Single-Shot Entanglement Manipulation of States and Channels Revisited

Thomas Theurer, Kun Fang, *Member, IEEE*, and Gilad Gour

***Abstract*—We study entanglement distillation and dilution of states and channels in the single-shot regime. With the help of a recently introduced conversion distance, we provide compact closed-form expressions for the dilution and distillation of pure states and show how this can be used to efficiently calculate these quantities on multiple copies of pure states. These closed-form expressions also allow us to obtain second-order asymptotics. We then prove that the ε-single-shot entanglement cost of mixed states is given exactly in terms of an expression containing a suitably smoothed version of the conditional max-entropy. For pure states, this expression reduces to the smoothed max-entropy of the reduced state, for which we provide a closed-form expression. Analogously, we provide a closed-form expression for the smoothed min-entropy and connect it to the ε-single-shot distillable entanglement. Based on these results, we bound the single-shot entanglement cost of channels. We then turn to the one-way entanglement distillation of states and channels and provide bounds in terms of a quantity we denote coherent information of entanglement.**

***Index Terms*—Quantum entanglement, single-shot entanglement distillation, single-shot entanglement cost, min-entropy, max-entropy.**

I. Introduction

QUANTUM entanglement [1], [2] plays a fundamental role in many technological applications that involve two (or more) spatially separated parties such as quantum teleportation [3], superdense coding [4], and secure quantum communication [5]. If these parties are restricted to local operations and classical communication (LOCC) [2], [6], they can manipulate and consume entanglement, but they cannot create it. Entanglement is thus a valuable quantum resource [7]. The fact that the previously mentioned protocols typically require specific entangled states, most often in the form of pure, maximally entangled states, makes the interconversion between entangled states via LOCC an important primitive. The interconversion between entangled states is traditionally studied in two different limits [2], either in the limit where one has access to (unboundedly) many identical and uncorrelated copies of a given initial state and tries to convert them to as many target states as possible or in the so-called single-shot regime: Here, one asks how well one can approximate a given target state with a single copy of an initial state and LOCC. Whilst the asymptotic regime provides ultimate bounds on the usefulness of a quantum state, the single-shot regime is closer to what is relevant from an experimental perspective, where one has only access to finitely many copies.

The maximally entangled states of dimension m play a prominent role, not only because they are required in many protocols, but also because they can be converted to all other states of lower or equal dimension [8]. The special cases of (approximate) entanglement interconversion where either the target or the initial state are maximally entangled are thus of special interest and studied under the name entanglement distillation and dilution [9], [10], [11], [12], [13], [14]. More precisely, single-shot entanglement dilution describes the task where two distant parties try to convert, up to a fixed error ε, a maximally entangled state of dimension m to a given target state via LOCC. The minimal m such that this is possible is then identified with the ε-single-shot entanglement cost of the target state. Conversely, the ε-single-shot distillable entanglement of an initial state ρ is determined by the maximal dimension of a maximally entangled state to which ρ can be converted (again up to an error ε and via LOCC). Importantly, there are many different (but topologically equivalent) ways to define the error ε, e.g., via the trace norm or the fidelity. In the following, we will consider two choices, one based on the star conversion distance recently introduced in [15], and the other based on the fidelity. After an introduction of the necessary notation in Sec. II, in Sec. III-A, we will provide closed-form expressions for the ε-single-shot distillable entanglement and the ε-single-shot entanglement cost of pure states with respect to the star conversion distance. Moreover, we show

Received 28 February 2024; revised 20 January 2025; accepted 9 May 2025. Date of publication 27 May 2025; date of current version 18 June 2025. The work of Kun Fang was supported in part by the National Natural Science Foundation of China under Grant 92470113 and Grant 12404569, in part by Shenzhen Science and Technology Program under Grant JCYJ20240813113519025, in part by Shenzhen Fundamental Research Program under Grant JCYJ20241202124023031, in part by the 1 CUHK-CUHK(SZ)-GDST Joint Collaboration Fund under Grant GRDP2025-022, and in part by the University Development Fund under Grant UDF01003565. The work of Gilad Gour was supported by Israel Science Foundation under Grant 1192/24. An earlier version of this paper was presented in part at the International Conference on Quantum Information Processing 2024 (QIP 2024). *(Corresponding author: Thomas Theurer.)*

Thomas Theurer was with the Department of Mathematics and Statistics and the Institute for Quantum Science and Technology, University of Calgary, Calgary, AB T2N 1N4, Canada. He is now with the Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark (e-mail: tth@math.ku.dk).

Kun Fang is with the School of Data Science, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China.

Gilad Gour was with the Department of Mathematics and Statistics and the Institute for Quantum Science and Technology, University of Calgary, Calgary, AB T2N 1N4, Canada. He is now with the Faculty of Mathematics, Technion—Israel Institute of Technology, Haifa 3200003, Israel.

Communicated by M. Berta, Associate Editor for Quantum.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2025.3570770.

Digital Object Identifier 10.1109/TIT.2025.3570770

that these quantities can be efficiently computed on multiple copies of a given state and provide explicit algorithms to do so. This allows us to derive analytical expressions for the second-order correction terms of the asymptotic expressions [16]. To conclude Sec. III-A, we provide formulas for the catalytic variants of ε-single-shot entanglement distillation and dilution and discuss generalizations in which we replace the maximally entangled states with arbitrary pure entangled states.

In Sec. III-B, we then move on to the ε-single-shot entanglement cost of mixed states, where we restrict the error using the square of the purified distance [17]. As one of our main results, we prove that this entanglement cost can be expressed *exactly* in terms of a smoothed version of the conditional max-entropy [18], [19], strengthening a result by Buscemi and Datta [17]. In Sec. III-C, we then show that on pure states, the two definitions of the ε-single-shot entanglement cost coincide and provide closed-form expressions for the smoothed min- and max-entropy [18], [19]. This allows us to express the ε-single-shot entanglement cost and the ε-single-shot distillable entanglement in terms of these two smoothed entropies, equipping them with an operational interpretation. We conclude the section with an comparison to previous results.

Historically, entanglement was primarily studied in the framework of so-called static resource theories [7] which focus on the value of quantum states [1], [2]. However, in typical applications in which we hope for quantum advantages, we are interested in performing a task (such as sending a secure message) that is done with the help of a quantum channel (called a dynamical resource). As demonstrated by quantum teleportation, with LOCC, we can convert static entanglement present in quantum states into channels outside of LOCC and thus indirectly quantify the value of channels via the value of states. From a conceptual point of view, it is however more natural to quantify the value of operations directly [20]. Since quantum states can be seen as a special case of quantum operations with no input and a fixed output, quantifying the value of operations is a unifying concept that can also be used to quantify properties of operations that cannot be reduced to static resources [20]. These observations have recently led to the development of dynamical resource theories [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31] and in particular dynamical resource theories of entanglement [22], [29], [30], [32], [33]. The distillation and dilution of the entanglement of channels is an important primitive in such theories for the same reasons as in the static case and has been studied under relaxations of LOCC such as complete PPT-preservation [11], [34], [35] in [22], [36], and [37] and under separability preservation in [38] and [39]. See also [31], [40], and [41] (as well as [42], which is still awaiting proof) for fundamental limitations on the distillation and dilution of channel resources and [43] for yield-cost relations in general resource theories. In Sec. III-D, we provide bounds on the ε-single-shot entanglement cost of quantum channels under LOCC which coincide in the zero-error limit.

An important subclass of LOCC is one-way LOCC in which classical communication is only allowed in one direction. In Sec. IV, we introduce the coherent information of entanglement, which is monotonic under one-way LOCC, and use it to bound the one-way ε-single shot distillable entanglement of both states and channels. We conclude with a discussion and outlook in Sec. V.

II. Notation and Preliminaries

Unless stated otherwise, proofs are provided in App. B. In this paper, we only consider finite-dimensional Hilbert spaces, which we denote with capital Latin letters such as C. The dimension of a Hilbert space C is denoted by $|C|$, and the set of density matrices acting on it by $\mathfrak{D}(C)$, with $\mathrm{Pure}(C)$ denoting the subset of pure states. Density matrices are represented by small Greek letters such as ρ^C, where the superscript indicates that ρ acts on C. For, e.g., a state $\rho^{AB} \in \mathfrak{D}(AB)$ we will also use the convention that $\rho^A = \mathrm{Tr}_B\left[\rho^{AB}\right]$ denotes the marginal on system A. Whenever we consider a copy of a Hilbert space A, we will denote it as $\tilde{A}$ and we reserve X to denote a classical system or register.

In the following, we will mainly be concerned with two spatially separated parties, Alice and Bob. To make clear which system is under the control of whom, we will use A and A' to denote Alice's systems and B and B' for Bob's. Quantum channels will be denoted by calligraphic large Latin letters such as $\mathcal{N}$ (with the exemption of the identity channel, which we denote by id) and the set of all quantum channels from A to B by $\mathrm{CPTP}(A \to B)$, which stands for completely positive and trace-preserving. Completely positive linear maps will be denoted by CP and a collection of CP maps $\{\mathcal{N}_x\}_{x=1}^n$ for which $\sum_{x=1}^n \mathcal{N}_x$ is a quantum channel will be called an instrument. Since we can always store the classical outcome x of any instrument in a classical system X, we will interchangeably also write $\sum_x \mathcal{N}_x \otimes |x\rangle\langle x|^X$ for the instrument.

The set of quantum channels that is implementable with local operations and classical communication will be denoted by LOCC. Since with LOCC, Alice and Bob can always attach and remove local auxiliary systems, for bipartite systems AB, we will assume in the following w.l.o.g. that $|A| = |B|$. Whenever there exists an $\mathcal{N} \in \mathrm{LOCC}$ such that $\mathcal{N}(\rho) = \sigma$, we write $\rho \xrightarrow{\mathrm{LOCC}} \sigma$. The set of channels that can be implemented with local operations and one-way classical communication *from Alice to Bob* will be denoted by LOCC_1. Moreover, we will denote superchannels [44], i.e., linear maps between quantum channels that can be implemented by concatenating the channel they act on with two other channels implementing a pre- and post-processing, with capital Greek letters such as Θ. We will be particularly concerned with LOCC superchannels, i.e., superchannels that can be implemented by a pre- and post-processing that are both in LOCC, see Fig. 1.

We utilize bold small Latin letters such as $\mathbf{p}$ for probability vectors, with p_x the x-th component of $\mathbf{p}$, and let $\mathrm{Prob}(d)$ be the set of probability vectors of length d. The subset that contains the d-dimensional probability vectors with non-increasing entries will be denoted as $\mathrm{Prob}^{\downarrow}(d)$ and for $\mathbf{p} \in \mathrm{Prob}(d)$, $\mathbf{p}^{\downarrow} \in \mathrm{Prob}^{\downarrow}(d)$ represents the vector that we obtain by reordering the elements of $\mathbf{p}$. For $k \in [d]$, where $[d]$ is a short-hand notation for $\{1, \ldots, d\}$, the k-th Ky-Fan norm of

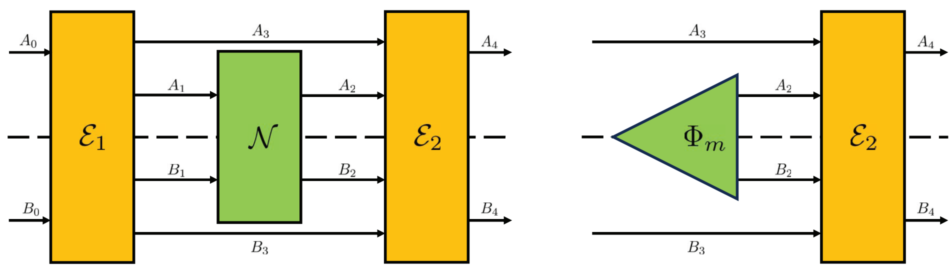

Fig. 1. Left: LOCC superchannel converting a channel $\mathcal{N}$ to a channel $\mathcal{M} = \mathcal{E}_2 \circ \mathcal{N} \circ \mathcal{E}_1$, where both $\mathcal{E}_1$ and $\mathcal{E}_2$ are in LOCC. Right: If $\mathcal{N}$ is replaced by a state, the superchannel simplifies. Solid lines represent quantum systems and the dashed line the spacial separation between Alice and Bob.

$\mathbf{p} \in \mathrm{Prob}(d)$ is defined as

$$\|\mathbf{p}\|_{(k)} = \sum_{x=1}^{k} p_x^{\downarrow}. \tag{1}$$

This definition can be extended to positive semidefinite operators A by denoting with $\|A\|_{(k)}$ the sum of the k largest eigenvalues of A (i.e., the Ky-Fan norm of its eigenvalues). For $\mathbf{p}, \mathbf{q} \in \mathrm{Prob}(d)$, we write $\mathbf{p} \succ \mathbf{q}$ if $\mathbf{p}$ majorizes $\mathbf{q}$ [45], i.e., if

$$\|\mathbf{p}\|_{(k)} \geq \|\mathbf{q}\|_{(k)} \; \forall k \in [d]. \tag{2}$$

If $\mathbf{p}, \mathbf{q}$ have different dimensions, we extend the definition by padding the shorter vector with zeros. By fixing an arbitrary orthonormal basis $|x\rangle^C$ for every Hilbert space C, any $\psi \in \mathrm{Pure}(AB)$ is then LOCC-equivalent to its standard form $\sum_x \sqrt{p_x}|xx\rangle^{AB}$, where $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|)$ contains the Schmidt coefficients of ψ, which, w.l.o.g., we always assume to be ordered non-increasingly. The number of non-zero Schmidt coefficients of ψ, i.e., its Schmidt rank, is denoted by $\mathrm{SR}(\psi)$ and $\Phi_m = \frac{1}{\sqrt{m}} \sum_{x=1}^{m} |xx\rangle^{AB}$ represents the maximally entangled state of dimension m, where w.l.o.g., we will always assume that $m = |A| = |B|$.

We use $H(\rho) := -\mathrm{Tr}(\rho \log \rho)$ to denote the von-Neumann entropy, and for $\rho \in \mathfrak{D}(ABE)$, let $H(A|B)_\rho := H(\rho^{AB}) - H(\rho^B)$ be the conditional von-Neumann entropy. For $\rho, \sigma \in \mathfrak{D}(C)$ and $\|\cdot\|_1$ the trace norm, the trace distance between ρ and σ is defined as $\frac{1}{2}\|\rho - \sigma\|_1$ and the relative entropy as $D(\rho\|\sigma) := \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma)$. The (trace) conversion distance under LOCC is commonly defined as

$$\begin{aligned} &T\left(\rho^{AB} \xrightarrow{\text{LOCC}} \sigma^{A'B'}\right) \\ &:= \min_{\tau \in \mathfrak{D}(A'B')} \left\{ \frac{1}{2}\left\|\tau^{A'B'} - \sigma^{A'B'}\right\|_1 : \rho^{AB} \xrightarrow{\text{LOCC}} \tau^{A'B'} \right\}. \end{aligned} \tag{3}$$

On pure states $\psi, \phi \in \mathrm{Pure}(AB)$, one can analogously define the star conversion distance [15] (see also [16]) as

$$\begin{aligned} &T_\star\left(\psi^{AB} \xrightarrow{\text{LOCC}} \phi^{AB}\right) \\ &:= \min_{\mathbf{r} \in \mathrm{Prob}(|A|)} \left\{ \frac{1}{2}\|\mathbf{r} - \mathbf{q}\|_1 \;:\; \mathbf{r} \succ \mathbf{p} \right\}, \end{aligned} \tag{4}$$

where $\mathbf{p}, \mathbf{q} \in \mathrm{Prob}^{\downarrow}(|A|)$ are the Schmidt coefficients of ψ and ϕ, respectively, and $\|\mathbf{p} - \mathbf{q}\|_1 = \sum_x |p_x - q_x|$. This definition can easily be extended to pure bipartite states that do not belong to the same Hilbert space [15]: Using LOCC, one can always add and remove separable auxiliary states such that the Hilbert spaces match. For $\psi \in \mathrm{Pure}(AB)$, $\phi \in \mathrm{Pure}(A'B')$, and $d = |A| = |B|$, $d' = |A'| = |B'|$, we thus define

$$\begin{aligned} &T_\star\left(\psi^{AB} \xrightarrow{\text{LOCC}} \phi^{A'B'}\right) \\ &:= T_\star\left(\psi^{AB} \otimes |11\rangle\langle 11|^{A'B'} \xrightarrow{\text{LOCC}} |11\rangle\langle 11|^{AB} \otimes \phi^{A'B'}\right). \end{aligned} \tag{5}$$

Importantly, the conversion distances discussed above are topologically equivalent in the sense that [15, Lem. 3]

$$\begin{aligned} &\frac{1}{2} T_\star^2\left(\psi^{AB} \xrightarrow{\text{LOCC}} \phi^{A'B'}\right) \leq T\left(\psi^{AB} \xrightarrow{\text{LOCC}} \phi^{A'B'}\right) \\ &\leq \sqrt{2 T_\star\left(\psi^{AB} \xrightarrow{\text{LOCC}} \phi^{A'B'}\right)} \end{aligned} \tag{6}$$

and according to [15, Thm. 4], $T_\star$ exhibits the following closed-form expression that is derived via the concept of approximate majorization [46]:

Theorem 1: Let $\psi \in \mathrm{Pure}(AB), \phi \in \mathrm{Pure}(A'B')$, and $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|), \mathbf{q} \in \mathrm{Prob}^{\downarrow}(|A'|)$ be their corresponding Schmidt coefficients. Then,

$$T_\star\left(\psi^{AB} \xrightarrow{\text{LOCC}} \phi^{A'B'}\right) = \max_{k \in [\mathrm{SR}(\psi^{AB})]} \left\{ \|\mathbf{p}\|_{(k)} - \|\mathbf{q}\|_{(k)} \right\}.$$

The purified distance between two states $\rho, \sigma \in \mathfrak{D}(AB)$ is defined as [47], [48], [49]

$$P(\rho, \sigma) := \sqrt{1 - F^2(\rho, \sigma)}, \tag{7}$$

where $F(\rho, \sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1$ is the fidelity. For $\mathbf{p}, \mathbf{q} \in \mathrm{Prob}^{\downarrow}(|A|)$, $P(\mathbf{r}, \mathbf{q}) := \sqrt{1 - F^2(\mathbf{r}, \mathbf{q})}$ is the classical version of the purified distance, and $F(\mathbf{r}, \mathbf{q}) := \sum_x \sqrt{r_x q_x}$. Analogously to what was done above with the trace distance, one can define a conversion distance based on the purified distance: For $\rho \in \mathfrak{D}(AB), \sigma \in \mathfrak{D}(A'B')$, the purified conversion distance is defined as

$$P\left(\rho \xrightarrow{\text{LOCC}} \sigma\right) := \min_{\tau \in \mathfrak{D}(A'B')} \left\{ P(\tau, \sigma) : \rho \xrightarrow{\text{LOCC}} \tau \right\}. \tag{8}$$

III. Single-Shot Entanglement Manipulation With LOCC

A. Pure State Entanglement Manipulation and the Star Conversion Distance

The star conversion distance can be used to define and calculate an ε-single-shot distillable entanglement, which we will do in the following. For any $\varepsilon \in [0,1]$ and $\psi \in \mathrm{Pure}(AB)$, the ε-single-shot distillable entanglement is defined as

$$\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) := \max_{m\in\mathbb{N}}\left\{\log m \;:\; T_\star\left(\psi^{AB} \xrightarrow{\text{LOCC}} \Phi_m\right) \le \varepsilon\right\}. \tag{9}$$

From the closed formula for $T_\star$, we get the following result.

Theorem 2: Let $\varepsilon \in [0,1)$, $\psi \in \mathrm{Pure}(AB)$, $d := \mathrm{SR}(\psi^{AB})$, and $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|)$ be the Schmidt coefficients of ψ^{AB}. The ε-single-shot distillable entanglement of ψ^{AB} is then given by

$$\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) = \min_{k\in\{\ell,\dots,d\}} \log\left\lfloor \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon}\right\rfloor, \tag{10}$$

where $\ell \in [d]$ is the integer satisfying $\|\mathbf{p}\|_{(\ell-1)} \le \varepsilon < \|\mathbf{p}\|_{(\ell)}$.

For any $\varepsilon \in [0,1]$ and $\psi \in \mathrm{Pure}(AB)$, the ε-single-shot entanglement cost is defined as

$$\mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right) := \min_{m\in\mathbb{N}}\left\{\log m \;:\; T_\star\left(\Phi_m \xrightarrow{\text{LOCC}} \psi^{AB}\right) \le \varepsilon\right\}. \tag{11}$$

From the closed formula for $T_\star$, we get the following result.

Theorem 3: Let $\varepsilon \in [0,1)$, $\psi \in \mathrm{Pure}(AB)$, $d := \mathrm{SR}(\psi^{AB})$, and $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|)$ be the Schmidt coefficients of ψ^{AB}. The ε-single-shot entanglement cost of ψ^{AB} is then given by

$$\mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right) = \log m, \tag{12}$$

where $m \in [d]$ is the integer satisfying $\|\mathbf{p}\|_{(m-1)} < 1-\varepsilon \le \|\mathbf{p}\|_{(m)}$.

As explained in detail in App. C, the above Theorems allow us to efficiently compute both the ε-single-shot distillable entanglement and the ε-single-shot entanglement cost for multiple copies of a given pure state.

Theorem 4: Let $n \in \mathbb{N}$, $\varepsilon \in [0,1)$, and $\psi \in \mathrm{Pure}(AB)$. This implies that both $\mathrm{Distill}^{\varepsilon}\left(\psi^{\otimes n}\right)$ and $\mathrm{Cost}^{\varepsilon}\left(\psi^{\otimes n}\right)$ can be computed efficiently.

The main idea behind the proof is to use that whilst $\mathbf{p}^{\otimes n}$ has a number of entries that is exponential in n, it only has a polynomial number of distinct entries, which allows to determine $\|\mathbf{p}^{\otimes n}\|_{(k)}$ and therefore $\mathrm{Distill}^{\varepsilon}\left(\psi^{\otimes n}\right)$ and $\mathrm{Cost}^{\varepsilon}\left(\psi^{\otimes n}\right)$. In App. C, we provide explicit algorithms to determine these quantities.

Moreover, Thms. 2 and 3 also allow us to obtain the second-order asymptotics of the entanglement cost and the distillable entanglement: Let the Gaussian cumulative distribution function with mean value μ and variance ν be denoted by

$$\Phi_{\mu,\nu}(x) := \frac{1}{\sqrt{2\pi\nu}}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\nu}}\,\mathrm{d}t \tag{13}$$

and the entropy variance $V(\mathbf{p})$ of a probability distribution $\mathbf{p}$ by

$$V(\mathbf{p}) := \sum_i p_i(-\log p_i - H(\mathbf{p}))^2, \tag{14}$$

where $H(\mathbf{p}) := -\sum_i p_i \log p_i$ is the Shannon entropy. As a shorthand notation, we will also use Φ to denote $\Phi_{0,1}$. We begin by stating the following Lemma, which is a consequence of [16, Lem. 15] (see also [50, Lem. 16]).

Lemma 1: For any distribution $\mathbf{p}$ such that $V(\mathbf{p}) > 0$, any natural number n, and $\varepsilon \in [0,1)$, let

$$f_{n,\varepsilon}(\mathbf{p}) := \min\left\{k : \|\mathbf{p}^{\otimes n}\|_{(k)} > \varepsilon\right\},$$

$$f'_{n,\varepsilon}(\mathbf{p}) := \min\left\{k : \|\mathbf{p}^{\otimes n}\|_{(k)} \ge \varepsilon\right\}. \tag{15}$$

Then we have that

$$\begin{aligned}\lim_{n\to\infty}\frac{\log f'_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}} &= \lim_{n\to\infty}\frac{\log f_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}} \\ &= \Phi^{-1}(\varepsilon).\end{aligned} \tag{16}$$

By combining this Lemma with Thm. 3, one immediately obtains the following Proposition.

