2 : B)_\rho + \log(\widehat{q}_{\min,1}) + \log(\widehat{q}_{\min,2}) \ .
\end{aligned}$$

Finally, since we had freedom in our choice of $\widehat{q}_{X_1}, \widehat{q}_{X_2}$, we may choose them to be the uniform distribution so that the minimum probability is $|\mathcal{X}_i|^{-1}$ for each i. ☐

B. Proofs for Theorem 10

Theorem 10 is largely about establishing the converse. This may be decomposed into two steps. The first step is to relate the down arrow smooth mutual information quantity to a third intermediary mutual information quantity introduced in this appendix via chain rules. This is done by extending results from [27]. Our technical contribution is the removal of needing the notion of a 'dual projector' to establish these results, which does not work well in the multipartite case. We do this by using the dominance property of $D_{\max}$ to bound the original quantity by instead bounding a quantity conditioned on a projected version of the state. The second step is to decompose the smooth version of the intermediary mutual information quantity into min-entropic terms. This is a straightforward generalization of results in [12], [13], [27].

1) Step 1: Relating Mutual Information Quantities: We state the following property of Rényi divergences, sometimes called 'dominance' [25].

Fact 18 (Dominance Property)**.** *Let $P, Q, Q' \in \mathrm{Pos}(A)$ such that $Q \preceq Q'$. Then $D_{\max}(P\|Q) \geq D_{\max}(P\|Q')$.*

We now define the intermediary mutual information quantity we will need.

Definition 19. *Let $\rho_{A_1^L B} \in \mathrm{D}(A_1^L B)$ and $\tau_{A_1^L} \in \mathrm{D}(A_1^L)$. The smooth 'up arrow' multipartite max mutual information is*

$$\begin{aligned}
&I^{\uparrow,\varepsilon}_{\max}(:A_1^L : B)_\rho \\
&\quad := \min_{\widetilde{\rho}\in\mathcal{B}^\varepsilon(\rho)} \min_{\sigma\in\mathrm{D}(B)} D_{\max}(\widetilde{\rho}\| \otimes_{i\in[L]} \widetilde{\rho}_{A_i} \otimes \sigma_B) \ .
\end{aligned}$$

We remark this quantity is denoted this way as it simplifies to $I^{\uparrow,\varepsilon}_{\max}$ in the notation of [30], [31] in the bipartite case.

Lemma 20. *Let $\rho \in \mathrm{D}(A_1^L B)$. For $\varepsilon \in (0,1)$, $\varepsilon' \geq 0$ such that $\varepsilon + \varepsilon' \in (0,1)$,*

$$\begin{aligned}
I^{\downarrow,\varepsilon+\varepsilon'}_{\max}(:A_1^L : B)_\rho &\leq I^{\uparrow,\varepsilon+\varepsilon'}_{\max}(:A_1^L : B)_\rho \\
&\leq I^{\downarrow,\varepsilon+\varepsilon'}_{\max}(:A_1^L : B)_\rho + f(L,\varepsilon,\varepsilon') \ ,
\end{aligned} \tag{14}$$

where $f(L,\varepsilon,\varepsilon') := \log\left(L(1-\sqrt{1-\varepsilon^2})^{-1} + (1-\varepsilon')^{-1}\right)$.

The basic idea of proving Lemma 20 was worked out in [27] and is to construct a state $\overline{\rho} \in \mathcal{B}^\varepsilon(\rho)$ with nice structure such that we can upper bound its max divergence with the down arrow mutual information of ρ. We will use the following two propositions.

Proposition 21. *[32, Lemma 17] For any $\rho \in \mathrm{D}_{\leq}(A)$ and any projector Π on A,*

$$P(\rho, \Pi\rho\Pi) \leq \sqrt{2\mathrm{Tr}[\Pi^\perp\rho] - \mathrm{Tr}[\Pi^\perp\rho]^2} \ , \tag{15}$$

where $\Pi^\perp = \mathbb{1} - \Pi$.