Proposition 1: For any pure state $\psi \in \mathrm{Pure}(AB)$ with Schmidt vector $\mathbf{p}$, $V(\mathbf{p}) > 0$, and $\varepsilon \in [0,1)$, it holds that

$$\mathrm{Cost}^{\varepsilon}(\psi^{\otimes n}) = nH(\mathbf{p}) - \Phi^{-1}(\varepsilon)\sqrt{nV(\mathbf{p})} + o(\sqrt{n}). \tag{17}$$

With slightly more work, it is also possible to obtain the second-order asymptotics of the distillable entanglement.

Proposition 2: For any pure state $\psi \in \mathrm{Pure}(AB)$ with Schmidt vector $\mathbf{p}$, $V(\mathbf{p}) > 0$, and $\varepsilon \in [0,1)$, it holds that

$$\mathrm{Distill}^{\varepsilon}(\psi^{\otimes n}) = nH(\mathbf{p}) + \Phi^{-1}(\varepsilon)\sqrt{nV(\mathbf{p})} + o(\sqrt{n}). \tag{18}$$

The above two Propositions are the analogues of [16, Eqs. (136) and (137)], where the error ε was measured with P instead of $T_\star$. Since the considered distance measures are topologically equivalent, one immediately recovers the asymptotic entanglement cost and distillable entanglement of pure entangled states [2].

Recently, the phenomenon of entanglement catalysis [51] has seen renewed interest, see for example the review article [52]. In particular, the authors of [53] showed that for $\psi, \phi \in \mathrm{Pure}(AB)$ with Schmidt coefficients $\mathbf{p}$ and $\mathbf{q}$, respectively, ψ can be converted to ϕ via an catalytic LOCC transformation if and only if

$$H(\mathbf{p}) \ge H(\mathbf{q}). \tag{19}$$

In [53], the authors consider a sequence of catalyst states that remain unchanged in the process but are allowed to build an asymptotically vanishing amount of correlations with the target state. Note that this allows for the compact condition above. If no error is allowed at any stage of the protocol, the allowed conversions are characterized by a continuous family of inequalities that can be expressed in terms of the Rényi divergences [54], [55]. In the following, we consider the framework of [53].

Motivated by the above discussion, we define the star conversion distance under catalytic LOCC (CLOCC) as

$$T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \phi^{AB}\right) := \min_{\mathbf{r}\in\mathrm{Prob}(|A|)}\left\{\frac{1}{2}\|\mathbf{r}-\mathbf{q}\|_1 \;:\; H(\mathbf{p}) \ge H(\mathbf{r})\right\}. \tag{20}$$

Denote by $\bar{\mathbf{q}}^{(\varepsilon)}$ the steepest ε-approximation of $\mathbf{q}$ [46], i.e., a state $\bar{\mathbf{q}}^{(\varepsilon)} \in \mathfrak{B}_\varepsilon(\mathbf{q}) := \left\{\mathbf{q}' \in \mathrm{Prob}(|A|) : \frac{1}{2}\|\mathbf{q}'-\mathbf{q}\|_1 \le \varepsilon\right\}$ which satisfies that $\bar{\mathbf{q}}^{(\varepsilon)} \succ \mathbf{q}'$ for all $\mathbf{q}' \in \mathfrak{B}_\varepsilon(\mathbf{q})$. According to [46], such a state exists and can be constructed explicitly.

Lemma 2: Let $\psi, \phi \in \text{Pure}(AB)$ with Schmidt coefficients $\mathbf{p}$ and $\mathbf{q}$, respectively. Then

$$T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \phi^{AB}\right) = \begin{cases} 0 \text{if } H(\mathbf{q}) \leq H(\mathbf{p}), \\ \varepsilon : H(\bar{\mathbf{q}}^{(\varepsilon)}) = H(\mathbf{p}) \text{else}. \end{cases}$$

Analogously to the case without catalysis, we define the ε-single-shot catalytic entanglement cost as

$$\begin{aligned} &\text{Cost}_c^\varepsilon\left(\psi^{AB}\right) \\ &:= \min_{m\in\mathbb{N}} \left\{ \log m \ : \ T_\star\left(\Phi_m \xrightarrow{\text{CLOCC}} \psi^{AB}\right) \leq \varepsilon \right\}, \end{aligned} \tag{21}$$

and provide a closed-form expression.

Theorem 5: Denote by $\mathbf{p}$ the vector containing the Schmidt coefficients of $\psi \in \text{Pure}(AB)$. Then

$$\text{Cost}_c^\varepsilon\left(\psi^{AB}\right) = \log \left\lceil 2^{H^\varepsilon(\mathbf{p})} \right\rceil, \tag{22}$$

where

$$H^\varepsilon(\mathbf{p}) = \min_{\mathbf{r}\in\mathfrak{B}_\varepsilon(\mathbf{p})} H(\mathbf{r}) = H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right) \tag{23}$$

is the ε-smoothed Shannon entropy and $\bar{\mathbf{p}}^{(\varepsilon)}$ denotes the steepest ε-approximation of $\mathbf{p}$ [46].

Next, we define the ε-single-shot catalytic distillable entanglement

$$\begin{aligned} &\text{Distill}_c^\varepsilon\left(\psi^{AB}\right) \\ &:= \max_{m\in\mathbb{N}} \left\{ \log m \ : \ T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \Phi_m\right) \leq \varepsilon \right\}, \end{aligned} \tag{24}$$

and show that one can evaluate it via an optimization over a finite set.

Theorem 6: Denote by $\mathbf{p}$ the vector containing the Schmidt coefficients of $\psi \in \text{Pure}(AB)$. For $\varepsilon \in [0, 1)$,

$$\text{Distill}_c^\varepsilon\left(\psi^{AB}\right) = \max_{m\in\mathbb{N}} \left\{ \log m : H^\varepsilon\left(\mathbf{u}_m\right) \leq H\left(\mathbf{p}\right) \right\},$$

and the optimization over m can be restricted to

$$\left\lfloor 2^{H(\mathbf{p})} \right\rfloor \leq m \leq \left\lfloor 2^{\frac{H(\mathbf{p})+h(\varepsilon)}{1-\varepsilon}} \right\rfloor, \tag{25}$$

where $h(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy.

To conclude this section, we want to quickly mention a generalization of entanglement distillation and dilution where one replaces the maximally entangled states with an arbitrary pure resource state. For any $\varepsilon \in [0, 1]$ and $\psi \in \text{Pure}(AB), \phi \in \text{Pure}(A'B')$ entangled, we define

$$\begin{aligned} &\text{Cost}^\varepsilon\left(\psi^{AB}\middle|\phi^{A'B'}\right) \\ &:= \min \left\{ m \in \mathbb{N}_0 \ : \ T_\star\left(\phi^{\otimes m} \xrightarrow{\text{LOCC}} \psi\right) \leq \varepsilon \right\} \end{aligned} \tag{26}$$

and

$$\begin{aligned} &\text{Distill}^\varepsilon\left(\psi^{AB}\middle|\phi^{A'B'}\right) \\ &:= \max \left\{ m \in \mathbb{N}_0 \ : \ T_\star\left(\psi \xrightarrow{\text{LOCC}} \phi^{\otimes m}\right) \leq \varepsilon \right\}, \end{aligned} \tag{27}$$

where we understand that $\phi^{\otimes 0}$ corresponds to a separable state.

Thanks to Thm. 1 and Algorithms 1 and 2, which provide an efficient way to evaluate $\left\|\mathbf{p}^{\otimes m}\right\|_{(k)}$, both of the above quantities can be evaluated exactly. To compute $\text{Cost}^\varepsilon(\psi|\phi)$,

Algorithm 1 Efficient Evaluation of Ky-Fan Norms

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \text{Prob}^\downarrow(d)$, integers $n, m \geq 1$
Output: $\|\mathbf{p}^{\otimes n}\|_{(m)}$

Compute the $r := \binom{n+d-1}{d-1}$ different terms $p_1^{n_1} p_2^{n_2} \cdots p_d^{n_d}$ where $n_1 + n_2 + \cdots + n_d = n$;
Sort the r terms in non-increasing order resulting in the vector $(s_1, s_2, \cdots, s_r)$. Let v_i be the number of times that s_i is repeated;
Let $N := 0$, $P := 0$;
foreach $k \in [r]$ **do**
  Let $N \leftarrow N + v_k$, and $P \leftarrow P + v_k s_k$;
  if $m < N$ **then**
    return $s_k(m - N) + P$;
  end
end

Algorithm 2 Efficient Evaluation of Ky-Fan Norms via Binary Search

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \text{Prob}^\downarrow(d)$, integers $n, m \geq 1$
Output: $\|\mathbf{p}^{\otimes n}\|_{(m)}$

Compute the $r := \binom{n+d-1}{d-1}$ different terms $p_1^{n_1} p_2^{n_2} \cdots p_d^{n_d}$ where $n_1 + n_2 + \cdots + n_d = n$;
Sort the r terms in non-increasing order resulting in the vector $(s_1, s_2, \cdots, s_r)$. Let v_i be the number of times that s_i is repeated;
Let $N_0 := 0$, $P_0 := 0$, and $\forall k \in [r]$, $N_k := \sum_{i=1}^k v_i$, and $P_k := \sum_{i=1}^k v_i s_i$;
Let $a = 0$, $b = r$;
while *True* **do**
  Let $c = \lfloor (a+b)/2 \rfloor$;
  if $N_c \leq m \leq N_{c+1}$ **then**
    return $s_{c+1}(m - N_c) + P_c$;
  else if $m < N_c$ **then**
    $b \leftarrow c$;
  else
    $a \leftarrow c + 1$;
  end
end

one can use the following algorithm: Let $m = 0$ and check if $T_\star\left(\phi^{\otimes m} \xrightarrow{\text{LOCC}} \psi\right) \leq \varepsilon$ (via Thm. 1 and Algorithms 1 or 2). If yes, return m, if no, increase m by one and repeat. The algorithm will terminate eventually: Let $\mathbf{p}, \mathbf{q}$ be the Schmidt coefficients of ψ, ϕ. A (crude) upper bound is for example

$$\text{Cost}^\varepsilon\left(\psi^{AB}\middle|\phi^{A'B'}\right) \leq \left\lceil \frac{\log\left(\frac{p_1}{\text{SR}(\psi)}\right)}{\log(q_1)} \right\rceil. \tag{28}$$

This can be seen as follows: First, notice that

$$T_\star\left(\phi^{\otimes m} \xrightarrow{\text{LOCC}} \psi\right) = \max_{k\in[\text{SR}(\phi^{\otimes m})]} \left\{ ||\mathbf{q}^{\otimes m}||_{(k)} - ||\mathbf{p}||_{(k)} \right\}.$$

Since for $k > \text{SR}(\psi)$

$$\left\{ ||\mathbf{q}^{\otimes m}||_{(k)} - ||\mathbf{p}||_{(k)} \right\} \leq 0 \tag{29}$$

and for $k \leq \mathrm{SR}(\psi)$ and $m > \left\lceil \log\left(\frac{p_1}{\mathrm{SR}(\psi)}\right) / \log(q_1) \right\rceil$,

$$\left\| \mathbf{q}^{\otimes m} \right\|_{(k)} \leq \mathrm{SR}(\psi) q_1^m < p_1 \leq \|\mathbf{p}\|_{(k)} , \tag{30}$$

$T_\star\left(\phi^{\otimes m} \xrightarrow{\text{LOCC}} \psi\right)$ will be zero for such m. This bound (which is independent of ε) thus provides a bound on the maximal number of copies of ϕ that is needed in order to *exactly* create a copy of ψ via LOCC.

To compute $\mathrm{Distill}^\varepsilon(\psi|\phi)$, set $m = 1$ and check if $T_\star\left(\psi \xrightarrow{\text{LOCC}} \phi^{\otimes m}\right) > \varepsilon$. If yes, return $m-1$, if no, increase m by one and repeat. The algorithm will terminate eventually, too: Let $\mathbf{p}, \mathbf{q}$ again be the Schmidt coefficients of ψ, ϕ and thus

$$\begin{aligned} T_\star\left(\psi \xrightarrow{\text{LOCC}} \phi^{\otimes m}\right) &= \max_{k \in [\mathrm{SR}(\psi)]} \left\{ \|\mathbf{p}\|_{(k)} - \|\mathbf{q}^{\otimes m}\|_{(k)} \right\} \\ &\geq \|\mathbf{p}\|_{(\mathrm{SR}(\psi))} - \|\mathbf{q}^{\otimes m}\|_{(\mathrm{SR}(\psi))} \\ &\geq 1 - \mathrm{SR}(\psi) q_1^m. \end{aligned} \tag{31}$$

This implies that

$$\mathrm{Distill}^\varepsilon\left(\psi^{AB} \middle| \phi^{A'B'}\right) \leq \left\lfloor \frac{\log\left(\frac{1-\varepsilon}{\mathrm{SR}(\psi)}\right)}{\log(q_1)} \right\rfloor. \tag{32}$$

B. Single-Shot Entanglement Cost of Mixed States

According to [17, Def. 1], for any $\varepsilon \in [0,1]$, the ε-single-shot entanglement cost is defined as

$$\begin{aligned} &\mathrm{Cost}^\varepsilon\left(\rho^{AB}\right) \\ &:= \min\left\{ \log m \ : \ P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \rho^{AB}\right) \leq \varepsilon \right\}. \end{aligned} \tag{33}$$

Note that at first glance, for pure states, this definition conflicts with Eq. (11). However, as we will show later, the two definitions coincide. In a bit of abuse of notation, we will thus not differentiate between the two, e.g., by adding a subscript to denote the distance according to which we determine the allowed error in the transformation.

To calculate the ε-single-shot entanglement cost, we will rely on a convenient characterization of the purified conversion distance: On bipartite states $\psi \in \mathrm{Pure}(AB)$, and for any $k \in [|A|]$, let

$$E_{(k)}(\psi^{AB}) := 1 - \|\mathbf{p}\|_{(k)} , \tag{34}$$

where $\mathbf{p}$ contains the Schmidt coefficients of ψ^{AB}. Applying a convex-roof extension, for mixed states $\rho \in \mathfrak{D}(AB)$, define

$$\begin{aligned} E_{(k)}\left(\rho^{AB}\right) &: = \inf \sum_x p_x E_{(k)}\left(\psi_x^{AB}\right) \\ &= \inf\left(1 - \sum_x p_x \left\| \mathrm{Tr}_A\left[\psi_x^{AB}\right] \right\|_{(k)} \right), \end{aligned} \tag{35}$$

where the infimums are over all pure-state decompositions $\rho^{AB} = \sum_x p_x \psi_x^{AB}$ [17], [56], [57], [58]. These quantities have an operational interpretation in the sense that for $\psi, \phi \in \mathrm{Pure}(AB)$, $\psi^{AB} \xrightarrow{\text{LOCC}} \phi^{AB}$ if and only if [8], [59],

$$E_{(k)}(\psi^{AB}) \geq E_{(k)}(\phi^{AB}) \ \ \forall\, k \in [|A|] . \tag{36}$$

Moreover, according to [57, Thm. 8.8 and Sec. B4] and [60, Lem. 5], for $\rho \in \mathfrak{D}(AB)$ and $m \in \mathbb{N}$,

$$P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \rho^{AB}\right) = E_{(m)}\left(\rho^{AB}\right), \tag{37}$$

where $E_{(m)}\left(\rho^{AB}\right)$ is defined in Eq. (35) with $k = m$ (see also [61, Eq. (48)] for the case of pure states).

To provide formulas for the ε-single-shot entanglement cost, it is convenient to utilize the concept of *classical extensions* of a bipartite density matrix $\rho \in \mathfrak{D}(AB)$: With any decomposition $\{p_x, \rho_x^{AB}\}_{x\in[k]}$ of ρ^{AB}, i.e., $\rho^{AB} = \sum_{x\in[k]} p_x \rho_x^{AB}$, one associates a classical-quantum-state (see [17, Eq. (3)])

$$\rho^{XAB} := \sum_{x\in[k]} p_x |x\rangle\langle x|^X \otimes \rho_x^{AB}, \tag{38}$$

where X is a classical system of dimension k. Such an extension will be called a *regular extension* of ρ^{AB} if all ρ_x^{AB} are *pure states*.

We further denote by $H_{\max}$ the conditional max-entropy, i.e.,

$$H_{\max}(A|B)_\rho := \max_{\tau \in \mathfrak{D}(B)} \log \mathrm{Tr}\left[\Pi_\rho^{AB}\left(I^A \otimes \tau^B\right)\right], \tag{39}$$

where Π_ρ^{AB} is the projection onto the support of $\rho^{AB} = \mathrm{Tr}_C\left[\rho^{ABC}\right]$, and by $H_{\max}^\varepsilon$ its smoothed version defined as

$$H_{\max}^\varepsilon(A|B)_\rho = \min_{\omega \in \mathfrak{B}_\varepsilon(\rho^{AB})} H_{\max}(A|B)_\omega, \tag{40}$$

where

$$\mathfrak{B}_\varepsilon\left(\rho^A\right) = \left\{\tau \in \mathfrak{D}(A) : \frac{1}{2}\left\|\tau^A - \rho^A\right\|_1 \leq \varepsilon\right\}. \tag{41}$$

For a classical extension

$$\rho^{XAB} = \sum_{x\in[k]} p_x |x\rangle\langle x|^X \otimes \rho_x^{AB} \tag{42}$$

of a bipartite state $\rho \in \mathfrak{D}(AB)$, this implies that

$$H_{\max}(A|X)_\rho = \max_{x\in[k]: p_x \neq 0} \log \mathrm{Tr}\left[\Pi_{\rho_x}^A\right], \tag{43}$$

where $\rho_x^A := \mathrm{Tr}_B\left[\rho_x^{AB}\right]$, and

$$H_{\max}^\varepsilon(A|X)_\rho = \min_{\omega \in \mathfrak{B}_\varepsilon(\rho^{XA})} \max_{x\in[k]: q_x \neq 0} \log \mathrm{Tr}\left[\Pi_{\omega_x}^A\right] , \tag{44}$$

with

$$\omega^{XA} = \sum_{x\in[k]} q_x |x\rangle\langle x|^X \otimes \omega_x^A. \tag{45}$$

It was pointed out in [17] that

$$\begin{aligned} &\mathrm{Cost}^\varepsilon\left(\rho^{AB}\right) \\ &= \inf_{\omega^{XAB}}\left\{H_{\max}(A|X)_\omega : \ P^2\left(\omega^{AB}, \rho^{AB}\right) \leq \varepsilon\right\}, \end{aligned} \tag{46}$$

where the infimum is over all regular extensions of ρ^{AB}.

This means that the ε-single-shot entanglement cost can be expressed in terms of a smoothed version of the conditional max-entropy. As the main result of [17], it was further shown in their Thm. 1 that the ε-single-shot entanglement cost can be lower and upper bounded by a slight modification of $H_{\max}^\varepsilon$ (take Eq. (40), but relax the requirement in Eq. (41) that τ is a density operator to the requirement that it is positive semidefinite). In the following, we will show that (with our slight change in the smoothing), the ε-single-shot entanglement cost can be expressed *exactly* in terms of the conditional max-entropy. To this end, we introduce some notation and a Lemma first.

For every classical-quantum-state

$$\rho^{XA} = \sum_{x\in[k]} p_x |x\rangle\langle x|^X \otimes \rho_x^A, \tag{47}$$

define

$$\rho^{(m)} = \sum_{x\in[k]} p_x |x\rangle\langle x|^X \otimes \rho_x^{(m)}, \tag{48}$$

with $\sigma^{(m)}$ an m-pruned version of σ^A defined as

$$\sigma^{(m)} := \frac{\Pi_m^A \sigma^A \Pi_m^A}{\mathrm{Tr}\left[\sigma^A \Pi_m^A\right]}, \tag{49}$$

where Π_m^A is a projection to a subspace spanned by m orthogonal eigenvectors corresponding to the m largest eigenvalues of σ^A.[1]

Lemma 3: Let $\rho \in \mathfrak{D}(XA)$ be a classical-quantum-state as in Eq. (47), and for any $m \in [|A|]$, let $\rho^{(m)}$ be as defined in Eq. (48). Then, for any $\varepsilon \in [0,1]$,

$$\begin{aligned} &H_{\max}^{\varepsilon}(A|X)_\rho \\ &= \min_{m\in[|A|]} \left\{\log m : \frac{1}{2}\left\|\rho^{(m)} - \rho^{XA}\right\|_1 \le \varepsilon\right\} \\ &= \min_{m\in[|A|]} \left\{\log m : \sum_{x\in[k]} p_x \left\|\rho_x^A\right\|_{(m)} \ge 1-\varepsilon\right\}. \end{aligned} \tag{50}$$

This Lemma shows that $H_{\max}^{\varepsilon}(A|X)_\rho$ can be directly evaluated by calculating $|A|$ trace distances, whilst a priori, it is defined as an optimization problem over an ε-ball. Whilst this might be of independent interest, it allows us to prove the promised Theorem.

Theorem 7: For $\rho \in \mathfrak{D}(AB)$, the ε-single-shot entanglement cost is given by

$$\mathrm{Cost}^{\varepsilon}\left(\rho^{AB}\right) = \inf_{\rho^{XAB}} H_{\max}^{\varepsilon}(A|X)_\rho, \tag{51}$$

where the infimum is over all classical systems X and all classical extensions ρ^{XAB} of ρ^{AB}. Moreover, the infimum is attained for a classical extension with $|X| = |AB|^2$ and can also be taken over all regular extensions of ρ^{AB}.

It is worth noticing that both in the above Theorem as well as in [17, Thm. 1], the ε in $H_{\max}^{\varepsilon}$ stands for smoothing with respect to the trace distance whilst the ε in $\mathrm{Cost}^{\varepsilon}$ corresponds to an error in conversion measured by the square of the purified distance. This is in contrast to Eq. (46) where also the conditional max-entropy is smoothed with respect to the square of the purified distance.

C. Single-Shot Entanglement Manipulation Characterized by Entropies

In this section, we show that the ε-single-shot distillable entanglement and the ε-single-shot entanglement cost of pure states can be expressed *exactly* in terms of a smoothed min-entropy and a smoothed max-entropy, respectively. We begin with the ε-single-shot entanglement cost as defined in Eq. (33): For a pure state $\rho^{AB} = \psi^{AB}$, the optimization over classical extensions ρ^{XAB} in Thm. 7 is trivial, since all of them are of the form $\rho^{XAB} = \sigma^X \otimes \psi^{AB}$. This allows us to simplify the expression for $\mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right)$: Let

$$H_{\max}(\rho) := \log \mathrm{Tr}[\Pi_\rho] \tag{52}$$

be the max-entropy [18] and

$$H_{\max}^{\varepsilon}(\rho^A) := \min_{\omega\in\mathfrak{B}_\varepsilon(\rho^A)} H_{\max}(\omega^A) \tag{53}$$

its ε-smoothed version [19]. Via Eqs. (40) and (43), we thus obtain the following Corollary.

Corollary 1: Let $\psi \in \mathrm{Pure}(AB)$ and $\rho^A = \mathrm{Tr}_B\left[\psi^{AB}\right]$. It then holds that

$$\mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right) = H_{\max}^{\varepsilon}(\rho^A). \tag{54}$$

This equips the smoothed max-entropy with an operational interpretation in terms of the ε-single-shot entanglement cost of pure states. Moreover, by choosing X to be a trivial classical system in Lem. 3, we obtain the following Corollary.

Corollary 2: Let $\rho \in \mathfrak{D}(A)$ and $\varepsilon \in [0,1]$. It then holds that

$$H_{\max}^{\varepsilon}(\rho^A) = \min_{m\in[|A|]} \left\{\log m : \left\|\rho^A\right\|_{(m)} \ge 1-\varepsilon\right\}. \tag{55}$$

Now notice that if $\psi \in \mathrm{Pure}(AB)$ and $\rho^A = \mathrm{Tr}_B\left[\psi^{AB}\right]$ as in Cor. 1, it holds that $\left\|\rho^A\right\|_{(m)} = \|\mathbf{p}\|_{(m)}$, where $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|)$ are the Schmidt coefficients of ψ^{AB}. We can thus rewrite Cor. 1 as

Corollary 3: Let $\psi \in \mathrm{Pure}(AB)$ and $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|)$ be the Schmidt coefficients of ψ^{AB}. It then holds that

$$\mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right) = \min_{m\in[|A|]} \left\{\log m : \|\mathbf{p}\|_{(m)} \ge 1-\varepsilon\right\}. \tag{56}$$

As claimed earlier, this also shows that the two definitions of $\mathrm{Cost}^{\varepsilon}$ provided in Eqs. (11) and (33) indeed coincide on pure states (see Thm. 3), and Prop. 1 thus applies to both. This is interesting, since in Eq. (11), ε was bounding the star conversion distance $T_\star$, whilst, in Eq. (33), it was bounding the square of the purified distance P^2. Moreover, we note that the previous discussion also implies that the quantity $f'_{n,\varepsilon}(\mathbf{p})$ in Lem. 1 has an operational interpretation in terms of the smoothed max-entropy.

By comparing Cor. 1 and Thm. 5, we note that

$$\begin{aligned} \mathrm{Cost}_c^{\varepsilon}\left(\psi^{AB}\right) &= \log\left\lceil 2^{H^{\varepsilon}(\mathbf{p})}\right\rceil \\ &\le \log\left\lceil 2^{H_{\max}^{\varepsilon}(\mathbf{p})}\right\rceil \\ &= \mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right). \end{aligned} \tag{57}$$

While this confirms our knowledge that catalysis does not increase the ε-single-shot entanglement cost, more importantly, it allows us to gain insides into the question when catalysis grants an advantage, i.e., when the inequality is strict. Consider for example the case $\varepsilon = 0$. If all non-zero entries of $\mathbf{p}$ are equal (i.e., if ψ^{AB} is separable or a maximally entangled state of arbitrary dimension), we have that $H_{\max}(\mathbf{p}) = H(\mathbf{p})$, and catalysis is useless. However, even if $H_{\max}(\mathbf{p}) \ne H(\mathbf{p})$, due to the ceiling function, it is possible

[1]Note that due to the possibility of degenerate eigenvalues, Π_m and thus $\sigma^{(m)}$ are not necessarily unique. However, for our purpose, this is irrelevant.

that catalysis does not provide an advantage. On the contrary, it is also easy to see that there exist ψ^{AB} such that catalysis provides an advantage. Interestingly, this depends not only on the difference between $H_{\max}(\mathbf{p})$ and $H(\mathbf{p})$, but also the value of $2^{H(\mathbf{p})}$. While $H_{\max}(\mathbf{p}) - H(\mathbf{p}) \geq 1$ guarantees an advantage, for $1 > H_{\max}(\mathbf{p}) - H(\mathbf{p}) > 0$, both having and not having an advantage is possible. Considering that $H^{\varepsilon}_{\max}(\mathbf{p})$ is not continuous in ε adds another layer of intricacy.

Next, we turn to the ε-single-shot distillable entanglement: Let

$$H_{\min}(\rho) := -\log \|\rho\|_\infty \tag{58}$$

(where $\|\rho\|_\infty$ denotes the operator norm of ρ, i.e., its largest eigenvalue) be the min-entropy [18] and

$$H^{\varepsilon}_{\min}(\rho^A) := \max_{\omega \in \mathfrak{B}_\varepsilon(\rho^A)} H_{\min}(\omega^A) \tag{59}$$

its ε-smoothed version [19]. Analogously to Cor. 2, we next provide a closed-form expression for $H^{\varepsilon}_{\min}(\rho^A)$ which might be of independent interest.

Lemma 4: Let $\rho \in \mathcal{D}(A)$ with $|A| = d$ and let $\mathbf{p} \in \mathrm{Prob}(d)$ contain the eigenvalues of ρ. Let further $\mathbf{u}^{(d)}$ be the flat distribution of dimension d. If $\frac{1}{2}\left\|\mathbf{p} - \mathbf{u}^{(d)}\right\|_1 \leq \varepsilon$, then

$$H^{\varepsilon}_{\min}(\rho^A) = \log d, \tag{60}$$

otherwise

$$H^{\varepsilon}_{\min}(\rho^A) = \log \min_{\ell \in [d]} \left\{ \frac{\ell}{\|\mathbf{p}\|_{(\ell)} - \varepsilon} \right\}. \tag{61}$$

In the Lemma above, we have two cases. With the help of the following Lemma, we will now see that this is indeed necessary, i.e., the first case is not included in the second.

Lemma 5: Let $\mathbf{q} \in \mathrm{Prob}(n)$ and $\mathbf{q} \neq \mathbf{u}^{(n)}$. Then

$$\left\|\mathbf{q} - \mathbf{u}^{(n)}\right\|_1 = 2 \max_{k \in [n]} \left(\|\mathbf{q}\|_{(k)} - \frac{k}{n} \right).$$

Using the notation of Lem. 4, we now assume that

$$0 < \frac{1}{2}\left\|\mathbf{p} - \mathbf{u}^{(d)}\right\|_1 < \varepsilon. \tag{62}$$

In case that $\mathbf{p} = \mathbf{u}^{(d)}$, we find that

$$\min_{y \in [d]} \frac{y}{\|\mathbf{p}\|_{(y)} - \varepsilon} = \min_{y \in [d]} \frac{y}{\frac{y}{d} - \varepsilon} > \min_{y \in [d]} \frac{y}{\frac{y}{d}} = d, \tag{63}$$

which shows that the two cases in Lem. 4 are needed. In case that $\mathbf{p} \neq \mathbf{u}^{(d)}$, it follows from our assumption and Lem. 5 that

$$\varepsilon > \max_{k \in [d]} \left(\|\mathbf{p}\|_{(k)} - \frac{k}{d} \right) \tag{64}$$

and thus

$$\min_{y \in [d]} \frac{y}{\|\mathbf{p}\|_{(y)} - \varepsilon} > \min_{y \in [d]} \frac{y}{\|\mathbf{p}\|_{(y)} - \left(\|\mathbf{p}\|_{(y)} - \frac{y}{d}\right)} = d, \tag{65}$$

i.e., the two cases are thus needed too. Importantly, this implies that in general, attaching an uncorrelated pure state ψ^C can change the ε-smoothed min-entropy of ρ^A in the sense that

$$H^{\varepsilon}_{\min}(\rho^A) < H^{\varepsilon}_{\min}(\rho^A \otimes \psi^C), \tag{66}$$

a property which we would not expect of an entropy. To resolve this, we define

$$\tilde{H}^{\varepsilon}_{\min}(\rho^A) := \sup_{\omega \in \mathfrak{B}_\varepsilon(\rho^A \otimes \psi^C)} H_{\min}\left(\omega^{AC}\right), \tag{67}$$

where it is understood that the supremum is also over all auxiliary systems C and ψ^C is an arbitrary pure state on this system. From Lem. 4, we obtain the following Corollary.

Corollary 4: Let $\rho \in \mathcal{D}(A)$ with $|A| = d$, $\varepsilon \in [0,1)$, and let $\mathbf{p} \in \mathrm{Prob}(d)$ contain the eigenvalues of ρ. It then holds that

$$\tilde{H}^{\varepsilon}_{\min}(\rho^A) = \log \min_{\ell \in [d]} \left\{ \frac{\ell}{\|\mathbf{p}\|_{(\ell)} - \varepsilon} \right\}. \tag{68}$$

Combining this Lemma with the closed-form expression in Thm. 2, we have the following entropic characterization of the ε-single-shot distillable entanglement.

Theorem 8: Let $\psi \in \mathrm{Pure}(AB)$ and $\varepsilon \in [0,1)$. The ε-single-shot distillable entanglement of ψ^{AB} is then given by

$$\mathrm{Distill}^{\varepsilon}(\psi^{AB}) = \log \left\lfloor 2^{\tilde{H}^{\varepsilon}_{\min}(\rho^A)} \right\rfloor,$$

where $\rho^A = \mathrm{Tr}_B(\psi^{AB})$ is the reduced density matrix of ψ^{AB}. It is now again interesting to note that also in the case of ε-single-shot entanglement distillation, there are cases in which catalysis provides an advantage and cases in which it does not. Consider for example two pure states ψ^{AB} and ϕ^{AB} with Schmidt coefficients $\mathbf{p} = (0.75, 0.15, 0.1)$ and $\mathbf{q} = (0.8, 0.2, 0)$, respectively. A straightforward calculation shows that

$$\mathrm{Distill}^{1/10}(\psi^{AB}) = 0 < 1 = \mathrm{Distill}^{1/10}_c(\psi^{AB}), \tag{69}$$

but

$$\mathrm{Distill}^{1/10}(\phi^{AB}) = 0 = \mathrm{Distill}^{1/10}_c(\phi^{AB}). \tag{70}$$

At this point, we want to mention that [14] investigated a variant of the ε-single-shot distillable entanglement too: The authors of [14] defined the fidelity of distillation as

$$F(\rho, m) := \sup_{\Lambda \in \mathrm{LOCC}} \mathrm{Tr}(\Lambda(\rho)\Phi_m) \tag{71}$$

and their variant of the ε-single-shot distillable entanglement as

$$E^{(1),\varepsilon}_D(\rho) := \log \max\{m \geq 2 : F(\rho, m) \geq 1 - \varepsilon\}. \tag{72}$$

Comparing this definition to our definition in Eq. (9), we see that they differ in the way in which the allowed error ε is introduced: Whilst we bound the error using the star conversion distance [15], in [14], it was demanded that $1 - F \leq \varepsilon$. Their definition does therefore not coincide with the one we used. Importantly, for pure states, there exists a closed-form formula for $E^{(1),\varepsilon}_D(\psi)$ too [14, Thm. 15, Cor. 16] which is given in terms of a distillation norm [62]. In App. D, we discuss in detail how this allows us to efficiently compute $E^{(1),\varepsilon}_D(\psi^{\otimes n})$ in a manner that is very similar to how we can compute $\mathrm{Distill}^{\varepsilon}\left(\psi^{\otimes n}\right)$. To conclude the discussion concerning the entanglement manipulation of quantum states, we want to mention that from Refs. [16] and [61], also analogs of our Thm. 3/Cor. 1 and Thm. 2 can be extracted, again with definitions of the error that differ from ours. Our choices of the conversion distance lead to particularly compact formulas.

D. Single-Shot Entanglement Cost of Channels

In the following section, we will bound the entanglement that it costs to simulate an arbitrary channel between two parties with a given precision. Since any quantum state can

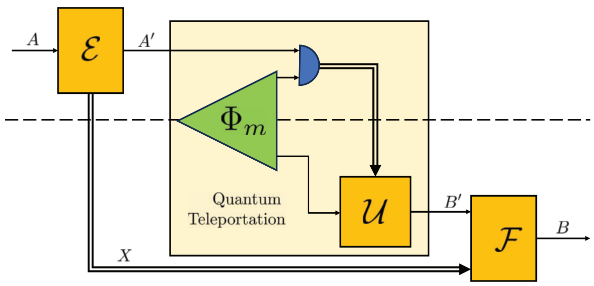

Fig. 2. Optimal simulation of a quantum channel $\mathcal{N}^{A\to B}$ if given access to Φ_m and LOCC using quantum teleportation. Solid lines represent quantum systems, double lines classical systems, and the dashed line the spacial separation between Alice and Bob.

be identified with its corresponding replacement channel, this can be seen as a generalization of the results presented in the previous section. Moreover, quantum teleportation [3] demonstrates that LOCC and one ebit can be used to simulate one identity qubit channel. It will therefore become apparent when we talk about optimal protocols that our results are closely related to teleportation too, see Fig. 2.

As in the state case, we will start by defining how we quantify the error of a simulation $\mathcal{N}'$ of a given channel $\mathcal{N}$: For any channel $\mathcal{N} \in \mathrm{CPTP}(A \to B)$ and a maximally entangled state Φ_m, we define the channel conversion fidelity as

$$F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{A\to B}\right) := \sup_{\Theta} \min_{\psi\in\mathrm{Pure}(A\tilde{A})} F\left(\Theta\left[\Phi_m\right](\psi^{A\tilde{A}}), \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right), \tag{73}$$

where the supremum is over all LOCC superchannels Θ that map the state Φ_m to a channel in $\mathrm{CPTP}(\tilde{A} \to B)$. Analogously to the ε-single-shot entanglement cost of a bipartite state defined in Eq. (33), we use the channel conversion fidelity to define the ε-single-shot entanglement cost of the channel $\mathcal{N}^{A\to B}$ as

$$\mathrm{Cost}^{\varepsilon}(\mathcal{N}) := \inf_{m\in\mathbb{N}} \left\{ \log m \;:\; P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}\right) \le \varepsilon \right\}, \tag{74}$$

where

$$P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}\right) := 1 - F^2\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}\right). \tag{75}$$

To bound $\mathrm{Cost}^{\varepsilon}(\mathcal{N})$, we will use that the supremum and minimum in the definition of the channel conversion fidelity can be exchanged: This is the content of the following Lemma, which can be shown with the help of [63, Prop. 8] and [64, Lem. II.3].

Lemma 6: Let $F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{A\to B}\right)$ be defined as in Eq. (73). It holds that

$$F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{A\to B}\right) = \min_{\psi\in\mathrm{Pure}(A\tilde{A})} \sup_{\Theta} F\left(\Theta[\Phi_m](\psi^{A\tilde{A}}), \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right) \tag{76}$$

where the supremum is again over all LOCC superchannels Θ that map the state Φ_m to a channel in $\mathrm{CPTP}(\tilde{A} \to B)$.

Having access to Φ_m and LOCC, Alice and Bob can use quantum teleportation to simulate the identity channel $\mathrm{id}^{A'\to B'}$, where $|A'| = |B'| = m$. Conversely, $\mathrm{id}^{A'\to B'}$ allows us to create Φ_m. Therefore, $\Phi_m \xleftrightarrow{\text{LOCC}} \mathrm{id}^{A'\to B'}$, i.e., the state Φ_m is equivalent to the channel $\mathrm{id}^{A'\to B'}$. In Eqs. (73) and (76), we can thus replace $\Theta\left[\Phi_m\right]$ with $\Theta\left[\mathrm{id}^{A'\to B'}\right]$ and take the supremums over all LOCC superchannels that map the identity channel $\mathrm{id}^{A'\to B'}$ to a channel in $\mathrm{CPTP}(\tilde{A} \to B)$. Importantly, every such superchannel can be expanded as

$$\begin{aligned}\Theta\left[\mathrm{id}^{A'\to B'}\right] &= \mathcal{F}^{B'X\to B} \circ \mathrm{id}^{A'\to B'} \circ \mathcal{E}^{\tilde{A}\to A'X} \\ &= \mathcal{F}^{B'X\to B} \circ \mathcal{E}^{\tilde{A}\to B'X} \\ &= \sum_{x\in[k]} \mathcal{F}_{(x)}^{B'\to B} \circ \mathcal{E}_x^{\tilde{A}\to B'},\end{aligned} \tag{77}$$

where X is a classical system (that can be exchanged via LOCC), $\mathcal{E}^{\tilde{A}\to A'X} \in \mathrm{CPTP}(\tilde{A} \to A'X)$, and $\mathcal{F}^{B'X\to B} \in \mathrm{CPTP}(B'X \to B)$ (see Fig. 3). In the last line, we utilized that this can be seen as the average of an instrument $\{\mathcal{E}_x^{\tilde{A}\to B'}\}_x$ and a channel $\mathcal{F}_{(x)}^{B'\to B}$ conditioned on its classical outcome x. From this follows that the superchannels that we need to consider in Eqs. (73) and (76) are of the form shown in Fig. 2, highlighting the close relation to quantum teleportation.

The Theorem below shows that the conversion fidelity is closely related to the monotones from Eq. (35), which can be extended to quantum channels in the usual manner, i.e.,

$$E_{(k)}\left(\mathcal{N}^{A\to B}\right) := \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(k)}\left(\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right). \tag{78}$$

Using this definition, we get the following result, which is the channel analog of Eq. (37).

Lemma 7: Let $\mathcal{N} \in \mathrm{CPTP}(A \to B)$ be a quantum channel. It then holds that

$$\begin{aligned}&1 - E_{(m)}\left(\mathcal{N}^{A\to B}\right) \\ &\le F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{A\to B}\right) \\ &\le \sqrt{1 - E_{(m)}\left(\mathcal{N}^{A\to B}\right)}.\end{aligned} \tag{79}$$

This allows us to provide the promised bounds on $\mathrm{Cost}^{\varepsilon}(\mathcal{N}^{A\to B})$.

Theorem 9: Let $\mathcal{N} \in \mathrm{CPTP}(A \to B)$ be a quantum channel and $\varepsilon \in [0, 1)$. Then

$$\begin{aligned}&\max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\sigma^{XAB}} H_{\max}^{\varepsilon}(A|X)_{\sigma} \\ &\le \mathrm{Cost}^{\varepsilon}(\mathcal{N}^{A\to B}) \\ &\le \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\sigma^{XAB}} H_{\max}^{\varepsilon/2}(A|X)_{\sigma},\end{aligned} \tag{80}$$

where the infimums are over all classical systems X and all classical extensions σ^{XAB} of $\sigma^{AB} = \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})$. Again, the infimums are attained for a regular/classical extension with $|X| = |AB|^2$.

In the limit of ε approaching zero, this Theorem provides the zero-error single-shot entanglement cost of a quantum channel. With the help of Thm. 7, we can express the Theorem as

$$\max_{\psi\in\mathrm{Pure}(A\tilde{A})} \mathrm{Cost}^{\varepsilon}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)$$

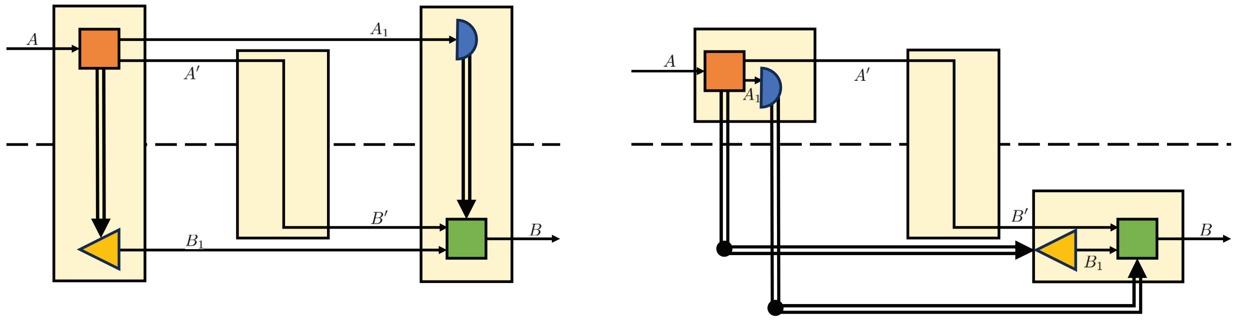

Fig. 3. Left: LOCC superchannel converting $\mathrm{id}^{A'\to B'}$ (center) to a channel $\mathcal{N} \in \mathrm{CPTP}(A \to B)$. Right: Any such superchannel can be realized by the following protocol: Alice applies a quantum instrument, sends the quantum output through the identity channel, and Bob applies a channel on its output that is conditioned on the classical outcome of Alice's instrument. Solid lines represent quantum systems, double lines classical systems, and the dashed line the spacial separation between Alice and Bob.

$$\begin{aligned}&\le \mathrm{Cost}^{\varepsilon}(\mathcal{N}^{A\to B})\\ &\le \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \mathrm{Cost}^{\varepsilon/2}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right).\end{aligned} \tag{81}$$

This shows that $\mathrm{Cost}^{\varepsilon}(\mathcal{N}^{A\to B})$ is lower bounded by the cost of the most expensive state that we can create with its help. This is to be expected for a consistent definition, since otherwise one might be able to build an entanglement perpetuum mobile. It is however not obvious that this should also be an upper bound since we intend to simulate a channel on an unknown input state and potentially do not have access to another correlated system (such as system A in Eq. (73)): Simply replacing the input state with (an approximation of) the corresponding output state of the channel we intend to simulate is thus not an option. We conclude this section by noting that with the help of postselection techniques [65], one can recover the asymptotic entanglement cost of a quantum channel [66] from Thm. 9.

IV. One-Way Single-Shot Entanglement Manipulation

A. State Distillation

In the following, we will explore a resource measure with respect to one-way LOCC. While this measure may not exhibit monotonicity under arbitrary LOCC, it can still be a valuable tool for providing bounds on the distillable entanglement of mixed bipartite states, as we will see in the following.

The most general one-way LOCC operation that Alice and Bob can perform is for Alice to apply a quantum instrument $\{\mathcal{E}_x\}_{x\in[n]}$, with $\mathcal{E}_x \in \mathrm{CP}(A \to A')$ and $\sum_{x=1}^{n} \mathcal{E}_x \in \mathrm{CPTP}(A \to A')$, send the outcome x to Bob, who then applies a quantum channel $\mathcal{F}_{(x)} \in \mathrm{CPTP}(B \to B')$ that depends on the outcome x received from Alice [67]. The overall operation can be described by the quantum channel

$$\mathcal{N}^{AB\to A'B'} := \sum_{x=1}^{n} \mathcal{E}_x^{A\to A'} \otimes \mathcal{F}_{(x)}^{B\to B'} . \tag{82}$$

Definition 1: The *coherent information of entanglement* of a state $\rho \in \mathfrak{D}(AB)$ is defined as

$$E_{\to}\left(\rho^{AB}\right) := \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I\big(A\rangle BX\big)_{\mathcal{E}(\rho)} , \tag{83}$$

where the supremum includes a supremum over all classical systems X with arbitrary dimension and

$$I(A\rangle B)_{\rho} := -H(A|B)_{\rho} \tag{84}$$

is the coherent information [68].

As promised, we will now show that the coherent information of entanglement is a resource measure with respect to one-way LOCC.

Theorem 10: Let $\rho \in \mathfrak{D}(AB)$ with $m = |A| = |B|$, $\sigma \in \mathfrak{D}(A'B')$, and $\mathcal{N} \in \mathrm{LOCC}_1(AB \to A'B')$. The coherent information of entanglement $E_{\to}$ is

1) monotonic under one-way LOCC, i.e.,
$$E_{\to}\left(\mathcal{N}^{AB\to A'B'}\left(\rho^{AB}\right)\right) \le E_{\to}\left(\rho^{AB}\right) ,$$
2) non-negative, i.e., $E_{\to}(\rho^{AB}) \ge 0$, with equality if ρ^{AB} is separable,
3) strongly monotonic under one-way LOCC, i.e., for any ensemble $\{p_y, \sigma_y^{A'B'}\}$ that can be obtained from ρ^{AB} using one-way LOCC and subselection, it holds that
$$E_{\to}\left(\rho^{AB}\right) \ge \sum_y p_y E_{\to}\left(\sigma_y^{A'B'}\right),$$
4) convex,
5) bounded by $E_{\to}\left(\rho^{AB}\right) \le \log(m) = E_{\to}(\Phi_m)$,
6) and superadditive, i.e.,
$$E_{\to}\left(\rho^{AB}\otimes\sigma^{A'B'}\right) \ge E_{\to}\left(\rho^{AB}\right) + E_{\to}\left(\sigma^{A'B'}\right).$$

The ε-single-shot distillable entanglement under one-way LOCC (and the error bounded by the trace distance) is defined as

$$\begin{aligned}&\mathrm{Distill}^{\varepsilon}_{\to}\left(\rho^{AB}\right)\\ &:= \max\left\{\log m \;:\; T\left(\rho^{AB} \xrightarrow{\mathrm{LOCC}_1} \Phi_m\right) \le \varepsilon\right\}.\end{aligned} \tag{85}$$

A simple formula for the above expression is presently not available. However, we can provide an upper bound.