Proposition 22. *[33] For any $P, Q \in \mathrm{Pos}(A)$ and any projector Π on A,*

$$\|\sqrt{P}\sqrt{\Pi Q\Pi}\|_1 = \|\sqrt{\Pi P\Pi}\sqrt{\Pi Q\Pi}\|_1 \ . \tag{16}$$

Proof of Lemma 20. The lower bound is immediate, so we just prove the upper bound. First we construct our candidate (subnormalized) state in the ball. Define

$$\widetilde{\rho}_{A_1^L B} := (\otimes_i \Pi_{A_i})\, \rho_{A_1^L B}\, (\otimes_i \Pi_{A_i}) \ ,$$

where each Π_{A_i} is any projector such that $\mathrm{Tr}[(\mathbb{1}_{A_i} - \Pi_{A_i})\rho_{A_i}] \leq 1 - \sqrt{1-\varepsilon^2}$ for each i. Define $\Pi^\perp_{A_i} := (\mathbb{1}_{A_i} - \Pi_{A_i})$ and $\Pi^\perp_{A_1^L} := \otimes_{i\in[L]}\Pi^\perp_{A_i}$. Define $\Delta_{A_1^L} := \otimes_{i\in[L]}(\rho_{A_i} - \widetilde{\rho}_{A_i}) \succeq 0$. Now define $\overline{\rho}_{A_1^L B} := \widetilde{\rho}_{A_1^L B} + \Delta_{A_1^L} \otimes \overline{\sigma}_B$.

First, $\overline{\rho} \succeq \widetilde{\rho}$, so $\|\sqrt{\rho}\sqrt{\overline{\rho}}\|_1 \geq \|\sqrt{\rho}\sqrt{\widetilde{\rho}}\|_1$. Second,

$$\begin{aligned}
1 - \mathrm{Tr}[\overline{\rho}] =& 1 - \mathrm{Tr}[\widetilde{\rho}] - \prod_{i\in[L]}(\mathrm{Tr}[\rho_{A_i}] - \mathrm{Tr}[\widetilde{\rho}_{A_i}]) \\
\geq& 1 - \mathrm{Tr}[\widetilde{\rho}] - (\mathrm{Tr}[\rho_{A_1}] - \mathrm{Tr}[\widetilde{\rho}_{A_1}]) = 1 - \mathrm{Tr}[\rho] \ ,
\end{aligned}$$

where we used the multiplicativity of trace over tensor products and that $(\mathrm{Tr}[\rho_{A_i}] - \mathrm{Tr}[\widetilde{\rho}_{A_i}]) \in (0,1]$, so any power of it is smaller or the same value. Using these points, we obtain

$$
\begin{aligned}
F_*(\rho,\overline{\rho}) &= \left(\|\sqrt{\rho}\sqrt{\overline{\rho}}\|_1 + \sqrt{(1-\mathrm{Tr}[\rho])(1-\mathrm{Tr}[\overline{\rho}])}\right)^2 \\
&\geq \left(\|\sqrt{\rho}\sqrt{\widetilde{\rho}}\|_1 + 1 - \mathrm{Tr}[\rho]\right)^2 \\
&= (\mathrm{Tr}[\widetilde{\rho}] + 1 - \mathrm{Tr}[\rho])^2 \\
&= \left(1 - \mathrm{Tr}[\Pi^{\perp}_{A_1^L}\rho]\right)^2 \ ,
\end{aligned}
$$

where the second equality follows from Proposition 22 and the third is the definition $\widetilde{\rho}$. Finally, using

$$
\mathrm{Tr}[\Pi^{\perp}_{A_1^L}\rho] \leq \mathrm{Tr}[\Pi^{\perp}_{A_i} \otimes \mathbb{1}_{A_2^L}\rho] \leq 1 - \sqrt{1-\varepsilon^2} \ ,
$$

we may conclude $F_*(\rho,\overline{\rho}) \geq 1-\varepsilon^2$. Thus, using the definition of purified distance, $P(\rho,\overline{\rho}) \leq \varepsilon$.