Theorem 11: Let $\rho \in \mathfrak{D}(AB)$ and $\varepsilon \in (0, 1/2)$. Then, the one-way ε-single-shot distillable entanglement is bounded by

$$\mathrm{Distill}^{\varepsilon}_{\to}\left(\rho^{AB}\right) \le \frac{1}{1-2\varepsilon}E_{\to}\left(\rho^{AB}\right) + \frac{1+\varepsilon}{1-2\varepsilon}h\left(\frac{\varepsilon}{1+\varepsilon}\right) , \tag{86}$$

where $h(x) := -x\log x - (1-x)\log(1-x)$ is the binary Shannon entropy.

This result should be compared to [69, Lem. 4], where also an upper bound on a variant of the one-way ε-single-shot distillable entanglement is provided (again with a slightly different definition of the allowed error in terms of the fidelity). The main difference is that in our bound, the optimization over instruments (contained in $E_{\rightarrow}\left(\rho^{AB}\right)$) is independent of ε, whilst the corresponding optimization in their bound is not. The bound provided in Thm. 11 recovers the exact asymptotic solution given in [70, Thm. 13].

For lower bounds on the one-way ε-single-shot distillable entanglement, see again [69] as well as [71, Prop. 21]. Even though [71] defined the conversion distance using the fidelity, their bound holds for our definition too due to the following Lemma.

Lemma 8: Let $\rho \in \mathfrak{D}(AB)$, and $\Phi_m \in \mathfrak{D}(A'B')$ be the maximally entangled state with $m := |A'| = |B'|$. Then,

$$T\left(\rho \xrightarrow{\text{LOCC}_1} \Phi_m\right) = P^2\left(\rho \xrightarrow{\text{LOCC}_1} \Phi_m\right) = 1 - \sup_{\mathcal{N}\in\text{LOCC}_1} \text{Tr}\left[\Phi_m \mathcal{N}(\rho)\right] , \quad (87)$$

where the supremum is over all $\mathcal{N} \in \text{LOCC}_1(AB \to A'B')$, and P is the purified distance as given in Eq. (7).

B. Channel Distillation

In Def. 1, we introduced the coherent information of entanglement of a quantum state and subsequently showed that it is a measure of entanglement under one-way LOCC. The following Definition contains the generalization to quantum channels.

Definition 2: Let $\mathcal{N} \in \text{CPTP}(A \to B)$ be a quantum channel. Its coherent information of entanglement is then defined as

$$E_{\rightarrow}\left(\mathcal{N}^{A\to B}\right) := \max_{\psi\in\text{Pure}(A\tilde{A})} E_{\rightarrow}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) . \quad (88)$$

Similarly, one can define the coherent information of a quantum channel $\mathcal{N}^{A\to B}$ as

$$I(A\rangle B)_{\mathcal{N}} := \max_{\psi\in\text{Pure}(A\tilde{A})} I(A\rangle B)_{\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})} . \quad (89)$$

Interestingly, $E_{\rightarrow}\left(\mathcal{N}^{A\to B}\right)$ and $I(A\rangle B)_{\mathcal{N}}$ coincide.

Theorem 12: Let $\mathcal{N} \in \text{CPTP}(A \to B)$ be a quantum channel. It then holds that

$$E_{\rightarrow}\left(\mathcal{N}^{A\to B}\right) = I(A\rangle B)_{\mathcal{N}} . \quad (90)$$

Consider a channel $\mathcal{N} \in \text{CPTP}(A \to B)$. The most general bipartite state $\rho \in \mathfrak{D}(A_3B_3)$ to which this channel can be converted with the help of LOCC_1 is given by

$$\rho^{A_3B_3} = \mathcal{E}^{A_2B\to A_3B_3} \circ \mathcal{N}^{A\to B}\left(\sigma^{AA_2}\right) , \quad (91)$$

where $\mathcal{E}$ is in LOCC_1. Since we can purify σ by enlarging system A_2 and adapting $\mathcal{E}$ accordingly, w.l.o.g., we will assume that σ is pure. For any natural number m, let $|A_3| = |B_3| = m$ and we thus define the conversion distance

$$T\left(\mathcal{N}^{A\to B} \xrightarrow{\text{LOCC}_1} \Phi_m\right) := \frac{1}{2} \inf_{\substack{\mathcal{E}\in\text{LOCC}_1\\ \psi\in\text{Pure}}} \left\|\Phi_m^{A_3B_3} - \mathcal{E}^{A_2B\to A_3B_3} \circ \mathcal{N}^{A\to B}\left(\psi^{AA_2}\right)\right\|_1 , \quad (92)$$

where a priori, the infimum also includes an infimum over $|A_2|$. Since the Schmidt rank of ψ^{AA_2} cannot exceed $|A|$, w.l.o.g., we can fix $|A_2| = |A|$ (and adapt $\mathcal{E}$ accordingly). Observe that the equation above implies that (see Eq. (3))

$$T\left(\mathcal{N}^{A\to B} \xrightarrow{\text{LOCC}_1} \Phi_m\right) = \min_{\psi,\phi\in\text{Pure}} T\left(\mathcal{N}^{A\to B}\left(\psi^{A\tilde{A}}\right) \xrightarrow{\text{LOCC}_1} \Phi_m\right) . \quad (93)$$

The one-way ε-single-shot distillable entanglement is then defined as (cf. Eq. (85))

$$\text{Distill}_{\rightarrow}^{\varepsilon}(\mathcal{N}) := \max\left\{\log m : T\left(\mathcal{N} \xrightarrow{\text{LOCC}_1} \Phi_m\right) \le \varepsilon\right\} . \quad (94)$$

Therefore, from Eqs. (85) and (93), we get that

$$\text{Distill}_{\rightarrow}^{\varepsilon}\left(\mathcal{N}^{A\to B}\right) = \max_{\psi,\phi\in\text{Pure}} \text{Distill}^{\varepsilon}\left(\mathcal{N}^{A\to B}\left(\psi^{A\tilde{A}}\right)\right) . \quad (95)$$

Combining Thm. 11 and Thm. 12, we thus obtain

Theorem 13: Let $\mathcal{N} \in \text{CPTP}(A \to B)$ be a quantum channel. It then holds that

$$\text{Distill}_{\rightarrow}^{\varepsilon}(\mathcal{N}) \le \frac{1}{1-2\varepsilon} I(A\rangle B)_{\mathcal{N}} + \frac{1+\varepsilon}{1-2\varepsilon} h\left(\frac{\varepsilon}{1+\varepsilon}\right) . \quad (96)$$

V. Discussion and Outlook

In this work, we studied entanglement distillation and dilution of states and channels in the single-shot regime [9], [10], [13], [17], [22], [31], [36], [37], [38], [39], [40], [43], [72]. By restricting the allowed error ε with the recently introduced star conversion distance [15], we determined surprisingly compact closed-form expressions for the ε-single-shot entanglement cost and the ε-single-shot distillable entanglement of pure states, which allowed us to obtain second-order asymptotics [16] and efficient methods to calculate these quantities on multiple copies. Furthermore, we discussed generalizations of ε-single-shot entanglement distillation/dilution in which either catalysis [51] is allowed or the maximally entangled states are replaced by arbitrary pure entangled states. Since these results are based on (approximate) majorization [45], [46], we expect that similar results can be obtained in other majorization-based resource theories such as coherence [73], [74], non-uniformity [75], or quantum thermodynamics [76], [77], [78], [79], [80].

For mixed states, we expressed the ε-single-shot entanglement cost introduced in [17] in terms of a smoothed version of the conditional max-entropy [18], [19] and showed that on pure states, it coincides with the ε-single-shot entanglement cost based on the star conversion distance.

Furthermore, we provided closed-form expressions for the smoothed min- and max-entropy [18], [19] and equipped them with an operational interpretation in terms of the ε-single-shot distillable entanglement and entanglement cost of pure states, respectively. Based on these results, we provided bounds on the entanglement cost of quantum channels that coincide in the zero-error limit. Concerning entanglement distillation, we introduced the coherent information of entanglement and used it to upper bound both the one-way ε-single-shot distillable entanglement of states and channels.

Our work thus contributes to a better understanding of how entanglement can be optimally manipulated and used to implement a desired quantum channel and how this is related to entropic quantities. This is highly relevant to optimize technological applications in which entanglement plays a role, which will lead to a better understanding of the relevance of entanglement for quantum advantages.

Appendix A
Additional Remarks

In the Appendix, for completeness, we provide a few additional comments. Technically Eq. (6) was not shown in [15], but it follows directly from what was shown: As expected for a consistent definition, and because one can always append and remove separable states reversibly,

$$\begin{aligned}&T\left(\rho^{AB}\xrightarrow{\text{LOCC}}\sigma^{A'B'}\right)\\&=T\left(\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\sigma^{A'B'}\right).\end{aligned}\tag{97}$$

A technical proof is the following:

$$\begin{aligned}&T\left(\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\sigma^{A'B'}\right)\\&=\inf_{\tau\in\mathfrak{D}(AA'BB')}\left\{\frac{1}{2}\left\|\tau^{AA'BB'}-|11\rangle\langle 11|^{AB}\otimes\sigma^{A'B'}\right\|_1:\right.\\&\qquad\left.\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}\tau^{AA'BB'}\right\}\\&\le\inf_{\tau\in\mathfrak{D}(A'B')}\left\{\frac{1}{2}\left\||11\rangle\langle 11|^{AB}\otimes\tau^{A'B'}-|11\rangle\langle 11|^{AB}\otimes\sigma^{A'B'}\right\|_1:\right.\\&\qquad\left.\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\tau^{A'B'}\right\}\\&=\inf_{\tau\in\mathfrak{D}(A'B')}\left\{\frac{1}{2}\left\|\tau^{A'B'}-\sigma^{A'B'}\right\|_1:\rho^{AB}\xrightarrow{\text{LOCC}}\tau^{A'B'}\right\}\\&=T\left(\rho^{AB}\xrightarrow{\text{LOCC}}\sigma^{A'B'}\right)\end{aligned}\tag{98}$$

and

$$\begin{aligned}&T\left(\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\sigma^{A'B'}\right)\\&=\inf_{\tau\in\mathfrak{D}(AA'BB')}\left\{\frac{1}{2}\left\|\tau^{AA'BB'}-|11\rangle\langle 11|^{AB}\otimes\sigma^{A'B'}\right\|_1:\right.\\&\qquad\left.\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}\tau^{AA'BB'}\right\}\\&\ge\inf_{\tau\in\mathfrak{D}(AA'BB')}\left\{\frac{1}{2}\left\|\mathrm{Tr}_{AB}\,\tau^{AA'BB'}-\sigma^{A'B'}\right\|_1:\right.\\&\qquad\left.\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}\tau^{AA'BB'}\right\}\\&\ge\inf_{\tau\in\mathfrak{D}(AA'BB')}\left\{\frac{1}{2}\left\|\mathrm{Tr}_{AB}\,\tau^{AA'BB'}-\sigma^{A'B'}\right\|_1:\right.\\&\qquad\left.\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}\mathrm{Tr}_{AB}\,\tau^{AA'BB'}\right\}\\&=\inf_{\tau\in\mathfrak{D}(A'B')}\left\{\frac{1}{2}\left\|\tau^{A'B'}-\sigma^{A'B'}\right\|_1:\rho^{AB}\xrightarrow{\text{LOCC}}\tau^{A'B'}\right\}\\&=T\left(\rho^{AB}\xrightarrow{\text{LOCC}}\sigma^{A'B'}\right).\end{aligned}\tag{99}$$

We, therefore, find with the help of [15, Lem. 3] that

$$\begin{aligned}&\frac{1}{2}T_\star^2\left(\psi^{AB}\xrightarrow{\text{LOCC}}\phi^{A'B'}\right)\\&=\frac{1}{2}T_\star^2\left(\psi^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\phi^{A'B'}\right)\\&\le T\left(\psi^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\phi^{A'B'}\right)\\&=T\left(\psi^{AB}\xrightarrow{\text{LOCC}}\phi^{A'B'}\right)\\&=T\left(\psi^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\phi^{A'B'}\right)\\&\le\sqrt{2T_\star\left(\psi^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\phi^{A'B'}\right)}\\&=\sqrt{2T_\star\left(\psi^{AB}\xrightarrow{\text{LOCC}}\phi^{A'B'}\right)}.\end{aligned}\tag{100}$$

Analogously to Eq. (97), and as expected, it also holds that

$$\begin{aligned}&P\left(\rho^{AB}\xrightarrow{\text{LOCC}}\sigma^{A'B'}\right)\\&=P\left(\rho^{AB}\otimes|11\rangle\langle 11|^{A'B'}\xrightarrow{\text{LOCC}}|11\rangle\langle 11|^{AB}\otimes\sigma^{A'B'}\right),\end{aligned}\tag{101}$$

which follows from exactly the same arguments.

Appendix B
Proofs of the Results in the Main Text

In the following, we provide the proofs of the results presented in the main text, which we repeat for readability. Note that Thm. 1 was stated in [15] for the case $AB=A'B'$. The generalization to $AB\neq A'B'$ is trivial. For completeness, we provide it nevertheless.

Theorem 1: Let $\psi\in\mathrm{Pure}(AB),\phi\in\mathrm{Pure}(A'B')$, and $\mathbf{p}\in\mathrm{Prob}^{\downarrow}(|A|),\mathbf{q}\in\mathrm{Prob}^{\downarrow}(|A'|)$ be their corresponding Schmidt coefficients. Then,

$$T_\star\left(\psi^{AB}\to\phi^{A'B'}\right)=\max_{k\in[\mathrm{SR}(\psi^{AB})]}\left\{\|\mathbf{p}\|_{(k)}-\|\mathbf{q}\|_{(k)}\right\}.\tag{102}$$

Proof: We notice that the (by convention ordered) Schmidt coefficients of

$$\psi^{AB}\otimes|11\rangle\langle 11|^{A'B'}\tag{103}$$

are given by

$$\mathbf{e}_1^{(|A'|)}\otimes\mathbf{p},\tag{104}$$

where $\mathbf{e}_1^{(|A'|)}=(1,0,\dots,0)\in\mathrm{Prob}^{\downarrow}(|A'|)$. The Schmidt coefficients of

$$|11\rangle\langle 11|^{AB}\otimes\phi^{A'B'}\tag{105}$$

on the other hand, are given by

$$\mathbf{e}_1^{(|A|)}\otimes\mathbf{q}.\tag{106}$$

The claim then follows from Eq. (5) and [15, Thm. 4].

Theorem 2: Let $\varepsilon \in [0,1)$, $\psi \in \mathrm{Pure}(AB)$, $d := \mathrm{SR}(\psi^{AB})$, and $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|)$ be the Schmidt coefficients of ψ^{AB}. The ε-single-shot distillable entanglement of ψ^{AB} is then given by

$$\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) = \min_{k\in\{\ell,\dots,d\}} \log\left\lfloor \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon}\right\rfloor , \tag{107}$$

where $\ell \in [d]$ is the integer satisfying $\|\mathbf{p}\|_{(\ell-1)} \le \varepsilon < \|\mathbf{p}\|_{(\ell)}$.

Proof: From Thm. 1, we find that for any $m \in \mathbb{N}$

$$T_{\star}\left(\psi^{AB} \to \Phi_m\right) = \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \|\Phi_m\|_{(k)}\right\}. \tag{108}$$

For $m \ge d$, this implies that

$$T_{\star}\left(\psi^{AB} \to \Phi_m\right) = \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{k}{m}\right\} . \tag{109}$$

Next, consider the case $m < d$, i.e.,

$$T_{\star}\left(\psi^{AB} \to \Phi_m\right) = \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{\min\{k,m\}}{m}\right\}. \tag{110}$$

Now assume that there exists an optimizer $k^{\star}$ in the above expression that satisfies $k^{\star} \ge m$. In this case, it must be equal to d, since $\|\mathbf{p}\|_{(k)}$ is strictly increasing for $k \in \{m,\dots,d\}$, whilst $\frac{\min\{k,m\}}{m} = 1$ is constant. This implies that $T_{\star}\left(\psi^{AB} \to \Phi_m\right) = 0$. Assume on the contrary that there only exist optimizers $k^{\star} < m$, and thus $T_{\star}\left(\psi^{AB} \to \Phi_m\right) > 0$ (otherwise $k^{\star} = d$ would be an optimizer too). In this case, we have that

$$\begin{aligned} T_{\star}\left(\psi^{AB} \to \Phi_m\right) &= \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{\min\{k,m\}}{m}\right\} \\ &= \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{k}{m}\right\}, \end{aligned} \tag{111}$$

since the maximum in the first line is reached for some $k < m$, and

$$\frac{\min\{k,m\}}{m} \le \frac{k}{m} \tag{112}$$

with equality for $k \le m$.

In summary, we thus find that for all $m \in \mathbb{N}$, either $T_{\star}\left(\psi^{AB} \to \Phi_m\right) = 0$, or

$$T_{\star}\left(\psi^{AB} \to \Phi_m\right) = \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{k}{m}\right\} . \tag{113}$$

Now let $\tilde{m}$ be the largest m such that $T_{\star}\left(\psi^{AB} \to \Phi_m\right) = 0$ (and note that $\tilde{m} \le d$, since LOCC cannot increase the Schmidt rank of any state). Remembering that by definition

$$\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) := \max_{m\in\mathbb{N}} \left\{\log m \;:\; T_{\star}\left(\psi^{AB} \to \Phi_m\right) \le \varepsilon\right\},$$

this implies that

$$\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) \ge \log\tilde{m}, \tag{114}$$

with equality iff $T_{\star}\left(\psi^{AB} \to \Phi_{\tilde{m}+1}\right) > \varepsilon$. Assuming that this is not the case, i.e., that $\varepsilon \ge T_{\star}\left(\psi^{AB} \to \Phi_{\tilde{m}+1}\right) > 0$, it follows from Eq. (113) that

$$\begin{aligned} &\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) \\ &= \max_{m\in\mathbb{N}} \left\{\log m \;:\; \|\mathbf{p}\|_{(k)} - \frac{k}{m} \le \varepsilon \;\; \forall\, k \in [d]\right\} \\ &= \max_{m\in\mathbb{N}} \left\{\log m \;:\; \|\mathbf{p}\|_{(k)} - \frac{k}{m} \le \varepsilon \;\; \forall\, k \in \{\ell,\dots,d\}\right\} \\ &= \max_{m\in\mathbb{N}} \left\{\log m \;:\; m \le \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon} \;\; \forall\, k \in \{\ell,\dots,d\}\right\} \\ &= \max_{m\in\mathbb{N}} \left\{\log m \;:\; m \le \min_{k\in\{\ell,\dots,d\}} \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon}\right\} \\ &= \min_{k\in\{\ell,\dots,d\}} \log\left\lfloor \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon}\right\rfloor . \end{aligned} \tag{115}$$

Moreover, by definition,

$$\tilde{m} = \max\left\{m \in [d] : \|\mathbf{p}\|_{(k)} - \frac{\min\{k,m\}}{m} \le 0 \;\forall k \in [d]\right\}.$$

Since $\|\mathbf{p}\|_{(k)} \le 1$, for $k > m$, $\|\mathbf{p}\|_{(k)} - \frac{\min\{k,m\}}{m} = \|\mathbf{p}\|_{(k)} - 1 \le 0$ and thus

$$\begin{aligned} \tilde{m} &= \max\left\{m \in [d] : \|\mathbf{p}\|_{(k)} - \frac{k}{m} \le 0 \;\forall k \in [m]\right\} \\ &= \max\left\{m \in [d] : \|\mathbf{p}\|_{(k)} - \frac{k}{m} \le 0 \;\forall k \in [d]\right\}, \end{aligned} \tag{116}$$

from which follows that

$$\max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{k}{\tilde{m}}\right\} \le 0 \le \varepsilon. \tag{117}$$

To conclude the proof, assume that $\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) = \log\tilde{m}$. According to Eq. (114), this implies that

$$T_{\star}\left(\psi^{AB} \to \Phi_{\tilde{m}+1}\right) = \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{k}{\tilde{m}+1}\right\} > \varepsilon ,$$

and consequently

$$\tilde{m} = \max\left\{m \in \mathbb{N} \;:\; \max_{k\in[d]} \left\{\|\mathbf{p}\|_{(k)} - \frac{k}{m}\right\} \le \varepsilon\right\}. \tag{118}$$

From this follows again that

$$\begin{aligned} &\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) = \log\tilde{m} \\ &= \max_{m\in\mathbb{N}} \left\{\log m : \|\mathbf{p}\|_{(k)} - \frac{k}{m} \le \varepsilon \;\forall\, k \in [d]\right\} \end{aligned} \tag{119}$$

Continuing with the same steps as in Eq. (115) finishes the proof.

Theorem 3: Let $\varepsilon \in [0,1)$, $\psi \in \mathrm{Pure}(AB)$, $d := \mathrm{SR}(\psi^{AB})$, and $\mathbf{p} \in \mathrm{Prob}^{\downarrow}(|A|)$ be the Schmidt coefficients of ψ^{AB}. The ε-single-shot entanglement cost of ψ^{AB} is then given by

$$\mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right) = \log m, \tag{120}$$

where $m \in [d]$ is the integer satisfying $\|\mathbf{p}\|_{(m-1)} < 1-\varepsilon \le \|\mathbf{p}\|_{(m)}$.

Proof: Let

$$b_k := \frac{k}{\|\mathbf{p}\|_{(k)} + \varepsilon} \qquad \forall\, k \in [m] . \tag{121}$$

Since for all $k \in [m-1]$, it holds that

$$\|\mathbf{p}\|_{(k)} + \varepsilon \ge \|\mathbf{p}\|_{(k)} \ge k p_k \ge k p_{k+1}, \tag{122}$$

we find that

$$\begin{aligned} \frac{b_{k+1}}{b_k} &= \frac{k+1}{k} \frac{\|\mathbf{p}\|_{(k)} + \varepsilon}{\|\mathbf{p}\|_{(k)} + p_{k+1} + \varepsilon} \\ &= \frac{k(\|\mathbf{p}\|_{(k)} + \varepsilon) + \|\mathbf{p}\|_{(k)} + \varepsilon}{k(\|\mathbf{p}\|_{(k)} + \varepsilon) + k p_{k+1}} \ge 1 \end{aligned} \tag{123}$$

and thus

$$b_1 \le b_2 \le \cdots \le b_m . \tag{124}$$

From Thm. 1, it follows that for any $m \in \mathbb{N}$

$$T_\star\left(\Phi_m \to \psi^{AB}\right) = \max_{k\in[m]}\left\{\frac{k}{m} - \|\mathbf{p}\|_{(k)}\right\}. \tag{125}$$

In combination, it, therefore, holds that

$$\begin{aligned}&\mathrm{Cost}^{\varepsilon}\left(\psi^{AB}\right) &(126)\\ &= \min_{m\in\mathbb{N}}\left\{\log m \;:\; \frac{k}{m} - \|\mathbf{p}\|_{(k)} \le \varepsilon \quad \forall\, k\in[m]\right\} &(127)\\ &= \min_{m\in\mathbb{N}}\left\{\log m \;:\; m \ge \frac{k}{\|\mathbf{p}\|_{(k)}+\varepsilon} \quad \forall\, k\in[m]\right\} &(128)\\ &= \min_{m\in\mathbb{N}}\left\{\log m \;:\; m \ge \frac{m}{\|\mathbf{p}\|_{(m)}+\varepsilon}\right\} &(129)\\ &= \min_{m\in\mathbb{N}}\left\{\log m \;:\; \|\mathbf{p}\|_{(m)} \ge 1-\varepsilon\right\}. &(130)\end{aligned}$$

Noticing that $\|\mathbf{p}\|_{(d)} = 1$ completes the proof.

The proof of Thm. 4 can be found in App. C.

Lemma 9: For any distribution $\mathbf{p}$ such that $V(\mathbf{p}) > 0$, any natural number n, and $\varepsilon \in [0, 1)$, let

$$\begin{aligned}f_{n,\varepsilon}(\mathbf{p}) &:= \min\left\{k : \|\mathbf{p}^{\otimes n}\|_{(k)} > \varepsilon\right\},\\ f'_{n,\varepsilon}(\mathbf{p}) &:= \min\left\{k : \|\mathbf{p}^{\otimes n}\|_{(k)} \ge \varepsilon\right\}. \end{aligned}\tag{131}$$

Then we have that

$$\begin{aligned}&\lim_{n\to\infty}\frac{\log f'_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}}\\ &= \lim_{n\to\infty}\frac{\log f_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}}\\ &= \Phi^{-1}(\varepsilon).\end{aligned}\tag{132}$$

Proof: According to [16, Lem. 15] (see also [50, Lem. 16]), it holds that for any distribution $\mathbf{p}$ such that $V(\mathbf{p}) > 0$,

$$\lim_{n\to\infty}\|\mathbf{p}^{\otimes n}\|_{(k_n(x))} = \lim_{n\to\infty}\sum_{i=1}^{k_n(x)}(\mathbf{p}^{\otimes n})_i^{\downarrow} = \Phi(x), \tag{133}$$

where

$$k_n(x) := \left\lfloor \exp\left(H(\mathbf{p}^{\otimes n}) + x\sqrt{V(\mathbf{p}^{\otimes n})}\right)\right\rfloor. \tag{134}$$

In the following, we will use this to prove the Lemma.

For any $\delta > 0$, let $x = \Phi^{-1}(\varepsilon + 2\delta)$. By Eq. (133), we have for sufficiently large n that

$$\|\mathbf{p}^{\otimes n}\|_{(k_n(x))} \ge \Phi(x) - \delta = \varepsilon + \delta > \varepsilon. \tag{135}$$

By the definition of $f_{n,\varepsilon}(\mathbf{p})$ and $k_n(x)$, this implies that

$$\log f_{n,\varepsilon}(\mathbf{p}) \le \log k_n(x) \le H(\mathbf{p}^{\otimes n}) + x\sqrt{V(\mathbf{p}^{\otimes n})}. \tag{136}$$

Taking $n \to \infty$, we get

$$\limsup_{n\to\infty}\frac{\log f_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}} \le x = \Phi^{-1}(\varepsilon + 2\delta). \tag{137}$$

Since the above inequality holds for any $\delta > 0$, we have

$$\limsup_{n\to\infty}\frac{\log f_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}} \le \Phi^{-1}(\varepsilon). \tag{138}$$

Next, we prove the other direction by contradiction. Suppose

$$\liminf_{n\to\infty}\frac{\log f_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}} < \Phi^{-1}(\varepsilon), \tag{139}$$

i.e., there exists a value r such that

$$\liminf_{n\to\infty}\frac{\log f_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}} \le r < \Phi^{-1}(\varepsilon). \tag{140}$$

For any

$$0 < \delta < \Phi^{-1}(\varepsilon) - r, \tag{141}$$

there thus exists a subsequence of n (denoted as n as well) such that

$$\frac{\log f_{n,\varepsilon}(\mathbf{p}) - nH(\mathbf{p})}{\sqrt{nV(\mathbf{p})}} \le r + \delta \tag{142}$$

for sufficiently large n. This is equivalent to

$$\begin{aligned}f_{n,\varepsilon}(\mathbf{p}) &\le \exp\left(nH(\mathbf{p}) + (r+\delta)\sqrt{nV(\mathbf{p})}\right)\\ &\le \left\lfloor \exp\left(nH(\mathbf{p}) + (r+\delta)\sqrt{nV(\mathbf{p})}\right)\right\rfloor + 1.\end{aligned}\tag{143}$$

Since $\|\mathbf{p}^{\otimes n}\|_{(k)}$ is non-decreasing in k, we have by the definition of k_n in Eq. (134) that

$$\begin{aligned}\|\mathbf{p}^{\otimes n}\|_{(f_{n,\varepsilon}(\mathbf{p}))} &\le \|\mathbf{p}^{\otimes n}\|_{(k_n(r+\delta)+1)}\\ &= \|\mathbf{p}^{\otimes n}\|_{(k_n(r+\delta))} + (\mathbf{p}^{\otimes n})^{\downarrow}_{(k_n(r+\delta)+1)}.\end{aligned}\tag{144}$$

Taking $n \to \infty$ on both sides and using Eq. (133) as well as Eq. (141), we have

$$\liminf_{n\to\infty}\|\mathbf{p}^{\otimes n}\|_{(f_{n,\varepsilon}(\mathbf{p}))} \le \Phi(r+\delta) < \varepsilon. \tag{145}$$

However, by the definition of $f_{n,\varepsilon}(\mathbf{p})$, we always have $\|\mathbf{p}^{\otimes n}\|_{(f_{n,\varepsilon}(\mathbf{p}))} > \varepsilon$, which forms a contradiction to Eq. (145). Since we are working in the asymptotic regime, the proof for $f'_{n,\varepsilon}(\mathbf{p})$ works exactly analogously.

Proposition 3: For any pure state $\psi \in \mathrm{Pure}(AB)$ with Schmidt vector $\mathbf{p}$, $V(\mathbf{p}) > 0$, and $\varepsilon \in [0, 1)$, it holds that

$$\mathrm{Cost}^{\varepsilon}(\psi^{\otimes n}) = nH(\mathbf{p}) - \Phi^{-1}(\varepsilon)\sqrt{nV(\mathbf{p})} + o(\sqrt{n}). \tag{146}$$

Proof: By Thm. 3, we have that

$$\mathrm{Cost}^{\varepsilon}(\psi^{\otimes n}) = \log f'_{n,1-\varepsilon}(\mathbf{p}). \tag{147}$$

The Proposition is thus simply a rewriting of Lem. 1 where we used that $\Phi^{-1}(1-\varepsilon) = -\Phi^{-1}(\varepsilon)$.

Proposition 4: For any pure state $\psi \in \mathrm{Pure}(AB)$ with Schmidt vector $\mathbf{p}$, $V(\mathbf{p}) > 0$, and $\varepsilon \in [0, 1)$, it holds that

$$\mathrm{Distill}^{\varepsilon}(\psi^{\otimes n}) = nH(\mathbf{p}) + \Phi^{-1}(\varepsilon)\sqrt{nV(\mathbf{p})} + o(\sqrt{n}). \tag{148}$$

Proof: For any $\delta > 0$, let $x = \Phi^{-1}(\varepsilon + 2\delta)$. Using Eq. (133), for any sufficiently large n, we have $\|\mathbf{p}^{\otimes n}\|_{(k_n(x))} \ge \Phi(x) - \delta = \varepsilon + \delta > \varepsilon$ where $k_n(x)$ is defined in Eq. (134). Due to Thm. 2, we find that

$$\begin{aligned}\mathrm{Distill}^{\varepsilon}(\psi^{\otimes n}) &\le \log\left\lfloor\frac{k_n(x)}{\|\mathbf{p}^{\otimes n}\|_{(k_n(x))} - \varepsilon}\right\rfloor\\ &\le \log\frac{k_n(x)}{\|\mathbf{p}^{\otimes n}\|_{(k_n(x))} - \varepsilon}\\ &= \log k_n(x) - \log(\|\mathbf{p}^{\otimes n}\|_{(k_n(x))} - \varepsilon)\\ &\le nH(\mathbf{p}) + x\sqrt{nV(\mathbf{p})} - \log\left(\Phi(x) - \delta - \varepsilon\right)\\ &= nH(\mathbf{p}) + x\sqrt{nV(\mathbf{p})} - \log(\delta).\end{aligned}\tag{149}$$

Note that $\Phi^{-1}(\cdot)$ is continuously differentiable and thus $x = \Phi^{-1}(\varepsilon + 2\delta) = \Phi^{-1}(\varepsilon) + O(\delta)$. Considering $\delta = 1/n$, we have

$$\text{Distill}^{\varepsilon}(\psi^{\otimes n}) \leq nH(\mathbf{p}) + \Phi^{-1}(\varepsilon)\sqrt{nV(\mathbf{p})} + o(\sqrt{n}). \quad (150)$$

Next, we prove the converse direction. For n large enough, we have

$$\begin{aligned}\text{Distill}^{\varepsilon}(\psi^{\otimes n}) &= \min_{k\in\{\ell,\dots,d\}} \log\left\lfloor \frac{k}{\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon}\right\rfloor \\ &\geq \min_{k\in\{\ell,\dots,d\}} \left[\log\left(\frac{k}{\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon}\right)\right] - 1 \\ &\geq \min_{k\in\{\ell,\dots,d\}} \left[\log\left(\frac{k}{1-\varepsilon}\right)\right] - 1 \\ &= \log f_{n,\varepsilon}(\mathbf{p}) - \log(1-\varepsilon) - 1 \\ &\geq nH(\mathbf{p}) + \Phi^{-1}(\varepsilon)\sqrt{nV(\mathbf{p})} + o(\sqrt{n}), \quad (151)\end{aligned}$$

where the first inequality follows from $\log\lfloor x\rfloor \geq (\log x) - 1$, the second from $\|\mathbf{p}\|_{(k)} \leq 1$, and the last from Lem. 1. This completes the proof.

Lemma 10: Let $\psi, \phi \in \text{Pure}(AB)$ with Schmidt coefficients $\mathbf{p}$ and $\mathbf{q}$, respecitvely. Then

$$T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \phi^{AB}\right) = \begin{cases} 0 \text{if } H(\mathbf{q}) \leq H(\mathbf{p}), \\ \varepsilon : H(\bar{\mathbf{q}}^{(\varepsilon)}) = H(\mathbf{p}) \text{else}. \end{cases}$$

Proof: The first case is obvious from Eq. (19). If $H(\mathbf{q}) > H(\mathbf{p})$, let $\mathbf{r}^\star$ be a minimizer in Eq. (20) and let

$$\varepsilon = T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \phi^{AB}\right). \quad (152)$$

This implies that $\mathbf{r}^\star \in \mathfrak{B}_\varepsilon(\mathbf{q})$, and thus $\bar{\mathbf{q}}^{(\varepsilon)} \succ \mathbf{r}^\star$ and $H(\bar{\mathbf{q}}^{(\varepsilon)}) \leq H(\mathbf{r}^\star)$ (because the Shannon entropy is Schur-concave). It follows that $\bar{\mathbf{q}}^{(\varepsilon)}$ is a minimizer too (by definition, $\frac{1}{2}\left\|\bar{\mathbf{q}}^{(\varepsilon)} - \mathbf{q}\right\|_1 \leq \varepsilon$), and we can thus restrict the optimization without loss of generality to steepest ε-approximations, i.e.,

$$\begin{aligned}&T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \phi^{AB}\right) \\ &= \min\left\{\varepsilon \in [0,1] : H(\bar{\mathbf{q}}^{(\varepsilon)}) \leq H(\mathbf{p})\right\}. \quad (153)\end{aligned}$$

However, for $\tilde{\varepsilon} \geq \varepsilon$, it follows that $\bar{\mathbf{q}}^{(\varepsilon)} \in \mathfrak{B}_{\tilde{\varepsilon}}(\mathbf{q})$ and thus $\bar{\mathbf{q}}^{(\tilde{\varepsilon})} \succeq \bar{\mathbf{q}}^{(\varepsilon)}$ and $H(\bar{\mathbf{q}}^{(\tilde{\varepsilon})}) \leq H(\bar{\mathbf{q}}^{(\varepsilon)})$. $H(\bar{\mathbf{q}}^{(\varepsilon)})$ is thus monotonically decreasing in ε, and by the construction of $\bar{\mathbf{q}}^{(\varepsilon)}$, also continuous in ε. Noting that $H(\bar{\mathbf{q}}^{(1)}) = 0 \leq H(\mathbf{p})$ finishes the proof.

Theorem 5: Denote by $\mathbf{p}$ the vector containing the Schmidt coefficients of $\psi \in \text{Pure}(AB)$. Then

$$\text{Cost}_c^{\varepsilon}\left(\psi^{AB}\right) = \log\left\lceil 2^{H^{\varepsilon}(\mathbf{p})}\right\rceil, \quad (154)$$

where

$$H^{\varepsilon}(\mathbf{p}) = \min_{\mathbf{r}\in\mathfrak{B}_\varepsilon(\mathbf{p})} H(\mathbf{r}) = H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right) \quad (155)$$

is the ε-smoothed Shannon entropy and $\bar{\mathbf{p}}^{(\varepsilon)}$ denotes the steepest ε-approximation of $\mathbf{p}$ [46].

Proof: We first note that $\log\left\lceil 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rceil$ is an upper bound on the cost, since for $m = \left\lceil 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rceil$, $T_\star\left(\Phi_m \xrightarrow{\text{CLOCC}} \psi^{AB}\right) = 0$ if $\log m \geq H(\mathbf{p})$ (see Lem. 2), and if not, $T_\star\left(\Phi_m \xrightarrow{\text{CLOCC}} \psi^{AB}\right) = \delta$, where δ is such that

$$H\left(\bar{\mathbf{p}}^{(\delta)}\right) = \log\left\lceil 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rceil \geq H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right) \quad (156)$$

and thus $\delta \leq \varepsilon$ (see the proof of Lem. 2).

To show that this is not only an upper bound, we first note that $H\left(\bar{\mathbf{p}}^{(\delta)}\right) < H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right)$ implies $\delta > \varepsilon$. In the proof of Lem. 2, we argued that $H\left(\bar{\mathbf{p}}^{(\delta)}\right) < H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right)$ implies $\delta \geq \varepsilon$. However, we cannot have $\delta = \varepsilon$, because then we find $H\left(\bar{\mathbf{p}}^{(\delta)}\right) = H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right)$.

Now assume that we have an integer m with $m < \left\lceil 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rceil$ and consider two cases.

i) Case $2^{H(\bar{\mathbf{p}}^{(\varepsilon)})} \in \mathbb{N}$. This implies that

$$\log m < H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right) \leq H(\mathbf{p}) \quad (157)$$

and according to Lem. 2, $T_\star\left(\Phi_m \xrightarrow{\text{CLOCC}} \psi^{AB}\right)$ is given by δ such that

$$H\left(\bar{\mathbf{p}}^{(\delta)}\right) = \log m < H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right) \quad (158)$$

and thus $\delta > \varepsilon$.

ii) Case $2^{H(\bar{\mathbf{p}}^{(\varepsilon)})} \notin \mathbb{N}$. The assumption $m < \left\lceil 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rceil$ then implies that $m \leq \left\lfloor 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rfloor$ and therefore

$$\log m \leq \log\left\lfloor 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rfloor < \log\left(2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right) \leq H(\mathbf{p}) \quad (159)$$

and according to Lem. 2, $T_\star\left(\Phi_m \xrightarrow{\text{CLOCC}} \psi^{AB}\right)$ is again given by δ such that

$$H\left(\bar{\mathbf{p}}^{(\delta)}\right) = \log m < H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right) \quad (160)$$

and thus $\delta > \varepsilon$.

In summary, we showed that for $m < \left\lceil 2^{H(\bar{\mathbf{p}}^{(\varepsilon)})}\right\rceil$, it follows that $T_\star\left(\Phi_m \xrightarrow{\text{CLOCC}} \psi^{AB}\right) > \varepsilon$, which finishes the first part of the proof. To conclude, we note that

$$H\left(\bar{\mathbf{p}}^{(\varepsilon)}\right) = H^{\varepsilon}(\mathbf{p}) \quad (161)$$

because of Schur-convexity (see [46, Prop. 3]).

Theorem 6: Denote by $\mathbf{p}$ the vector containing the Schmidt coefficients of $\psi \in \text{Pure}(AB)$. For $\varepsilon \in [0,1)$,

$$\text{Distill}_c^{\varepsilon}\left(\psi^{AB}\right) = \max_{m\in\mathbb{N}}\left\{\log m : H^{\varepsilon}(\mathbf{u}_m) \leq H(\mathbf{p})\right\},$$

and the optimization over m can be restricted to

$$\left\lfloor 2^{H(\mathbf{p})}\right\rfloor \leq m \leq \left\lfloor 2^{\frac{H(\mathbf{p})+h(\varepsilon)}{1-\varepsilon}}\right\rfloor, \quad (162)$$

where $h(x) = -x\log(x) - (1-x)\log(1-x)$ is the binary entropy.

Proof: Using Lem. 2, we find that

$$T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \Phi_m^{AB}\right) = \begin{cases} 0 \text{if } \log m \leq H(\mathbf{p}), \\ \delta : H\left(\bar{\mathbf{u}}_m^{(\delta)}\right) = H(\mathbf{p}) \text{else}. \end{cases}$$

Let $m \in \mathbb{N}$ be such that $H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right) \leq H(\mathbf{p})$. If $\log m \leq H(\mathbf{p})$, $T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \Phi_m^{AB}\right) = 0$, if not, $T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \Phi_m^{AB}\right) = \delta$ with δ such that

$$H\left(\bar{\mathbf{u}}_m^{(\delta)}\right) = H(\mathbf{p}) \geq H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right) \quad (163)$$

and thus $\delta \leq \varepsilon$. In either case, $T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \Phi_m^{AB}\right) \leq \varepsilon$ and such m can thus be achieved. If $m \in \mathbb{N}$ is however such that $H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right) > H(\mathbf{p})$, we have

$$\log m = H(\mathbf{u}_m) \geq H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right) > H(\mathbf{p}) \tag{164}$$

and thus $T_\star\left(\psi^{AB} \xrightarrow{\text{CLOCC}} \Phi_m^{AB}\right) = \delta$ with δ such that

$$H\left(\bar{\mathbf{u}}_m^{(\delta)}\right) = H(\mathbf{p}) < H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right) \tag{165}$$

and therefore $\delta > \varepsilon$, which shows that such m cannot be achieved.

For $\varepsilon = 0$, we therefore get that $\text{Distill}_c^0\left(\psi^{AB}\right) = \log\left\lfloor 2^{H(\mathbf{p})}\right\rfloor$. Since $\text{Distill}_c^0\left(\psi^{AB}\right) \leq \text{Distill}_c^\varepsilon\left(\psi^{AB}\right)$, $\left\lfloor 2^{H(\mathbf{p})}\right\rfloor$ is a lower bound on the m over which we need to optimize. If m is an integer such that $m > \left\lfloor 2^{\frac{H(p)+h(\varepsilon)}{1-\varepsilon}}\right\rfloor$, this implies that $m > 2^{\frac{H(p)+h(\varepsilon)}{1-\varepsilon}}$ and in combination with (see Refs. [81], [82])

$$\begin{aligned}\log m &- H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right)\\ &= \left|H(\mathbf{u}_m) - H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right)\right|\\ &\leq h(\varepsilon) + \varepsilon \log m\end{aligned} \tag{166}$$

we find

$$H^\varepsilon(\mathbf{u}_m) = H\left(\bar{\mathbf{u}}_m^{(\varepsilon)}\right) \geq (1-\varepsilon)\log m - h(\varepsilon) > H(\mathbf{p}). \tag{167}$$

Lemma 11: Let $\rho \in \mathfrak{D}(XA)$ be a classical-quantum-state as in Eq. (47), and for any $m \in [|A|]$, let $\rho^{(m)}$ be as defined in Eq. (48). Then, for any $\varepsilon \in [0,1]$,

$$\begin{aligned}&H_{\max}^\varepsilon(A|X)_\rho\\ &= \min_{m\in[|A|]}\left\{\log m : \frac{1}{2}\left\|\rho^{(m)} - \rho^{XA}\right\|_1 \leq \varepsilon\right\}\\ &= \min_{m\in[|A|]}\left\{\log m : \sum_{x\in[k]} p_x \left\|\rho_x^A\right\|_{(m)} \geq 1-\varepsilon\right\}.\end{aligned} \tag{168}$$

Proof: If $\rho^{(m)} \neq \rho^{XA}$, there exists an $x \in [|A|]$ such that $p_x \neq 0$ and the rank of $\rho_x^{(m)}$ equals m, from which follows that (see Eq. (43))

$$H_{\max}(A|X)_{\rho^{(m)}} = \log m. \tag{169}$$

Using the definition of $H_{\max}^\varepsilon(A|X)_\rho$ (see Eq. (40)), we thus obtain

$$\begin{aligned}&H_{\max}^\varepsilon(A|X)_\rho\\ &= \min_{\omega\in\mathfrak{D}(XA)}\left\{H_{\max}(A|X)_\omega : \frac{1}{2}\left\|\omega^{XA} - \rho^{XA}\right\|_1 \leq \varepsilon\right\}\\ &\leq \min_{m\in[|A|]}\left\{H_{\max}(A|X)_{\rho^{(m)}} : \frac{1}{2}\left\|\rho^{(m)} - \rho^{XA}\right\|_1 \leq \varepsilon\right\}\\ &= \min_{m\in[|A|]}\left\{\log m : \frac{1}{2}\left\|\rho^{(m)} - \rho^{XA}\right\|_1 \leq \varepsilon\right\}.\end{aligned} \tag{170}$$

To show that indeed we have an equality in the above equation, we notice that by definition, $\rho_x^{(m)}$ and ρ_x^A commute. Denoting the eigenvalues of ρ_x^A by $\{\lambda_{y|x}\}$ we get that the only potentially non-zero eigenvalues of $\rho_x^{(m)}$ are

$$\left\{\lambda_{y|x}^\downarrow / \left\|\rho_x^A\right\|_{(m)}\right\}_{y\in[m]}. \tag{171}$$

From this follows that

$$\begin{aligned}&\frac{1}{2}\left\|\rho_x^A - \rho_x^{(m)}\right\|_1\\ &= \frac{1}{2}\left[\sum_{y=1}^{m}\left|\lambda_{y|x}^\downarrow - \lambda_{y|x}^\downarrow / \left\|\rho_x^A\right\|_{(m)}\right| + \sum_{y=m+1}^{|A|}\lambda_{y|x}^\downarrow\right]\\ &= \frac{1}{2}\left[\left\|\rho_x^A\right\|_{(m)}\left(1/\left\|\rho_x^A\right\|_{(m)} - 1\right) + 1 - \left\|\rho_x^A\right\|_{(m)}\right]\\ &= 1 - \left\|\rho_x^A\right\|_{(m)}\end{aligned} \tag{172}$$

and thus

$$\frac{1}{2}\left\|\rho^{(m)} - \rho^{XA}\right\|_1 = 1 - \sum_{x\in[k]} p_x \left\|\rho_x^A\right\|_{(m)}. \tag{173}$$

This proves that the two optimization problems in the Lemma have the same solution, i.e.,

$$\begin{aligned}&\min_{m\in[|A|]}\left\{\log m : \frac{1}{2}\left\|\rho^{(m)} - \rho^{XA}\right\|_1 \leq \varepsilon\right\}\\ &= \min_{m\in[|A|]}\left\{\log m : \sum_{x\in[k]} p_x \left\|\rho_x^A\right\|_{(m)} \geq 1-\varepsilon\right\}.\end{aligned} \tag{174}$$

Now suppose by contradiction that

$$H_{\max}^\varepsilon(A|X)_\rho < \min_{m\in[|A|]}\left\{\log m : \sum_{x\in[k]} p_x \left\|\rho_x^A\right\|_{(m)} \geq 1-\varepsilon\right\} \tag{175}$$

and let m be the smallest natural number such that $\sum_x p_x \left\|\rho_x^A\right\|_{(m)} \geq 1-\varepsilon$. Further, let

$$\omega^{XA} = \sum_x q_x |x\rangle\langle x|^X \otimes \omega_x^A \in \mathfrak{B}_\varepsilon\left(\rho^{XA}\right) \tag{176}$$

be such that $H_{\max}^\varepsilon(A|X)_\rho = H_{\max}(A|X)_\omega$. Then, according to the assumption,

$$\log m > H_{\max}(A|X)_\omega = \log \max_{x\in[k]:q_x\neq 0} \operatorname{Tr}\left[\Pi_{\omega_x}^A\right] \tag{177}$$

so that $\operatorname{Tr}\left[\Pi_{\omega_x}^A\right] < m$ for all $x \in [k] : q_x \neq 0$. In particular, this means that for any $x \in [k] : q_x \neq 0$, $\left\|\rho_x^A\right\|_{(m-1)} \geq \operatorname{Tr}\left[\rho_x^A \Pi_{\omega_x}^A\right]$ since $\Pi_{\omega_x}^A$ has a rank strictly smaller than m. Therefore, with

$$\Pi_\omega^{XA} := \sum_{x\in[k]:q_x\neq 0} |x\rangle\langle x|^X \otimes \Pi_{\omega_x}^A, \tag{178}$$

we have

$$\begin{aligned}&\sum_x p_x \left\|\rho_x^A\right\|_{(m-1)}\\ &\quad\geq \operatorname{Tr}\left[\rho^{XA}\Pi_\omega^{XA}\right]\\ &\quad= \operatorname{Tr}\left[\omega^{XA}\Pi_\omega^{XA}\right] + \operatorname{Tr}\left[\left(\rho^{XA} - \omega^{XA}\right)\Pi_\omega^{XA}\right]\\ &\quad\geq 1 - \operatorname{Tr}\left[\left(\rho^{XA} - \omega^{XA}\right)_-\Pi_\omega^{XA}\right]\\ &\quad\geq 1-\varepsilon,\end{aligned} \tag{179}$$

where we used the fact that $\Pi_\omega^{XA} \leq I^{XA}$ and $\operatorname{Tr}\left(\rho^{XA} - \omega^{XA}\right)_- = \frac{1}{2}\|\omega^{XA} - \rho^{XA}\|_1 \leq \varepsilon$. This is in contradiction to the definition of m as the smallest natural number such that $\sum_x p_x \left\|\rho_x^A\right\|_{(m)} \geq 1-\varepsilon$, which completes the proof.