Next, we start to bound our candidate state. Note that by linearity of trace, $\overline{\rho}_{A_i} = \rho_{A_i}$ for every $i \in [L]$. Then,

$$
\begin{aligned}
&\exp\left(\min_{\sigma_B \in \mathrm{D}(B)} D_{\max}(\overline{\rho}\| \otimes_{i\in[L]} \overline{\rho}_{A_i} \otimes \sigma_B)\right) \\
=&\left\|\left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2} \overline{\rho} \left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2}\right\|_\infty \\
\leq&\Big\| \left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2} (\widetilde{\rho} + \Delta_{A_1^L} \otimes \sigma_B) \\
&\qquad \cdot \left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2} \Big\|_\infty \\
=&\Big\| \left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2} \widetilde{\rho} \left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2} \\
&\qquad + \otimes_i \rho^0_{A_i} \otimes \sigma^0_B \Big\|_\infty \\
\leq&\left\|\left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2} \widetilde{\rho} \left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2}\right\|_\infty + 1 \ ,
\end{aligned} \tag{17}
$$

where the first inequality is operator dominance and the second is the triangle inequality.

Next we want to upper bound the remaining infinity norm quantity in (17) with the exponential of $I^{\downarrow}_{\max}(:A_1^L:B)_\rho$ and a correction term. To do this, let $\log(\lambda) = I^{\downarrow}_{\max}(:A_1^L:B)_\rho$. This means $\lambda \otimes_i \overline{\sigma}_{A_i} \otimes \overline{\sigma}_B \succeq \rho$, where $(\log(\lambda), \{\overline{\sigma}_{A_i}\}, \overline{\sigma}_B)$ is the optimizer. We will use by Fact 18 and that we are considering projectors that

$$
\begin{aligned}
\exp(D_{\max}(\widetilde{\rho}\|(\otimes_{i\in[L]}\Pi_{A_i})Q(\otimes_{i\in[L]}\Pi_{A_i}))) \\
\geq \exp(D_{\max}(\widetilde{\rho}\|Q)) \ ,
\end{aligned}
$$

so it is sufficient to optimize the projected version.

For notational simplicity for any local operator Q_{A_i}, let $Q^{\Pi}_{A_i} := \Pi_{A_i} Q_{A_i} \Pi_{A_i}$. Then,

$$
\begin{aligned}
&\left\|\left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2} \widetilde{\rho} \left(\otimes_{i\in[L]}\rho_{A_i} \otimes \sigma_B\right)^{-1/2}\right\|_\infty \\
\leq& \exp(D_{\max}(\widetilde{\rho}\| \otimes_i \rho^{\Pi}_{A_i} \otimes \sigma_B)) \\
=& \left\|\left(\otimes_{i\in[L]}\rho^{\Pi}_{A_i} \otimes \sigma_B\right)^{-1/2} \widetilde{\rho} \left(\otimes_{i\in[L]}\rho^{\Pi}_{A_i} \otimes \sigma_B\right)^{-1/2}\right\|_\infty \\
=&\Big\| \left(\otimes_{i\in[L]}\rho^{\Pi}_{A_i} \otimes \sigma_B\right)^{-1/2} (\otimes_i \Pi_{A_i})\rho(\otimes_i \Pi_{A_i}) \\
&\qquad \cdot \left(\rho^{\Pi}_{A_i} \otimes \sigma_B\right)^{-1/2} \Big\|_\infty \\
\leq&\lambda \Big\| \left(\otimes_{i\in[L]}\rho^{\Pi}_{A_i} \otimes \sigma_B\right)^{-1/2} \otimes_i \Pi_{A_i}\sigma_{A_i}\Pi_{A_i} \otimes \sigma_B \\
&\qquad \cdot \left(\otimes_{i\in[L]}\rho^{\Pi}_{A_i} \otimes \sigma_B\right)^{-1/2} \Big\|_\infty \\
=&\lambda \prod_{i\in[L]} \|\Pi_{A_i}\rho^{-1/2}_{A_i}\Pi_{A_i}\sigma_{A_i}\Pi_{A_i}\rho^{-1/2}_{A_i}\Pi_{A_i}\|_\infty \\
\leq&\lambda \prod_{i\in[L]} \|\Pi_{A_i}\rho^{-1/2}_{A_i}\sigma_{A_i}\rho^{-1/2}_{A_i}\Pi_{A_i}\|_\infty \qquad (18) \\
:=&\lambda \prod_{i\in[L]} \|\Pi_{A_i}\Gamma_{A_i}\Pi_{A_i}\|_\infty \ , \qquad (19)
\end{aligned}
$$