Theorem 7: For $\rho \in \mathfrak{D}(AB)$, the ε-single-shot entanglement cost is given by

$$\mathrm{Cost}^{\varepsilon}\left(\rho^{AB}\right) = \inf_{\rho^{XAB}} H^{\varepsilon}_{\max}(A|X)_{\rho}\,, \tag{180}$$

where the infimum is over all classical systems X and all classical extensions ρ^{XAB} of ρ^{AB}. Moreover, the infimum is attained for a classical extension with $|X| = |AB|^2$ and can also be taken over all regular extensions of ρ^{AB}.

Proof: We begin by showing a useful mathematical identity. For $\varepsilon \in [0, 1]$, let

$$I := \min_{m\in[|A|]} \left\{\log m : \max_{\rho^{XAB}} \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon\right\},$$
$$J := \min_{\rho^{XAB}} \min_{m\in[|A|]} \left\{\log m : \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon\right\}, \tag{181}$$

where the optimizations over ρ^{XAB} run over all regular extensions ρ^{XAB} of ρ^{AB} with $|X| = |AB|^2$. Using the same convention, let

$$\mathfrak{Y} := \left\{\log m : m \in [|A|], \max_{\rho^{XAB}} \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon\right\},$$
$$\mathfrak{Y}_{\rho} := \left\{\log m : m \in [|A|], \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon\right\}. \tag{182}$$

Since $\max_{\rho^{XAB}} \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon$, it follows that $\mathfrak{Y}_{\rho} \subset \mathfrak{Y}$ and thus $\min \mathfrak{Y} \leq \min \mathfrak{Y}_{\rho}$ for all regular extensions ρ^{XAB}. We therefore find that

$$I = \min \mathfrak{Y} \leq \min_{\rho^{XAB}} \min \mathfrak{Y}_{\rho} = J. \tag{183}$$

From the definition of $I =: \log m^{\star}$, it further follows that there exists a regular extension $\tilde{\rho}^{XAB}$ of ρ^{AB} such that

$$\sum_x \tilde{p}_x \left\|\tilde{\rho}^A_x\right\|_{(m^{\star})} \geq 1-\varepsilon. \tag{184}$$

Now assume that $I < J$, i.e.,

$$\begin{aligned}\log m^{\star} &< \min_{\rho^{XAB}} \min_{m\in[|A|]} \left\{\log m : \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon\right\}\\ &\leq \min_{m\in[|A|]} \left\{\log m : \sum_x \tilde{p}_x \left\|\tilde{\rho}^A_x\right\|_{(m)} \geq 1-\varepsilon\right\}\\ &\leq \log m^{\star},\end{aligned} \tag{185}$$

which is a contradiction. We thus showed that $I = J$.

Next, we remember that according to Eq. (37)

$$P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \rho^{AB}\right) = E_{(m)}\left(\rho^{AB}\right). \tag{186}$$

Now recall that $\Phi_m \xrightarrow{\text{LOCC}} \rho^{AB}$ if and only if ρ^{AB} can be decomposed into an ensemble of pure states with Schmidt rank no larger than m (sufficiency follows from Eq. (36), necessity from the fact that the Schmidt number cannot increase [49], [56], [83]). Moreover, according to Carathéodory's theorem, one can restrict consideration to ensembles consisting of at most $|AB|^2$ pure states. Following along the proof of Eq. (37) in [57, Sec. B4], it then becomes clear that the infimums in the definition of $E_{(m)}$ (see Eq. (35)) are achieved for a decomposition into at most $|AB|^2$ pure states and we thus find that

$$\begin{aligned}&P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \rho^{AB}\right)\\ &= E_{(m)}\left(\rho^{AB}\right) = \min_{\rho^{XAB}}\left(1-\sum_x p_x \left\|\rho^A_x\right\|_{(m)}\right)\\ &= 1-\max_{\rho^{XAB}} \sum_x p_x \left\|\rho^A_x\right\|_{(m)},\end{aligned} \tag{187}$$

where the optimizations are over all classical systems X and all regular extensions ρ^{XAB} of ρ^{AB}, with the optimal values attained for $|X| = |AB|^2$. The ε-single-shot entanglement cost as defined in Eq. (33) can thus be expressed as

$$\begin{aligned}&\mathrm{Cost}^{\varepsilon}\left(\rho^{AB}\right)\\ &= \min_{m\in[|A|]}\left\{\log m : \max_{\rho^{XAB}} \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon\right\}\\ &= \min_{\rho^{XAB}} \min_{m\in[|A|]}\left\{\log m : \sum_x p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon\right\}\\ &= \min_{\rho^{XAB}} H^{\varepsilon}_{\max}(A|X)_{\rho},\end{aligned} \tag{188}$$

where we used Lem. 3 in the last line.

To conclude the proof, we show that the infimum can also be taken over all classical extensions. Since the set of regular extensions ρ^{XAB} is a subset of the set of classical extensions $\tilde{\rho}^{XAB}$,

$$\inf_{\rho^{XAB}} H^{\varepsilon}_{\max}(A|X)_{\rho} \geq \inf_{\tilde{\rho}^{XAB}} H^{\varepsilon}_{\max}(A|X)_{\tilde{\rho}}. \tag{189}$$

The proof of the converse is very similar to the proof of Lem. 3: Let ρ^{XAB} be a *regular* extension of ρ^{AB} satisfying $H^{\varepsilon}_{\max}(A|X)_{\rho} = \inf_{\rho^{XAB}} H^{\varepsilon}_{\max}(A|X)_{\rho}$ (which exists according to the previous parts of the proof) and let m be the smallest integer such that $\sum_{x\in[|A|]} p_x \left\|\rho^A_x\right\|_{(m)} \geq 1-\varepsilon$. Now assume by contradiction that there exists a *classical* extension $\tilde{\rho}^{XAB}$ of ρ^{AB} such that $H^{\varepsilon}_{\max}(A|X)_{\tilde{\rho}} < H^{\varepsilon}_{\max}(A|X)_{\rho} = \log m$ (see Lem. 3). Denoting by

$$\omega^{XA} = \sum_x q_x |x\rangle\langle x|^X \otimes \omega^A_x \in \mathfrak{B}_{\varepsilon}\left(\tilde{\rho}^{XA}\right) \tag{190}$$

a state satisfying $H^{\varepsilon}_{\max}(A|X)_{\tilde{\rho}} = H_{\max}(A|X)_{\omega}$, we find that

$$\log m > H_{\max}(A|X)_{\omega} = \log \max_{x\in[|A|]:q_x\neq 0} \mathrm{Tr}\left[\Pi^A_{\omega_x}\right] \tag{191}$$

and thus $m > \mathrm{Tr}\left[\Pi^A_{\omega_x}\right] \forall x \in [|A|] : q_x \neq 0$. With Π^{XA}_{ω} defined as in Eq. (178), this implies again that

$$\begin{aligned}&\sum_x p_x \left\|\rho^A_x\right\|_{(m-1)}\\ &\geq \mathrm{Tr}\left[\rho^{XA}\Pi^{XA}_{\omega}\right]\\ &= \mathrm{Tr}\left[\omega^{XA}\Pi^{XA}_{\omega}\right] + \mathrm{Tr}\left[\left(\rho^{XA}-\omega^{XA}\right)\Pi^{XA}_{\omega}\right]\\ &\geq 1-\mathrm{Tr}\left[\left(\rho^{XA}-\omega^{XA}\right)_{-}\Pi^{XA}_{\omega}\right]\\ &\geq 1-\varepsilon\,,\end{aligned} \tag{192}$$

which is a contradiction to the assumption.

The proofs of Cor. 1, Cor. 2, and Cor. 3 were provided in the main text.

Lemma 12: Let $\rho \in \mathcal{D}(A)$ with $|A| = d$ and let $\mathbf{p} \in \text{Prob}(d)$ contain the eigenvalues of ρ. Let further $\mathbf{u}^{(d)}$ be the flat distribution of dimension d. If $\frac{1}{2}\left\|\mathbf{p} - \mathbf{u}^{(d)}\right\|_1 \le \varepsilon$, then

$$H_{\min}^{\varepsilon}(\rho^A) = \log d, \tag{193}$$

otherwise

$$H_{\min}^{\varepsilon}(\rho^A) = \log \min_{\ell \in [d]} \left\{ \frac{\ell}{\|\mathbf{p}\|_{(\ell)} - \varepsilon} \right\}. \tag{194}$$

Proof: Let $\omega \in \mathfrak{B}_\varepsilon(\rho^A)$ be optimal in the sense that $H_{\min}(\omega^A) = H_{\min}^{\varepsilon}(\rho^A)$. We first show that without loss of generality, we can assume that ω and ρ commute. To this end, let $\Delta \in \text{CPTP}(A \to A)$ be the completely dephasing channel in the eigenbasis of ρ. This implies that $\Delta(\rho) = \rho$ and thus

$$\frac{1}{2}\|\Delta(\omega) - \rho\|_1 = \frac{1}{2}\|\Delta(\omega) - \Delta(\rho)\|_1 \le \frac{1}{2}\|\omega - \rho\|_1 \le \varepsilon, \tag{195}$$

where we made use of the data-processing inequality and the choice of ω. Hence, $\Delta(\omega)$ is also in $\mathfrak{B}_\varepsilon(\rho^A)$. Moreover, since Δ is unital and the min-entropy is non-decreasing under such channels [18],

$$H_{\min}(\Delta(\omega)) \ge H_{\min}(\omega). \tag{196}$$

This implies that also $\Delta(\omega)$ is optimal. Moreover, by construction, ρ and $\Delta(\omega)$ commute, which proofs our claim.

Next, we notice that this allows us to reduce our analysis to probability vectors: First, we extend our definitions in the usual manner, i.e., for $\mathbf{p}, \mathbf{q} \in \text{Prob}(d)$, let

$$H_{\min}(\mathbf{p}) = -\log \|\mathbf{p}\|_\infty, \tag{197}$$

where $\|\mathbf{p}\|_\infty$ denotes the largest entry of $\mathbf{p}$. Denoting by $\mathbf{p}$ the probability vector containing the eigenvalues of ρ, we can thus express $H_{\min}^{\varepsilon}(\rho^A)$ as

$$H_{\min}^{\varepsilon}(\rho^A) = \max_{\mathbf{q} \in \mathfrak{B}_\varepsilon(\mathbf{p})} H_{\min}(\mathbf{q}). \tag{198}$$

This allows us to use a result from approximate majorization presented in [46] concerning the existence of the so-called flattest ε-approximation: There exists a vector $\underline{\mathbf{p}}^{(\varepsilon)} \in \mathfrak{B}_\varepsilon(\mathbf{p})$ that is minimal with respect to the majorization relation in the sense that $\mathbf{q} \succ \underline{\mathbf{p}}^{(\varepsilon)}$ for all $\mathbf{q} \in \mathfrak{B}_\varepsilon(\mathbf{p})$. Since $\mathbf{r} \succ \mathbf{s}$ implies that

$$H_{\min}(\mathbf{r}) \le H_{\min}(\mathbf{s}), \tag{199}$$

we obtain that

$$H_{\min}^{\varepsilon}(\rho) = H_{\min}(\underline{\mathbf{p}}^{(\varepsilon)}) = -\log \|\underline{\mathbf{p}}^{(\varepsilon)}\|_\infty. \tag{200}$$

From now on, we assume without loss of generality that $\mathbf{p} = \mathbf{p}^{\downarrow}$.

First, we notice that for $\varepsilon = 0$, the Lemma is trivially true, since

$$H_{\min}^{0}(\rho) = -\log \|\rho\|_\infty = -\log p_1 \tag{201}$$

and

$$\frac{\|\rho\|_{(y)}}{y} = \frac{\sum_{x=1}^{y} p_x}{y} \le \frac{y p_1}{y} = p_1 = \frac{\|\rho\|_{(1)}}{1}. \tag{202}$$

We will thus assume from now on that $\varepsilon > 0$.

If

$$\varepsilon \ge \frac{1}{2}\|\mathbf{p} - \mathbf{u}^{(d)}\|_1, \tag{203}$$

$\underline{\mathbf{p}}^{(\varepsilon)}$ is given by $\mathbf{u}^{(d)}$, since the flat distribution is majorized by all other distributions of equal dimension. This implies that

$$H_{\min}^{\varepsilon}(\rho) = \log d. \tag{204}$$

If

$$\varepsilon < \frac{1}{2}\|\mathbf{p} - \mathbf{u}^{(d)}\|_1, \tag{205}$$

there exist unique $a, b \in [0, 1]$ and $k, l \in [d]$ such that [46, Eq. (9)]

$$\underline{p}_x^{(\varepsilon)} = \begin{cases} a & \text{if } x \in [k], \\ p_x & \text{if } k < x < l, \\ b & \text{if } x \in \{l, \dots, d\} \end{cases} \tag{206}$$

and

$$\underline{\mathbf{p}}_x^{(\varepsilon)} = \left(\underline{\mathbf{p}}_x^{(\varepsilon)}\right)^{\downarrow}. \tag{207}$$

Using in addition that $\underline{\mathbf{p}}_x^{(\varepsilon)}$ corresponds to a flattening of $\mathbf{p}$ (see [46, Fig. 1]), we conclude that

$$p_{k+1} < a \le p_k. \tag{208}$$

Moreover, according to [46, Eq. (7) and comment below]) and taking into account the factor of 2 in the definition of the error,

$$\varepsilon = \|\mathbf{p}\|_{(k)} - ka. \tag{209}$$

In combination with Eq. (208), this implies that

$$\|\mathbf{p}\|_{(k)} - kp_k \le \varepsilon < \|\mathbf{p}\|_{(k)} - kp_{k+1}. \tag{210}$$

Now for $y \in [d]$, let

$$t_y := \frac{\|\mathbf{p}\|_{(y)} - \varepsilon}{y} \tag{211}$$

and define l as the largest integer in $[d]$ such that

$$t_l = \max_{y \in [d]} t_y. \tag{212}$$

We will now consider three different cases:

(i) Case $1 < l < d$: In this case, by definition of l, we have

$$\begin{aligned} 0 < t_l - t_{l+1} &= \frac{\|\mathbf{p}\|_{(l)} - lp_{l+1} - \varepsilon}{l(l+1)} \\ 0 \le t_l - t_{l-1} &= \frac{lp_l - \|\mathbf{p}\|_{(l)} + \varepsilon}{l(l-1)} \end{aligned} \tag{213}$$

and thus

$$\|\mathbf{p}\|_{(l)} - lp_l \le \varepsilon < \|\mathbf{p}\|_{(l)} - lp_{l+1}. \tag{214}$$

Comparing with Eq. (210), we thus find that $l = k$.

(ii) Case $l = 1$: This implies that

$$\|\mathbf{p}\|_{(1)} - \varepsilon = t_1 > t_2 = \frac{\|\mathbf{p}\|_{(2)} - \varepsilon}{2}, \tag{215}$$

which is equivalent to

$$\varepsilon < p_1 - p_2. \tag{216}$$

Comparing with Eq. (210), we find again that $l = k$.

(iii) Case $l = d$: Here, we find that

$$\frac{1 - \varepsilon}{d} = t_d \ge t_{d-1} = \frac{1 - p_d - \varepsilon}{d - 1} \tag{217}$$

and thus

$$\varepsilon \ge \|\mathbf{p}\|_{(d)} - dp_d. \tag{218}$$

Comparing with Eq. (210), the only possibility is again $l = k$.

Due to Eq. (209), we finally conclude that

$$a = t_k = \max_{y\in[d]} \frac{\|\mathbf{p}\|_{(y)} - \varepsilon}{y} \tag{219}$$

and thus (see Eq. (200)),

$$\begin{aligned} H^{\varepsilon}_{\min}(\rho) &= -\log \max_{y\in[d]} \left\{ \frac{\|\mathbf{p}\|_{(y)} - \varepsilon}{y} \right\} \\ &= \log \min_{y\in[d]} \left\{ \frac{y}{\|\mathbf{p}\|_{(y)} - \varepsilon} \right\}. \end{aligned} \tag{220}$$

Lemma 13: Let $\mathbf{q} \in \mathrm{Prob}(n)$ and $\mathbf{q} \neq \mathbf{u}^{(n)}$. Then

$$\left\| \mathbf{q} - \mathbf{u}^{(n)} \right\|_1 = 2 \max_{k\in[n]} \left(\|\mathbf{q}\|_{(k)} - \frac{k}{n} \right).$$

Proof: Assume without loss of generality that $\mathbf{q} = \mathbf{q}^{\downarrow}$ and let $z \in [n-1]$ be such that $q_z > \frac{1}{n}$ and $q_{z+1} \leq \frac{1}{n}$. Observe that

$$0 = \sum_{x=1}^{n} \left(q_x - \frac{1}{n} \right) = \sum_{x=1}^{z} \left(q_x - \frac{1}{n} \right) + \sum_{x=z+1}^{n} \left(q_x - \frac{1}{n} \right) \tag{221}$$

and thus

$$\sum_{x=1}^{z} \left(q_x - \frac{1}{n} \right) = \sum_{x=z+1}^{n} \left(\frac{1}{n} - q_x \right). \tag{222}$$

Moreover,

$$\max_{k\in[n]} \sum_{x=1}^{k} \left(q_x - \frac{1}{n} \right) = \sum_{x=1}^{z} \left(q_x - \frac{1}{n} \right) \tag{223}$$

and thus

$$\begin{aligned} \left\| \mathbf{q} - \mathbf{u}^{(n)} \right\|_1 &= \sum_{x=1}^{n} \left| q_x - \frac{1}{n} \right| \\ &= \sum_{x=1}^{z} \left(q_x - \frac{1}{n} \right) + \sum_{x=z+1}^{n} \left(\frac{1}{n} - q_x \right) \\ &= 2 \max_{k\in[n]} \sum_{x=1}^{k} \left(q_x - \frac{1}{n} \right) \\ &= 2 \max_{k\in[n]} \left(\|\mathbf{q}\|_{(k)} - \frac{k}{n} \right). \end{aligned} \tag{224}$$

Corollary 5: Let $\rho \in \mathcal{D}(A)$ with $|A| = d$, $\varepsilon \in [0, 1)$, and let $\mathbf{p} \in \mathrm{Prob}(d)$ contain the eigenvalues of ρ. It then holds that

$$\tilde{H}^{\varepsilon}_{\min}(\rho^A) = \log \min_{\ell\in[d]} \left\{ \frac{\ell}{\|\mathbf{p}\|_{(\ell)} - \varepsilon} \right\}. \tag{225}$$

Proof: If $\varepsilon = 0$, the Lemma follows directly from the definition. We thus assume from here on that $\varepsilon > 0$ and let $\mathbf{1}^{\mathbf{C}} \in \mathrm{Prob}^{\downarrow}(|\mathbf{C}|)$ be the probability vector of which the first entry is equal to one. The eigenvalues of $\rho^A \otimes \psi^C$ are given by $\mathbf{p} \otimes \mathbf{1}^{\mathbf{C}}$ and

$$\lim_{|C|\to\infty} \frac{1}{2} \left\| \mathbf{p} \otimes \mathbf{1}^{\mathbf{C}} - \mathbf{u}^{(\mathbf{d}|\mathbf{C}|)} \right\|_1 = 1. \tag{226}$$

According to Lem. 4, for a fixed system C that is large enough (and remembering that $\varepsilon < 1$), this implies that

$$\begin{aligned} &\max_{\omega\in\mathfrak{B}_{\varepsilon}(\rho^A\otimes\psi^C)} H_{\min}\left(\omega^{AC}\right) \\ &= \log \min_{\ell\in[d|C|]} \left\{ \frac{\ell}{\|\mathbf{p} \otimes \mathbf{1}^{\mathbf{C}}\|_{(\ell)} - \varepsilon} \right\} \\ &= \log \min_{\ell\in[d]} \left\{ \frac{\ell}{\|\mathbf{p}\|_{(\ell)} - \varepsilon} \right\}, \end{aligned} \tag{227}$$

where we used that $\|\mathbf{p} \otimes \mathbf{1}^{\mathbf{C}}\|_{(\ell)} = \mathbf{1}$ for $\ell \geq d$. In case that $|C|$ is not large enough in the sense that

$$\frac{1}{2} \left\| \mathbf{p} \otimes \mathbf{1}^{\mathbf{C}} - \mathbf{u}^{(\mathbf{d}|\mathbf{C}|)} \right\|_1 < \varepsilon \tag{228}$$

(if this happens for any C), the discussion above the Lemma (see Eq. (65)) shows that the supremum will be reached for larger C. This finishes the proof.

Theorem 16: Let $\psi \in \mathrm{Pure}(AB)$ and $\varepsilon \in [0, 1)$. The ε-single-shot distillable entanglement of ψ^{AB} is then given by

$$\mathrm{Distill}^{\varepsilon}(\psi^{AB}) = \log \left\lfloor 2^{\tilde{H}^{\varepsilon}_{\min}(\rho^A)} \right\rfloor,$$

where $\rho^A = \mathrm{Tr}_B(\psi^{AB})$ is the reduced density matrix of ψ^{AB}.

Proof: According to the first line of Eq. (115),

$$\begin{aligned} &\mathrm{Distill}^{\varepsilon}\left(\psi^{AB}\right) \\ &= \max_{m\in\mathbb{N}} \left\{ \log m \ : \ \|\mathbf{p}\|_{(k)} - \frac{k}{m} \leq \varepsilon \ \ \forall\, k \in [d] \right\} \\ &= \max_{m\in\mathbb{N}} \left\{ \log m \ : \ m \leq \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon} \ \ \forall\, k \in [d] \right\} \\ &= \max_{m\in\mathbb{N}} \left\{ \log m \ : \ m \leq \min_{k\in[d]} \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon} \right\} \\ &= \log \left\lfloor \min_{k\in[d]} \frac{k}{\|\mathbf{p}\|_{(k)} - \varepsilon} \right\rfloor, \end{aligned} \tag{229}$$

which finishes the proof by invoking Cor. 4

Lemma 14: Let $F\left(\Phi_m \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{A\to B}\right)$ be defined as in Eq. (73). It holds that

$$\begin{aligned} &F\left(\Phi_m \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{A\to B}\right) \\ &= \min_{\psi\in\mathrm{Pure}(A\tilde{A})} \sup_{\Theta} F\left(\Theta[\Phi_m](\psi^{A\tilde{A}}), \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right) \end{aligned} \tag{230}$$

where the supremum is again over all LOCC superchannels Θ that map the state Φ_m to a channel in $\mathrm{CPTP}(\tilde{A} \to B)$.

Proof: We begin by noting that

$$\tilde{D}_{1/2}(\rho\|\sigma) = -2\log F(\rho, \sigma), \tag{231}$$

where $\tilde{D}_{1/2}$ denotes the sandwiched Rényi relative entropy of order 1/2 [84], [85]. The data-processing inequality of $\tilde{D}_{1/2}$ follows directly from the analogous property of the fidelity and that $\tilde{D}_{1/2}$ obeys the direct-sum property, i.e.,

$$\begin{aligned} &\tilde{D}_{1/2}\left(\sum_i p_i |i\rangle\langle i| \otimes \rho_i \middle\| \sum_i p_i |i\rangle\langle i| \otimes \sigma_i \right) \\ &= \sum_i p_i \tilde{D}_{1/2}(\rho_i\|\sigma_i) \end{aligned} \tag{232}$$

too. According to (the proofs of) [63, Prop. 8] and [64, Lem. II.3], it therefore holds that

$$\tilde{D}_{1/2}(\mathcal{N}^{A'\to B}(\phi^{AA'}_{\rho})\|\mathcal{M}^{A'\to B}(\phi^{AA'}_{\rho}))$$

is concave in $\rho^{A'}$, where $\phi_\rho^{AA'}$ is a purification of $\rho^{A'}$ and $\mathcal{M}^{A'\to B}, \mathcal{N}^{A'\to B}$ are quantum channels. In short, their argument is the following: For any convex combination $\rho^{A'} = \sum_i p_i \rho_i^{A'}$, suppose $\rho_i^{A'}$ has a purification $|\phi_i^{AA'}\rangle$. Then

$$|\psi^{PAA'}\rangle = \sum_i \sqrt{p_i}|i^P\rangle \otimes |\phi_i^{AA'}\rangle \tag{233}$$

is a purification of the average state $\rho^{A'}$. Since all purifications are related by an isometry, there exists an isometric channel $\mathcal{W}^{A\to PA}$ such that $\mathcal{W}^{A\to PA}(\phi_\rho^{AA'}) = \psi^{PAA'}$. Then we have

$$\begin{aligned}
&\tilde{D}_{1/2}\Big(\mathcal{N}^{A'\to B}(\phi_\rho^{AA'})\Big\|\mathcal{M}^{A'\to B}(\phi_\rho^{AA'})\Big)\\
&= \tilde{D}_{1/2}\Big(\mathcal{W}^{A\to PA}\mathcal{N}^{A'\to B}(\phi_\rho^{AA'})\Big\|\mathcal{W}^{A\to PA}\mathcal{M}^{A'\to B}(\phi_\rho^{AA'})\Big)\\
&= \tilde{D}_{1/2}\Big(\mathcal{N}^{A'\to B}\mathcal{W}^{A\to PA}(\phi_\rho^{AA'})\Big\|\mathcal{M}^{A'\to B}\mathcal{W}^{A\to PA}(\phi_\rho^{AA'})\Big)\\
&= \tilde{D}_{1/2}\Big(\mathcal{N}^{A'\to B}(\psi^{PAA'})\Big\|\mathcal{M}^{A'\to B}(\psi^{PAA'})\Big)\\
&\geq \tilde{D}_{1/2}\Bigg(\sum_i p_i|i\rangle\langle i|^P \otimes \mathcal{N}^{A'\to B}(\phi_i^{AA'})\Bigg\|\\
&\qquad\qquad \sum_i p_i|i\rangle\langle i|^P \otimes \mathcal{M}^{A'\to B}(\phi_i^{AA'})\Bigg)\\
&= \sum_i p_i \tilde{D}_{1/2}\left(\mathcal{N}^{A'\to B}(\phi_i^{AA'})\|\mathcal{M}^{A'\to B}(\phi_i^{AA'})\right),
\end{aligned} \tag{234}$$

where the first equality follows from the isometric invariance of the divergence, the second equality follows as $\mathcal{W}^{A\to PA}$ commutes with $\mathcal{N}^{A'\to B}$ and $\mathcal{M}^{A'\to B}$, the inequality follows from the data-processing inequality of $\tilde{D}_{1/2}$ under the dephasing channel $\sum_i |i\rangle\langle i|^P \cdot |i\rangle\langle i|^P$, and the last equality follows from the direct-sum property of $\tilde{D}_{1/2}$.

We thus find that

$$\begin{aligned}
&\inf_{\Theta\in\mathrm{LOCC}} \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \tilde{D}_{1/2}\left(\Theta[\Phi_m](\psi^{A\tilde{A}})\Big\|\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right)\\
&= \inf_{\Theta\in\mathrm{LOCC}} \max_{\rho^A\in\mathfrak{D}(A)} \tilde{D}_{1/2}\left(\Theta[\Phi_m](\phi_\rho^{A\tilde{A}})\Big\|\mathcal{N}^{\tilde{A}\to B}(\phi_\rho^{A\tilde{A}})\right)\\
&= \max_{\rho^A\in\mathfrak{D}(A)} \inf_{\Theta\in\mathrm{LOCC}} \tilde{D}_{1/2}\left(\Theta[\Phi_m](\phi_\rho^{A\tilde{A}})\Big\|\mathcal{N}^{\tilde{A}\to B}(\phi_\rho^{A\tilde{A}})\right)\\
&= \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\Theta\in\mathrm{LOCC}} \tilde{D}_{1/2}\left(\Theta[\Phi_m](\psi^{A\tilde{A}})\Big\|\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right),
\end{aligned} \tag{235}$$

where the first and last equalities follow because the objective function is invariant with respect to the purification. The second equality follows from the above argument showing that the objective function is concave in ρ^A, the fact that the objective function is convex in Θ, and that we are thus allowed to apply Sion's minimax theorem (clearly $\mathfrak{D}(A)$ is compact and both sets over which we optimize are convex) [86, Cor. 3.3]. Remembering that $\tilde{D}_{1/2}(\rho\|\sigma) = -2\log F(\rho,\sigma)$ and thus, e.g.,

$$\begin{aligned}
&\inf_{\Theta\in\mathrm{LOCC}} \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \tilde{D}_{1/2}\left(\Theta[\Phi_m](\psi^{A\tilde{A}})\Big\|\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right)\\
&= \inf_{\Theta\in\mathrm{LOCC}} \max_{\psi\in\mathrm{Pure}(A\tilde{A})} -2\log F\Big(\Theta[\Phi_m](\psi^{A\tilde{A}}), \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\Big)\\
&= \inf_{\Theta\in\mathrm{LOCC}} \max_{\psi\in\mathrm{Pure}(A\tilde{A})} -2\log F\Big(\Theta[\Phi_m](\psi^{A\tilde{A}}), \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\Big)\\
&= -2\log \sup_{\Theta\in\mathrm{LOCC}} \min_{\psi\in\mathrm{Pure}(A\tilde{A})} F\Big(\Theta[\Phi_m](\psi^{A\tilde{A}}), \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\Big)\\
&= -2\log F(\Phi_m \to \mathcal{N}^{A\to B})
\end{aligned} \tag{236}$$

completes the proof.

Lemma 15: Let $\mathcal{N} \in \mathrm{CPTP}(A\to B)$ be a quantum channel. It then holds that

$$\begin{aligned}
&1 - E_{(m)}\left(\mathcal{N}^{A\to B}\right)\\
&\leq F\left(\Phi_m \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{A\to B}\right)\\
&\leq \sqrt{1 - E_{(m)}\left(\mathcal{N}^{A\to B}\right)}\,.
\end{aligned} \tag{237}$$

Proof: We begin by showing the upper bound. To this end, let $\mathfrak{T}$ be the set of all LOCC superchannels Θ for which $\Theta[\Phi_m]$ is a channel from $\tilde{A}$ to B and consider an arbitrary $\psi \in \mathrm{Pure}(A\tilde{A})$. Since $\psi^{A\tilde{A}}$ can be prepared locally, for any $\Theta \in \mathfrak{T}$, there exists an $\mathcal{M} \in \mathrm{LOCC}(A'B' \to AB)$ such that

$$\Theta[\Phi_m]\left(\psi^{A\tilde{A}}\right) = \mathcal{M}\left(\Phi_m^{A'B'}\right), \tag{238}$$

see Fig. 4. This implies that

$$\begin{aligned}
&\sup_{\mathcal{M}\in\mathrm{LOCC}(A'B'\to AB)} F\left(\mathcal{M}\left(\Phi_m^{A'B'}\right), \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)\\
&\geq \sup_{\Theta\in\mathfrak{T}} F\left(\Theta[\Phi_m]\left(\psi^{A\tilde{A}}\right), \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)\\
&\geq \min_{\psi\in\mathrm{Pure}(A\tilde{A})} \sup_{\Theta\in\mathfrak{T}} F\left(\Theta[\Phi_m]\left(\psi^{A\tilde{A}}\right), \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)\\
&= F\left(\Phi_m \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{A\to B}\right),
\end{aligned} \tag{239}$$

where we used Lem. 6 in the last line. Moreover, recalling first that the purified conversion distance can be expressed in terms of $E_{(m)}$ (see Eq. (37)) and then its definition provided in Eq. (8), we note that

$$\begin{aligned}
&E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)\\
&= P^2\left(\Phi_m^{A'B'} \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)\\
&= \min_{\tau\in\mathfrak{D}(AB)} \left\{P^2\left(\tau, \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) : \Phi_m^{A'B'} \xrightarrow{\mathrm{LOCC}} \tau\right\}\\
&= 1 - \sup_{\mathcal{M}\in\mathrm{LOCC}(A'B'\to AB)} F^2\left(\mathcal{M}\left(\Phi_m^{A'B'}\right), \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)\\
&\leq 1 - F^2\left(\Phi_m \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{A\to B}\right).
\end{aligned} \tag{240}$$

Since this is true for any $\psi \in \mathrm{Pure}(A\tilde{A})$, it also holds that

$$\begin{aligned}
E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\right) &= \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)\\
&\leq 1 - F^2\left(\Phi_m \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{A\to B}\right),
\end{aligned} \tag{241}$$

which concludes the first part of the proof.

To complete the proof, we must show that $F\left(\Phi_m \xrightarrow{\mathrm{LOCC}} \mathcal{N}^{A\to B}\right) \geq 1 - E_{(m)}\left(\mathcal{N}^{A\to B}\right)$. To this end, let

$$\mathcal{N}^{\tilde{A}\to B}(\cdot) = \sum_{x\in[n+1]} N_x(\cdot)N_x^* \tag{242}$$

be an operator sum representation of $\mathcal{N}^{\tilde{A}\to B}$, where we assume for $x = n+1$ that $\mathcal{N}_{n+1}^{\tilde{A}\to B} = 0$. Consider a superchannel Θ

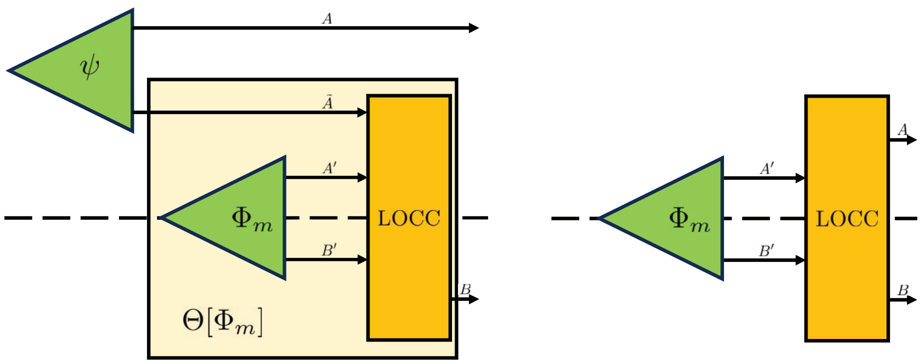

Fig. 4. Figure for the proof of Lem. 7. For fixed $\psi^{A\tilde{A}}$, every state τ^{AB} that can be obtained with a circuit as shown on the left can also be obtained with a circuit as shown on the right by including the local preparation of $\psi^{A\tilde{A}}$ into the LOCC protocol.

of the form given in Eq. (77) where we choose $k = n+1$, $\mathcal{F}_{(x)}^{B'\to B}(\cdot) = V_x(\cdot)V_x^*$ to be isometries, and $\mathcal{E}_x^{\tilde{A}\to B'}(\cdot) = M_x(\cdot)M_x^*$ with

$$M_x := \begin{cases} V_x^* N_x & \text{for } x \in [n], \\ \sqrt{I^B - \sum_{x\in[n]} N_x^* P_x N_x} & \text{for } x = n+1, \end{cases} \tag{243}$$

where $P_x := V_x V_x^* \in \mathrm{Pos}(B)$ is a projection to an m-dimensional subspace. Since

$$\sum_{x\in[n]} N_x^* P_x N_x \le \sum_{x\in[n]} N_x^* N_x = I^B , \tag{244}$$

this ensures that $\{M_x\}_{x\in[n+1]}$ is a valid instrument. Moreover, let $\psi \in \mathrm{Pure}(A\tilde{A})$ be fixed but arbitrary and $\rho^{\tilde{A}} = \mathrm{Tr}_A\left[\psi^{A\tilde{A}}\right]$. Remembering that $F(P_0 + P_1, Q_0 + Q_1) \ge F(P_0, Q_0) + F(P_1, Q_1)$ whenever $P_0, P_1, Q_0, Q_1 \ge 0$, see, e.g., [87, Thm. 3.25], we get that

$$\begin{aligned} &F\left(\sum_{x\in[n+1]} \mathcal{F}_{(x)}^{B'\to B} \circ \mathcal{E}_x^{\tilde{A}\to B'}(\psi^{A\tilde{A}}), \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right) \\ &\ge \sum_{x\in[n]} F\left(\mathcal{F}_{(x)}^{B'\to B} \circ \mathcal{E}_x^{\tilde{A}\to B'}(\psi^{A\tilde{A}}), \mathcal{N}_x^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right) \\ &= \sum_{x\in[n]} \sqrt{\mathrm{Tr}\left[\mathcal{N}_x^{\tilde{A}\to B}(\psi^{A\tilde{A}})\mathcal{F}_{(x)}^{B'\to B} \circ \mathcal{E}_x^{\tilde{A}\to B'}(\psi^{A\tilde{A}})\right]} \\ &= \sum_{x\in[n]} \left|\mathrm{Tr}\left[\rho^{\tilde{A}} N_x^* V_x M_x\right]\right| \\ &= \sum_{x\in[n]} \mathrm{Tr}\left[\rho^{\tilde{A}} N_x^* P_x N_x\right] , \end{aligned} \tag{245}$$

where we used in the third line that $\mathcal{N}_x^{\tilde{A}\to B}(\psi^{A\tilde{A}})$ is pure. Next, we recall that for a positive semidefinite operator A, $\|A\|_{(m)}$ denotes the sum of its m largest eigenvalues. Taking in Eq. (245) P_x to be the projection to the m-dimensional eigen-subspace corresponding to the m largest eigenvalues of $N_x\rho^{\tilde{A}}N_x^*$ and utilizing Lem. 6 thus gives

$$\begin{aligned} &F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{A\to B}\right) \\ &\ge \min_{\rho\in\mathfrak{D}(\tilde{A})} \sup_{\{N_x\}} \sum_{x\in[k]} \left\|N_x\rho^{\tilde{A}}N_x^*\right\|_{(m)} \\ &= 1 - \max_{\rho\in\mathfrak{D}(\tilde{A})} \inf_{\{N_x\}} \left(1 - \sum_{x\in[k]} \left\|N_x\rho^{\tilde{A}}N_x^*\right\|_{(m)}\right), \end{aligned} \tag{246}$$

where the optimizations involving $\{N_x\}$ are over all operator-sum representations of $\mathcal{N}$.

Next, we will show that the right-hand side of the above equation can be expressed in terms of $E_{(m)}$. To this end, for a fixed $\psi^{A\tilde{A}} \in \mathrm{Pure}(A\tilde{A})$, let

$$\rho^{AB} = \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right).$$

Now every unnormalized pure state decomposition of $\rho^{AB} = \sum_x \xi_x^{AB}$ corresponds to an operator-sum representation N_x of $\mathcal{N}$ in the sense that $\xi_x^{AB} = \mathcal{N}_x^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right) := N_x\psi^{A\tilde{A}}N_x^*$: That every operator-sum representation defines a pure state decomposition in this way is obvious. Now let $\{\xi_x^{AB}\}$ be a pure state decomposition corresponding to any operator-sum representation $\{N_x\}$ and let $\{\chi_y^{AB}\}$ be an arbitrary pure state decomposition. According to [88], this implies that there exists a unitary U such that $|\chi_y^{AB}\rangle = \sum_x U_{yx}|\xi_x^{AB}\rangle$. Due to the unitary freedom in operator-sum representations, $\left\{M_y = \sum_x U_{yx}N_x\right\}$ is an operator-sum representation of $\mathcal{N}$ too, and $\chi_y^{AB} = \mathcal{M}_y^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)$. This implies that

$$\begin{aligned} &\inf_{\{N_x\}}\left(1 - \sum_{x\in[k]} \left\|\mathrm{Tr}_A\left[\mathcal{N}_x^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right]\right\|_{(m)}\right) \\ &= \inf\left(1 - \sum_{x\in[k]} p_x \left\|\mathrm{Tr}_A\left[\xi_x^{AB}\right]\right\|_{(m)}\right), \end{aligned} \tag{247}$$

where the second infimum is over all normalized pure state decompositions $\{p_x, \xi_x^{AB}\}$ of $\rho^{AB} = \mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right) = \sum_x p_i\xi_x^{AB}$.

Moreover, remember that according to Eq. (35),

$$E_{(m)}\left(\rho^{AB}\right) = \inf\left(1 - \sum_x p_x \left\|\mathrm{Tr}_A\left[\psi_x^{AB}\right]\right\|_{(m)}\right), \tag{248}$$

where the infimum is over all pure-state decompositions $\rho^{AB} = \sum_x p_x\psi_x^{AB}$. In combination, this shows that

$$\begin{aligned} &E_{(m)}\left(\mathcal{N}^{A\to B}\right) \\ &= \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \end{aligned}$$

$$= \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\{N_x\}} \left(1 - \sum_{x\in[k]} \left\| \mathrm{Tr}_A \left[\mathcal{N}_x^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right]\right\|_{(m)}\right)$$
$$= \max_{\rho\in\mathfrak{D}(\tilde{A})} \inf_{\{N_x\}} \left(1 - \sum_{x\in[k]} \left\|\mathcal{N}_x^{\tilde{A}\to B}\left(\rho^{\tilde{A}}\right)\right\|_{(m)}\right). \quad (249)$$

Inserting this into Eq. (246), we find that

$$F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{A\to B}\right) \geq 1 - E_{(m)}\left(\mathcal{N}^{A\to B}\right), \quad (250)$$

which finishes the proof.

Theorem 17: Let $\mathcal{N} \in \mathrm{CPTP}(A \to B)$ be a quantum channel and $\varepsilon \in [0, 1)$. Then

$$\max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\sigma^{XAB}} H^{\varepsilon}_{\max}(A|X)_\sigma \leq \mathrm{Cost}^{\varepsilon}(\mathcal{N}^{A\to B}) \leq \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\sigma^{XAB}} H^{\varepsilon/2}_{\max}(A|X)_\sigma, \quad (251)$$

where the infimums are over all classical systems X and all classical extensions σ^{XAB} of $\sigma^{AB} = \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})$. Again, the infimums are attained for a regular/classical extension with $|X| = |AB|^2$.

Proof: Notice that for every $\psi \in \mathrm{Pure}(A\tilde{A})$, it holds that

$$\inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} \leq \inf_{m\in\mathbb{N}} \left\{\log m \ : \ \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\}, \quad (252)$$

which implies that

$$\max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} \leq \inf_{m\in\mathbb{N}} \left\{\log m \ : \ \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\}. \quad (253)$$

On the contrary, assuming that

$$\max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} < \inf_{m\in\mathbb{N}} \left\{\log m \ : \ \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} \quad (254)$$

leads to a contradiction: Let $m^\star \in \mathbb{N}$ be such that

$$\log m^\star = \inf_{m\in\mathbb{N}} \left\{\log m \colon \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\}. \quad (255)$$

This implies that there exists an $\chi \in \mathrm{Pure}(A\tilde{A})$ such that $E_{(m^\star)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\chi^{A\tilde{A}}\right)\right) \leq \varepsilon$ and $E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\chi^{A\tilde{A}}\right)\right) > \varepsilon$ for $m = m^\star - 1$ (and thus all $m < m^\star$, since $E_{(k)} \geq E_{(k+1)}$). Consequently,

$$\log m^\star > \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} \geq \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\chi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} = \log m^\star, \quad (256)$$

resulting in the promised contradiction. We thus showed that

$$\max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} = \inf_{m\in\mathbb{N}} \left\{\log m \ : \ \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\}. \quad (257)$$

We now turn to the main part of the proof.

According to Lem. 7, we have

$$P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{A\to B}\right) \geq E_{(m)}\left(\mathcal{N}^{A\to B}\right) . \quad (258)$$

Combining this with the definition of the entanglement cost in Eq. (74) gives

$$\begin{aligned}
&\mathrm{Cost}^{\varepsilon}(\mathcal{N}) \\
&\geq \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{A\to B}\right) \leq \varepsilon\right\} \\
&= \inf_{m\in\mathbb{N}} \left\{\log m \ : \ \max_{\psi\in\mathrm{Pure}(A\tilde{A})} E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} \\
&\underset{(i)}{=} \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right) \leq \varepsilon\right\} \\
&\underset{(ii)}{=} \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ P^2\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right) \leq \varepsilon\right\} \\
&= \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \mathrm{Cost}^{\varepsilon}\left(\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right) \\
&\underset{(iii)}{=} \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\sigma^{XAB}} H^{\varepsilon}_{\max}(A|X)_\sigma ,
\end{aligned} \quad (259)$$

where (i) follows from Eq. (257), (ii) from Eq. (37), (iii) from Thm. 7, and the infimum in the last line is over all classical systems X and all regular extensions σ^{XAB} of $\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})$.

For the converse inequality, expressing the cost in terms of the fidelity gives

$$\begin{aligned}
&\mathrm{Cost}^{\varepsilon}(\mathcal{N}) \\
&= \inf_{m\in\mathbb{N}} \left\{\log m \ : \ F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}\right) \geq \sqrt{1-\varepsilon}\right\} \\
&\underset{(i)}{\leq} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ F\left(\Phi_m \xrightarrow{\text{LOCC}} \mathcal{N}\right) \geq 1 - \varepsilon/2\right\} \\
&\underset{(ii)}{\leq} \inf_{m\in\mathbb{N}} \left\{\log m \ : \ E_{(m)}\left(\mathcal{N}\right) \leq \varepsilon/2\right\} \\
&= \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \mathrm{Cost}^{\varepsilon/2}\left(\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})\right) \\
&= \max_{\psi\in\mathrm{Pure}(A\tilde{A})} \inf_{\sigma^{XAB}} H^{\varepsilon/2}_{\max}(A|X)_\sigma ,
\end{aligned} \quad (260)$$

where (i) follows from $\sqrt{1-\varepsilon} \leq 1 - \varepsilon/2$, (ii) from Lem. 7, the last two lines follow from the same steps as in Eq. (259), and the infimum is again over all classical systems X and all regular extensions σ^{XAB} of $\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})$. The fact that the infimums are achieved and can be taken over all classical extensions follows from the analogous statements in Thm. 7.