where the first inequality is Fact 18, the second equality is expanding the definition $\widetilde{\rho}$, the second inequality is the definition of λ and $I^{\downarrow}_{\max}(:A_1^L:B)_\rho$, the third equality is multiplicativity over tensor products, the fourth inequality is the projector could have only decreased the operator norm, and the final equality is $\Gamma_{A_i} := \rho^{-1/2}_{A_i}\sigma_{A_i}\rho^{-1/2}_{A_i}$.

To upperbound $\|\Pi_{A_i}\Gamma_{A_i}\Pi_{A_i}\|_\infty$, one may choose Π_{A_i} to be the minimum rank projector on the eigenvalues of Γ_{A_i} such that $\mathrm{Tr}[(\mathbb{1}_{A_i} - \Pi_{A_i})\rho_{A_i}] \leq 1 - \sqrt{1-\varepsilon^2}$ for each i so that we still satisfy the assumptions we used to bound the purified distance. Then by the argument in [27], which uses [32, Lemma 19], for each i,

$$
\|\Pi_{A_i}\Gamma_{A_i}\Pi_{A_i}\|_\infty \leq \left(1 - \sqrt{1-\varepsilon^2}\right)^{-1} \ . \tag{20}
$$

Combining this with (17),(19) we obtain

$$
\begin{aligned}
&\min_{\sigma_B \in \mathrm{D}(B)} D_{\max}(\overline{\rho}\| \otimes_{i\in[L]} \overline{\rho}_{A_i} \otimes \sigma_B) \\
\leq& \log\left(\lambda L \left(1 - \sqrt{1-\varepsilon^2}\right)^{-1} + 1\right) \\
\leq& \log\left(\lambda\left(L\left(1 - \sqrt{1-\varepsilon^2}\right)^{-1} + \frac{1}{\mathrm{Tr}[\rho]}\right)\right) \\
=& I^{\downarrow}_{\max}(:A_1^L:B)_\rho \\
&\qquad + \log\left(L(1-\sqrt{1-\varepsilon^2})^{-1} + \mathrm{Tr}[\rho]^{-1}\right) \ ,
\end{aligned}
$$

where the second inequality uses that $\lambda \geq \mathrm{Tr}[\rho]$ as may be verified via DPI and the normalization property of $D_{\max}$.

Finally, we make the replacement $\rho \to \rho'$ such that $I^{\downarrow,\varepsilon'}_{\max}(:A_1^L:B)_\rho = I^{\downarrow}_{\max}(:A_1^L:B)_{\rho'}$ where $\varepsilon' \in (0,1)$, so that, by purified distance being a metric,

$$
\begin{aligned}
&I^{\uparrow,\varepsilon+\varepsilon'}(:A_1^L:B)_\rho \\
\leq& \min_{\sigma_B \in \mathrm{D}(B)} D_{\max}(\overline{\rho}'\| \otimes_{i\in[L]} \overline{\rho}'_{A_i} \otimes \sigma_B) \\
\leq& I^{\downarrow,\varepsilon}_{\max}(:A_1^L:B)_\rho + \log\left(L(1-\sqrt{1-\varepsilon^2})^{-1} + \mathrm{Tr}[\rho']^{-1}\right) \\
\leq& I^{\downarrow,\varepsilon}_{\max}(:A_1^L:B)_\rho + \log\left(L(1-\sqrt{1-\varepsilon^2})^{-1} + (1-\varepsilon')^{-1}\right) \ .
\end{aligned}
$$

This completes the proof. □

2) Smooth Chain Rule: We have now related the smooth up arrow and down arrow multipartite mutual information quantities. However, this is not enough to establish a strong AEP as such a property is not known for these measures.

Instead, we establish a lower bound on the smooth up arrow max mutual information quantity in terms of smooth min-entropic terms, which are known to admit a strong AEP [25].

We first state the following fact.

Proposition 23. *For any $\rho_{A_1^L B} \in \mathrm{D}(A_1^L B)$, $I_{\max}^{\varepsilon}(:A_1^L:B)_\rho$ is achieved by a state $\widetilde{\rho} \in \mathcal{B}^{\varepsilon} \cap \mathrm{D}(A_1^L B)$.*

Proof. The proof is identical to that of [27, Lemma 22]. $\square$

We will now need the following decomposition into min-entropies, which is a straightforward generalization of [13, Lemma B.10].

Lemma 24. *Let $\rho_{A_1^L B} \in \mathrm{D}_{\leq}(A_1^L B)$. Then,*

$$I_{\max}^{\uparrow}(:A_1^L:B)_\rho \geq \sum_{i\in[L]} H_{\min}(A_i)_\rho - H_{\min}(A_1^L|B)_\rho \ . \quad (21)$$

Proof. The proof is a a straightforward generalization of the proof of [13, Lemma B.10], which we provide for completeness. Let $I_{\max}^{\uparrow}(:A_1^L:B)_\rho = \log(\lambda)$. That is, for optimizer σ_B, $\lambda \otimes_{i\in[L]} \rho_{A_i} \otimes \sigma_B \succeq \rho_{A_1^L B}$. Let μ be the minimal such that

$$\mu \prod_{i\in[L]} \|\rho_{A_i}\|_\infty \mathbb{1} \otimes \sigma_B \succeq \rho_{A_1^L B} \ . \quad (22)$$

Since $\prod_{i\in[L]} \|\rho_{A_i}\|_\infty \succeq \rho_{A_1^L}$, by DPI of $D_{\max}$ and Fact 18, $\lambda \geq \mu$. Let $H_{\min}(A_1^L|B)_{\rho|\sigma} = -\log(\nu)$ which means

$$\nu \mathbb{1}_{A_1^L} \otimes \sigma_B \succeq \rho_{A_1^L B} \ . \quad (23)$$

Combining (22) and (23), we have $\mu = \nu\left(\prod_{i\in[L]} \|\rho_{A_i}\|_\infty\right)^{-1}$. Combining all of these points, we obtain

$$\begin{aligned} I_{\max}^{\uparrow}(:A_1^L:B)_\rho &= \log(\lambda) \\ &\geq \log(\mu) \\ &= \sum_i -\log(\|\rho_{A_i}\|_\infty) + \log(\nu) \\ &= \sum_{i\in[L]} H_{\min}(A_i)_\rho - H_{\min}(A_1^L|B)_\rho \ , \end{aligned}$$

where the last inequality uses the definition of $H_{\min}$. $\square$

We now convert the above result into a chain rule for smooth entropic quantities. This uses the following fact.

Proposition 25. *[12] Let $\varepsilon \in (0,1)$ and $\rho \in \mathrm{D}(A)$. Then there exists $0 \leq \Pi \leq \mathbb{1}$ such that $[\Pi, \rho] = 0$, $P(\rho, \Pi\rho\Pi) \leq 2\sqrt{\varepsilon(1-\varepsilon)}$ and*

$$H_{\min}^{\varepsilon}(A)_\rho \leq H_{\min}(A)_{\Pi\rho\Pi} \ .$$

We now convert the above into a chain rule into a smoothed version that will apply for all $\varepsilon \in (0,1)$. This is a straightforward multipartite generalization of [12, Proposition 25], which is a straightforward extension of a result in [27].

Proposition 26. *Let $\rho_{AB} \in \mathrm{D}(A_1^L B)$. Let $0 < \varepsilon < \overline{\varepsilon} < 1$ such that such that $\varepsilon + \delta(\overline{\varepsilon}) \in (0,1)$ where $\delta(\overline{\varepsilon}) := 2\sqrt{\overline{\varepsilon}(1-\overline{\varepsilon})}$. Then it holds*

$$I_{\max}^{\uparrow,\varepsilon}(A_1^L:B)_\rho \geq \sum_{i\in[L]} H_{\min}^{\overline{\varepsilon}-\varepsilon}(A_i)_\rho - H_{\min}^{\varepsilon+\delta(\overline{\varepsilon})}(A_1^L|B)_\rho \ .$$

Proof. The proof is largely identical to proofs in [12], [27], but we provide it given the relevant multipartite nuance. By re-arranging Lemma 24 and maximizing over the smoothing ball, we have

$$\begin{aligned} &H_{\min}^{\varepsilon+\delta(\overline{\varepsilon})}(A_1^L|B)_\rho \\ &\quad\geq \max_{\widetilde{\rho}\in\mathcal{B}^{\varepsilon+\delta(\overline{\varepsilon})}(\rho)} \left[\sum_i H_{\min}(A_i)_{\widetilde{\rho}} - I_{\max}(:A_1^L:B)_{\widetilde{\rho}}\right] \ . \end{aligned}$$

Then,

$$\begin{aligned} &H_{\min}^{\varepsilon+\delta(\overline{\varepsilon})}(A_1^L|B)_\rho \\ &\geq \max_{\widetilde{\rho}\in\mathcal{B}^{\varepsilon+\delta(\overline{\varepsilon})}(\rho)} \left[\sum_i H_{\min}(A_i)_{\widetilde{\rho}} - I_{\max}(A:B)_{\widetilde{\rho}}\right] \\ &\geq \max_{\omega\in\mathcal{B}^{\varepsilon}(\rho)} \left[\max_{\Pi}\left[\sum_i H_{\min}(A_i)_{\Pi\omega\Pi} - I_{\max}(:A_1^L:B)_{\Pi\omega\Pi}\right]\right] \ , \end{aligned}$$

where the new maximization is over $\Pi_{A_1^L} := \prod_{i\in[L]} \Pi_{A_i}$ where for every i, $0 \leq \Pi_{A_i} \leq \mathbb{1}_{A_i}$ such that $\Pi\omega_{A_i}\Pi \approx_{\delta(\overline{\varepsilon})} \omega_{A_i}$ and $\Pi\omega\Pi \approx_{\delta(\overline{\varepsilon})} \omega$. This is a non-empty set as one may always use $\Pi_{A_i} = \mathbb{1}_{A_i}$ for every i. This is a lower bound because we restricted smoothing to $\mathcal{B}^{\varepsilon}(\rho)$, so $\omega_{A_1^L B} \approx_{\varepsilon} \rho_{A_1^L B}$ by the DPI for purified distance, which, using purified distance is a metric, implies $\Pi\omega\Pi \approx_{\varepsilon+\delta(\overline{\varepsilon})} \rho$ as needed for this to be a restriction. Let $\omega^\star \in \mathcal{B}^{\varepsilon}(\rho) \cap \mathrm{D}(A \otimes B)$ be the optimizer of $I_{\max}^{\varepsilon}(:A_1^L:B)_\rho$ which is normalized without loss of generality (Proposition 23). Then,

$$\begin{aligned} &H_{\min}^{\varepsilon+\delta(\overline{\varepsilon})}(A|B)_\rho \\ &\geq \max_{\Pi}\left[\sum_i H_{\min}(A_i)_{\Pi\omega\Pi} - I_{\max}(:A_1^L:B)_{\Pi\omega^\star\Pi}\right] \\ &\geq \max_{\Pi}\left[\sum_i H_{\min}(A_i)_{\Pi\omega\Pi}\right] - I_{\max}(:A_1^L:B)_{\omega^\star} \ , \end{aligned}$$

where the first we have chosen $\omega^\star$ rather than maximizing and the second we have used that $D_{\max}$ actually satisfies data-processing for CPTNI maps. Then as $\omega^\star$ is normalized and we range over Π such that $\Pi_{A_i}\omega_{A_i}^\star\Pi_{A_i} \approx_{\delta(\overline{\varepsilon})} \omega_{A_i}^\star$ for every i, we can bound the min-entropic terms as

$$\begin{aligned} H_{\min}^{\varepsilon+\delta(\overline{\varepsilon})}(A|B)_\rho &\geq \sum_i H_{\min}^{\overline{\varepsilon}}(A_i)_{\omega^\star} - I_{\max}(:A_1^L:B)_{\omega^\star} \\ &\geq \sum_i H_{\min}^{\overline{\varepsilon}-\varepsilon}(A_i)_\rho - I_{\max}^{\uparrow,\varepsilon}(:A_1^L:B)_\rho \ , \end{aligned}$$

where the first inequality uses we range over Π such that $\Pi_{A_i}\omega_{A_i}^\star\Pi_{A_i} \approx_{\delta(\overline{\varepsilon})} \omega_{A_i}^\star$ and Proposition 25. The second inequality is because if $\widetilde{\rho}_{A_i} \in \mathcal{B}^{\overline{\varepsilon}-\varepsilon}(\rho_{A_i})$, as $\omega_{A_i}^\star \approx_{\varepsilon} \rho_{A_i}$ for all i, we can conclude $\widetilde{\rho} \approx_{\overline{\varepsilon}} \omega^\star$ and thus is included in the previous line's optimization. As smooth min-entropy is maximized, this suffices. Re-ordering the terms completes the proof. □

3) Putting Things Together:

Proof of Theorem 10. (*Direct Part*) This follows directly from the AEP for smooth max divergence [25] and our assumption on the function $g_\varepsilon(\cdot)$. That is,

$$\begin{aligned}&\frac{1}{n}\left[D^{\varepsilon}_{\max}(\rho^{\otimes n}\|(\rho_{A_1}\otimes\rho_{A_2}\otimes\rho_B)^{\otimes n})\right]\\ \leq& D(\rho\|\otimes_i\rho_{A_i}\otimes\rho_B)\\ =& I(:A_1^L:B)_\rho\ ,\end{aligned}$$

and adding a function that is $o(n)$ does not change this.

(*Converse Part*) This will follow from our chain rules, our assumption on the function $f_\varepsilon(\cdot)$, and the strong AEP for smooth min-entropy [25]. For any $\varepsilon\in(0,1)$, we have

$$\begin{aligned}&I^{\downarrow,\varepsilon}_{\max}(:A_1^L:B)_\rho\\ \geq& I^{\uparrow,\varepsilon}_{\max}(:A_1^L:B)_\rho-f(L,\varepsilon,0)\\ \geq&\sum_{i\in[L]}H^{\bar{\varepsilon}-\varepsilon}_{\min}(A_i)_\rho+H^{\varepsilon+\delta(\bar{\varepsilon})}_{\min}(A_1^L|B)_\rho-f(L,\varepsilon,0)\ ,\end{aligned}$$

where the first inequality is Lemma 20 with the choice $\varepsilon'=0$ and the second is Lemma 26. Thus, regularizing, we have

$$\begin{aligned}&\lim_{n\to\infty}\left[\frac{1}{n}I^{\downarrow,\varepsilon}_{\max}(:(A_1^L)^{\otimes n}:B^{\otimes n})_{\rho^{\otimes n}}\right]\\ \geq&\lim_{n\to\infty}\frac{1}{n}\Bigg[\sum_{i\in[L]}H^{\bar{\varepsilon}-\varepsilon}_{\min}(A_i^{\otimes n})_{\rho^{\otimes n}}+H^{\varepsilon+\delta(\bar{\varepsilon})}_{\min}((A_1^L)^{\otimes n}|B^{\otimes n})_{\rho^{\otimes n}}\\ &\qquad-f(L,\varepsilon,0)\Bigg]\\ =&\sum_{i\in[L]}H(A_i)_\rho+H(A_1^L|B)_\rho\\ =&I(:A_1^L:B)_\rho\ ,\end{aligned}$$

where the first equality is the strong AEP for min-entropy and the second is a straightforward chain rule using the definition of relative entropy and the well-known identity $\log(P_A\otimes Q_B)=\log(P_A)\otimes\mathbb{1}_B+\mathbb{1}_A\otimes\log(Q_B)$ for positive semidefinite P and Q [24]. □