Theorem 24: Let $\rho \in \mathfrak{D}(AB)$ with $m = |A| = |B|$, $\sigma \in \mathfrak{D}(A'B')$, and $\mathcal{N} \in \mathrm{LOCC}_1(AB \to A'B')$. The coherent information of entanglement $E_\to$ is

1) monotonic under one-way LOCC, i.e.,

$$E_\to\left(\mathcal{N}^{AB\to A'B'}\left(\rho^{AB}\right)\right) \leq E_\to\left(\rho^{AB}\right) ,$$

2) non-negative, i.e., $E_{\to}(\rho^{AB}) \geq 0$, with equality if ρ^{AB} is separable,
3) strongly monotonic under one-way LOCC, i.e., for any ensemble $\{p_y, \sigma_y^{A'B'}\}$ that can be obtained from ρ^{AB} using one-way LOCC and subselection, it holds that

$$E_{\to}\left(\rho^{AB}\right) \geq \sum_y p_y E_{\to}\left(\sigma_y^{A'B'}\right),$$

4) convex,
5) bounded by $E_{\to}\left(\rho^{AB}\right) \leq E_{\to}(\Phi_m) = \log(m)$,
6) superadditive, i.e.,

$$E_{\to}\left(\rho^{AB} \otimes \sigma^{A'B'}\right) \geq E_{\to}\left(\rho^{AB}\right) + E_{\to}\left(\sigma^{A'B'}\right).$$

Proof: 1. First observe that since $\mathcal{N} \in \mathrm{LOCC}_1(AB \to A'B')$, for any $\mathcal{L} \in \mathrm{CPTP}(A' \to A'X)$ we have that $\mathcal{L} \circ \mathcal{N} \in \mathrm{LOCC}_1(AB \to A'B'X)$. Therefore,

$$\begin{aligned} &E_{\to}\left(\mathcal{N}^{AB\to A'B'}\left(\rho^{AB}\right)\right) \\ &= \sup_{\mathcal{L}\in\mathrm{CPTP}(A'\to A'X)} I\big(A'\rangle B'X\big)_{\mathcal{L}\circ\mathcal{N}(\rho)} \\ &\leq \sup_{\mathcal{M}\in\mathrm{LOCC}_1(AB\to A'B'X)} I\big(A'\rangle B'X\big)_{\mathcal{M}(\rho)}. \end{aligned} \tag{261}$$

According to Eq. (82), every $\mathcal{M} \in \mathrm{LOCC}_1(AB \to A'B'X)$ can be expressed as

$$\mathcal{M}^{AB\to A'B'X} := \sum_{y\in[n]} \mathcal{E}_y^{A\to A'X} \otimes \mathcal{F}_{(y)}^{B\to B'}, \tag{262}$$

with each $\mathcal{E}_y^{A\to A'X}$ being a CP map such that $\sum_{y\in[n]} \mathcal{E}_y \in \mathrm{CPTP}(A \to A'X)$, and each $\mathcal{F}_{(y)} \in \mathrm{CPTP}(B \to B')$. Let further be

$$q_y := \mathrm{Tr}\left[\mathcal{E}_y^{A\to A'X}\left(\rho^{AB}\right)\right] \tag{263}$$

and

$$\sigma_y^{A'BX} := \frac{1}{q_y}\mathcal{E}_y^{A\to A'X}\left(\rho^{AB}\right). \tag{264}$$

With these notations, we have

$$\mathcal{M}^{AB\to A'B'X}\left(\rho^{AB}\right) = \sum_{y\in[n]} q_y \mathcal{F}_{(y)}^{B\to B'}\left(\sigma_y^{A'BX}\right). \tag{265}$$

Remembering that

$$\begin{aligned} &D\left(\tau^{CD}\middle\| I^C \otimes \tau^D\right) \\ &= \mathrm{Tr}\left(\tau^{CD}\log\left(\tau^{CD}\right)\right) - \mathrm{Tr}\left(\tau^{CD}\log\left(I^C \otimes \tau^D\right)\right) \\ &= \mathrm{Tr}\left(\tau^{CD}\log\left(\tau^{CD}\right)\right) - \mathrm{Tr}\left(\tau^{CD} I^C \otimes \log\left(\tau^D\right)\right) \\ &= \mathrm{Tr}\left(\tau^{CD}\log\left(\tau^{CD}\right)\right) - \mathrm{Tr}\left(\tau^D\log\left(\tau^D\right)\right) \\ &= -H\left(\tau^{CD}\right) + H\left(\tau^D\right) \\ &= -H(C|D)_\tau = I(C\rangle D)_\tau, \end{aligned} \tag{266}$$

it follows from the joint convexity of the relative entropy, the data processing inequality, and the inequality in Eq. (261) that

$$\begin{aligned} &E_{\to}\left(\mathcal{N}^{AB\to A'B'}\left(\rho^{AB}\right)\right) \\ &\leq \sup_{\mathcal{M}\in\mathrm{LOCC}_1(AB\to A'B'X)} I\big(A'\rangle B'X\big)_{\mathcal{M}(\rho)} \\ &= \sup D\Big(\sum_{y\in[n]} q_y \mathcal{F}_{(y)}^{B\to B'}\left(\sigma_y^{A'BX}\right) \Big\| \\ &\qquad I^{A'} \otimes \sum_{y\in[n]} q_y \mathcal{F}_{(y)}^{B\to B'}\left(\sigma_y^{BX}\right)\Big) \\ &\leq \sup \sum_{y\in[n]} q_y D\Big(\mathcal{F}_{(y)}^{B\to B'}\left(\sigma_y^{A'BX}\right)\Big\| I^{A'} \otimes \mathcal{F}_{(y)}^{B\to B'}\left(\sigma_y^{BX}\right)\Big) \\ &\leq \sup \sum_{y\in[n]} q_y D\Big(\sigma_y^{A'BX}\Big\| I^{A'} \otimes \sigma_y^{BX}\Big) \\ &= \sup \sum_{y\in[n]} q_y I\big(A'\rangle BX\big)_{\sigma_y}, \end{aligned} \tag{267}$$

where the supremums are over all decompositions described above corresponding to an $\mathcal{M} \in \mathrm{LOCC}_1(AB \to A'B'X')$.

Let $\{\tau_i\} \subset \mathfrak{D}(C)$ be a set of states that are mutually orthogonal and $\{r_i\}$ a probability distribution. It is well known (and follows from a straightforward calculation) that this implies that

$$H\left(\sum_i r_i\tau_i\right) = \sum_i r_i H(\tau_i) - \sum_i r_i \log(r_i). \tag{268}$$

If $\rho^{A'BX} := \sum_{x\in[n]} p_x \rho_x^{A'B} \otimes |x\rangle\langle x|^X$, i.e., is a quantum-classical-state in $\mathfrak{D}(A'BX)$, this implies that

$$I(A'\rangle BX)_\rho = \sum_{x\in[n]} p_x I(A'\rangle B)_{\rho_x} : \tag{269}$$

Let $\rho_x^{A'BX} = \rho_x^{A'B} \otimes |x\rangle\langle x|$. From Eqs. (266) and (268), we thus find that

$$\begin{aligned} I(A'\rangle BX)_\rho &= H\left(\rho^{BX}\right) - H\left(\rho^{A'BX}\right) \\ &= \sum_{x\in[n]} p_x H\left(\rho_x^{BX}\right) - \sum_{x\in[n]} p_x \log(p_x) \\ &\quad - \sum_{x\in[n]} p_x H\left(\rho_x^{A'BX}\right) + \sum_{x\in[n]} p_x \log(p_x) \\ &= \sum_{x\in[n]} p_x I(A'\rangle B)_{\rho_x}, \end{aligned} \tag{270}$$

which proves the claim.

From Eq. (267), and remembering the notation from Eqs. (263) and (264), it then follows that

$$\begin{aligned} &E_{\to}\left(\mathcal{N}^{AB\to A'B'}\left(\rho^{AB}\right)\right) \\ &\leq \sup \sum_{y\in[n]} q_y I\big(A'\rangle BX\big)_{\sigma_y} \\ &= \sup I\big(A'\rangle BXY\big)_{\sum_{y\in[n]} q_y \sigma_y^{A'BX} \otimes |y\rangle\langle y|^Y} \\ &= \sup I\big(A'\rangle BXY\big)_{\sum_{y\in[n]} \mathcal{E}_y^{A\to A'X}\left(\rho^{AB}\right) \otimes |y\rangle\langle y|^Y}, \end{aligned} \tag{271}$$

where in the last line, the supremum is over all classical systems Y with arbitrary dimension n, all dimensions of X, and all instruments $\{\mathcal{E}_y^{A\to A'X}\}_{y\in[n]}$.

Now let $Z := XY$. Since $\mathcal{E}^{A\to A'Z}(\rho^A) := \sum_{y\in[n]} \mathcal{E}_y^{A\to A'X}\left(\rho^A\right) \otimes |y\rangle\langle y|^Y \in \mathrm{CPTP}(A \to A'Z)$ for any instrument $\{\mathcal{E}_y^{A\to A'X}\}_{y\in[n]}$, we find that

$$\begin{aligned} &E_{\to}\left(\mathcal{N}^{AB\to A'B'}\left(\rho^{AB}\right)\right) \\ &\leq \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to A'Z)} I\big(A'\rangle BZ\big)_{\mathcal{E}(\rho)}, \end{aligned} \tag{272}$$

where the supremum is also over all dimensions of the systems A' and Z. We will conclude the proof by showing that the second line in the above equation equals $E_{\to}\left(\rho^{AB}\right)$, i.e., that we can restrict the supremum without loss of generality to the case $A' = A$ (compare to Def. 1): Observe first that the coherent information remains invariant under local isometries. If $|A'| \leq |A|$, this implies that the supremum over CPTP($A \to AZ$) cannot be smaller than the supremum over CPTP($A \to A'Z$) (since we can always use an isometric embedding).

On the other hand, suppose that $|A'| > |A|$. In general, we notice that in Eq. (272) (and similarly in Eq. (83)), the supremum can be restricted to quantum channels of the form

$$\mathcal{E}^{A\to A'Z} = \sum_{x\in[n]} \mathcal{E}_x^{A\to A'} \otimes |x\rangle\langle x|^Z, \tag{273}$$

where each $\mathcal{E}_x^{A\to A'}$ is a CP map with a single Kraus operator. This follows from Eq. (266) and the data processing inequality of the relative entropy: An arbitrary $\mathcal{E}^{A\to A'Z} \in$ CPTP($A \to A'Z$) can be written as

$$\mathcal{E}^{A\to A'Z} = \sum_{x\in[n],y\in[m]} \mathcal{E}_{x,y}^{A\to A'} \otimes |x\rangle\langle x|^Z, \tag{274}$$

where each $\mathcal{E}_{x,y}^{A\to A'}$ is a CP map with a single Kraus operator $M_{x,y} : A \to A'$. Introducing another classical system $\tilde{Z}$ of dimension m,

$$\tilde{\mathcal{E}}^{A\to A'Z\tilde{Z}} := \sum_{x\in[n],y\in[m]} \mathcal{E}_{x,y}^{A\to A'} \otimes |x\rangle\langle x|^Z \otimes |y\rangle\langle y|^{\tilde{Z}}, \tag{275}$$

is an element of CPTP($A \to A'Z\tilde{Z}$) and

$$\mathcal{E}^{A\to A'Z} = \operatorname{Tr}_{Z'} \circ \tilde{\mathcal{E}}^{A\to A'Z\tilde{Z}}. \tag{276}$$

It thus follows that

$$\begin{aligned} &I(A'\rangle BZ\tilde{Z})_{\tilde{\mathcal{E}}(\rho)} \\ &= D\left(\tilde{\mathcal{E}}(\rho^{AB}) \middle\| I^{A'} \otimes \operatorname{Tr}_{A'} \circ \tilde{\mathcal{E}}(\rho^{AB})\right) \\ &\geq D\left(\operatorname{Tr}_{Z'} \circ \tilde{\mathcal{E}}(\rho^{AB}) \middle\| I^{A'} \otimes \operatorname{Tr}_{Z'} \circ \operatorname{Tr}_{A'} \circ \tilde{\mathcal{E}}(\rho^{AB})\right) \\ &= D\left(\mathcal{E}(\rho^{AB}) \middle\| I^{A'} \otimes \operatorname{Tr}_{A'} \circ \mathcal{E}(\rho^{AB})\right) \\ &= I(A'\rangle BZ)_{\mathcal{E}(\rho)}. \end{aligned} \tag{277}$$

Since the supremums include a supremum over the classical system anyway, this proves our claim.

Now consider a channel as in Eq. (273), i.e., each $\mathcal{E}_x^{A\to A'}(\cdot) = M_x(\cdot)M_x^*$ is a CP map with a single Kraus operator $M_x : A \to A'$. According to the polar decomposition, each M_x can be expressed as $M_x = V_x N_x$, where each $N_x : A \to A$ is an element of a generalized measurement, and each $V_x : A \to A'$ is an isometry. Defining an isometry

$$V^{AZ\to A'Z} := \sum_x V_x^{A\to A'} \otimes |x\rangle\langle x|^Z, \tag{278}$$

and a channel

$$\mathcal{N}^{A\to AZ}(\cdot) := \sum_{x\in[n]} N_x(\cdot)N_x^* \otimes |x\rangle\langle x|^Z, \tag{279}$$

it holds that $\mathcal{E}^{A\to A'Z} = \mathcal{V}^{AZ\to A'Z} \circ \mathcal{N}^{AZ\to AZ}$, where $\mathcal{V}^{A\to A'Z}(\cdot) = V(\cdot)V^*$. Using again that the coherent information is invariant under local isometries, we find that

$$I(A'\rangle BZ)_{\mathcal{E}(\rho)} = I(A\rangle BZ)_{\mathcal{N}(\rho)}, \tag{280}$$

which completes the proof of 1.

2. Let

$$\tilde{\mathcal{E}}^{A\to AX}(\rho^A) = \operatorname{Tr}\left(\rho^A\right)\ |0\rangle\langle 0|^A \otimes |0\rangle\langle 0|^X. \tag{281}$$

It then follows that

$$\begin{aligned} &E_{\to}\left(\rho^{AB}\right) \\ &= \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I(A\rangle BX)_{\mathcal{E}(\rho)} \\ &\geq I(A\rangle B)_{\tilde{\mathcal{E}}(\rho)} \\ &= H\left(\rho^B \otimes |0\rangle\langle 0|^X\right) - H\left(|0\rangle\langle 0|^A \otimes \rho^B \otimes |0\rangle\langle 0|^X\right) \\ &= 0. \end{aligned} \tag{282}$$

According to [89, Thm. 1] (see also the announcement in [90] and [91] for a review),

$$I(A\rangle B)_\rho = H(\rho^B) - H(\rho^{AB}) \tag{283}$$

is convex in ρ^{AB}. This immediately implies that if ρ^{AB} is separable, i.e.,

$$\rho^{AB} = \sum_i p_i \phi_i^A \otimes \psi_i^B, \tag{284}$$

then [92]

$$I(A\rangle B)_\rho \leq \sum_i p_i \left(H(\psi_i^B) - H(\phi_i^A \otimes \psi_i^B)\right) = 0. \tag{285}$$

For any $\mathcal{E} \in$ CPTP($A \to AX$), $\mathcal{E}(\rho^{AB})$ is separable if ρ^{AB} was and thus

$$E_{\to}\left(\rho^{AB}\right) = \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I(A\rangle BX)_{\mathcal{E}(\rho)} \leq 0 \tag{286}$$

on separable states. Together with Eq. (282), this finishes the proof of 2. It also follows immediately from the operational interpretation of the coherent information detailed in [93].

3. Let $\rho^{AB} \in \mathfrak{D}(AB)$ be a state. Any ensemble $\{p_y, \sigma_y^{A'B'}\}$ that can be obtained from it by means of one-way LOCC and subselection is of the form

$$\begin{aligned} p_y &= \operatorname{Tr}\left(\mathcal{M}_y^{A\to A'} \otimes \mathcal{F}_{(y)}^{B\to B'}\left(\rho^{AB}\right)\right), \\ \sigma_y^{A'B'} &= \frac{1}{p_y}\mathcal{M}_y^{A\to A'} \otimes \mathcal{F}_{(y)}^{B\to B'}\left(\rho^{AB}\right), \end{aligned} \tag{287}$$

i.e., Alice applies an instrument $\{\mathcal{M}_y\}_{y\in[n]}$, with $\mathcal{M}_y \in$ CP($A \to A'$) and $\sum_{y=1}^n \mathcal{M}_y \in$ CPTP($A \to A'$), sends the outcome y to Bob, who then, conditioned on y, applies a channel $\mathcal{F}_{(y)} \in$ CPTP($B \to B'$) [67].

Now define a channel $\mathcal{N} \in$ CPTP($AB \to A'A_sB'B_s$) as

$$\mathcal{N} = \sum_y \mathcal{M}_y^{A\to A'} \otimes |y\rangle\langle y|^{A_s} \otimes \mathcal{F}_{(y)}^{B\to B'} \otimes |y\rangle\langle y|^{B_s}, \tag{288}$$

where A_s is a classical system *in Alice's possession* and B_s a classical system *in Bob's possession* that are used to store the outcome of our instrument.

Let $\mathfrak{C}(A' \to A'X|A_s)$ be the set of channels that can be written as

$$\begin{aligned} &\mathcal{E}^{A'A_s\to A'A_sX}(\tau^{A'A_s}) \\ &= \sum_y \mathcal{E}_{(y)}^{A'\to A'X} \operatorname{Tr}_{A_s}\left(|y\rangle\langle y|^{A_s}\tau^{A'A_s}\right) \otimes |1\rangle\langle 1|^{A_s} \end{aligned} \tag{289}$$

where each $\mathcal{E}_{(y)} \in \mathrm{CPTP}(A \to A'X)$. Clearly, $\mathcal{N} \in \mathrm{LOCC}_1$, and $\mathfrak{C}(A' \to A'X|A_s) \subset \mathrm{CPTP}(A'A_s \to A'A_sX)$. Observing that

$$\mathcal{N}(\rho^{AB}) = \sum_y p_y \sigma_y^{A'B'} \otimes |y\rangle\langle y|^{A_s} \otimes |y\rangle\langle y|^{B_s}, \tag{290}$$

it thus follows from monotonicity and Eq. (269) that

$$\begin{aligned}
&E_\to\left(\rho^{AB}\right)\\
&\geq E_\to\left(\mathcal{N}\left(\rho^{AB}\right)\right)\\
&= \sup_{\mathcal{E}\in\mathrm{CPTP}(A'A_s\to A'A_sX)} I\big(A'A_s\rangle B'B_sX\big)_{\mathcal{E}\circ\mathcal{N}(\rho)}\\
&\geq \sup_{\mathcal{E}\in\mathfrak{C}(A'\to A'X|A_s)} I\big(A'A_s\rangle B'B_sX\big)_{\mathcal{E}\circ\mathcal{N}(\rho)}\\
&= \sup_{\{\mathcal{E}_{(y)}\in\mathrm{CPTP}(A'\to A'X)\}}\\
&\qquad I\big(A'A_s\rangle B'B_sX\big)_{\sum_y p_y\mathcal{E}_{(y)}\left(\sigma_y^{A'B'}\right)\otimes|1\rangle\langle 1|^{A_s}\otimes|y\rangle\langle y|^{B_s}}\\
&= \sup_{\{\mathcal{E}_{(y)}\in\mathrm{CPTP}(A'\to A'X)\}}\\
&\qquad \sum_y p_y I\big(A'A_s\rangle B'X\big)_{\mathcal{E}_{(y)}\left(\sigma_y^{A'B'}\right)\otimes|1\rangle\langle 1|^{A_s}}\\
&= \sum_y p_y \sup_{\mathcal{E}\in\mathrm{CPTP}(A'\to A'X)\}} I\big(A'\rangle B'X\big)_{\mathcal{E}\left(\sigma_y^{A'B'}\right)}\\
&= \sum_y p_y E_\to\left(\sigma^{A'B'}\right),
\end{aligned} \tag{291}$$

which finishes the proof.

4. Convexity is inherited from the convexity of $I\big(A\rangle B\big)$ [89, Thm. 1]: For $\rho^{AB} = t\sigma^{AB} + (1-t)\tau^{AB}$ and $t \in [0,1]$,

$$\begin{aligned}
&E_\to\left(\rho^{AB}\right)\\
&= \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I\big(A\rangle BX\big)_{\mathcal{E}(\rho)}\\
&= \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I\big(A\rangle BX\big)_{t\mathcal{E}(\sigma)+(1-t)\mathcal{E}(\tau)}\\
&\leq \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} \left(tI\big(A\rangle BX\big)_{\mathcal{E}(\sigma)} + (1-t)I\big(A\rangle BX\big)_{\mathcal{E}(\tau)}\right)\\
&\leq t\sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I\big(A\rangle BX\big)_{\mathcal{E}(\sigma)}\\
&\quad + (1-t)\sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I\big(A\rangle BX\big)_{\mathcal{E}(\tau)}\\
&= tE_\to\left(\sigma^{AB}\right) + (1-t)E_\to\left(\tau^{AB}\right).
\end{aligned} \tag{292}$$

5. Let $\Phi_m \in \mathfrak{D}(AB)$ be the maximally entangled state with $m := |A| = |B|$. This implies that

$$\begin{aligned}
&E_\to\left(\Phi^{AB}\right)\\
&\geq I\big(A\rangle BX\big)_{\Phi_m}\\
&= H\left(\Phi_m^B\right) - H\left(\Phi_m^{AB}\right)\\
&= \log(m).
\end{aligned} \tag{293}$$

Moreover, for any state $\rho \in \mathfrak{D}(AB)$ and any channel $\mathcal{E} \in \mathrm{CPTP}(A \to AX)$ with $m = |A| = |B|$, we have that $\mathcal{E}(\rho)$ is quantum-classical and we can thus apply Eq. (269) to find that

$$\begin{aligned}
&E_\to\left(\rho^{AB}\right)\\
&= \sup_{\mathcal{E}\in\mathrm{CPTP}(A\to AX)} I\big(A\rangle BX\big)_{\mathcal{E}(\rho^{AB})}\\
&\leq \sup_{\{p_x,\sigma_x^{AB}\}} \sum_x p_x I\big(A\rangle B\big)_{\sigma_x^{AB}}\\
&= \sup_{\{p_x,\sigma_x^{AB}\}} \sum_x p_x \left[H\left(\sigma_x^B\right) - H\left(\sigma_x^{AB}\right)\right]\\
&\leq \sup_{\{p_x,\sigma_x^{AB}\}} \sum_x p_x \log(m)\\
&= \log(m).
\end{aligned} \tag{294}$$

In combination, this proves that

$$E_\to\left(\Phi_m^{AB}\right) = \log(m). \tag{295}$$

6. From the additivity of the von Neumann entropy and Eq. (266), it follows that

$$I\big(AA'\rangle BB'\big)_{\rho^{AB}\otimes\sigma^{A'B'}} = I\big(A\rangle B\big)_{\rho^{AB}} + I\big(A'\rangle B'\big)_{\sigma^{A'B'}}. \tag{296}$$

This implies that

$$\begin{aligned}
&E_\to\left(\rho^{AB}\otimes\sigma^{A'B'}\right)\\
&= \sup_{\mathcal{E}\in\mathrm{CPTP}(AA'\to AA'X)} I\big(AA'\rangle BB'X\big)_{\mathcal{E}(\rho^{AB}\otimes\sigma^{A'B'})}\\
&= \sup_{\mathcal{E}\in\mathrm{CPTP}(AA'\to AA'XX')} I\big(AA'\rangle BB'XX'\big)_{\mathcal{E}(\rho^{AB}\otimes\sigma^{A'B'})}\\
&\geq \sup_{\substack{\mathcal{E}\in\mathrm{CPTP}(A\to AX)\\ \mathcal{E}'\in\mathrm{CPTP}(A'\to A'X')}} I\big(AA'\rangle BB'XX'\big)_{\mathcal{E}(\rho^{AB})\otimes\mathcal{E}'(\sigma^{A'B'})}\\
&= \sup_{\substack{\mathcal{E}\in\mathrm{CPTP}(A\to AX)\\ \mathcal{E}'\in\mathrm{CPTP}(A'\to A'X')}} \Big[I\big(A\rangle BX\big)_{\mathcal{E}(\rho^{AB})}\\
&\qquad\qquad + I\big(A'\rangle B'X'\big)_{\mathcal{E}'(\sigma^{A'B'})}\Big]\\
&= E_\to\left(\rho^{AB}\right) + E_\to\left(\sigma^{A'B'}\right).
\end{aligned} \tag{297}$$

Theorem 25: Let $\rho \in \mathfrak{D}(AB)$ and $\varepsilon \in (0, 1/2)$. Then, the one-way ε-single-shot distillable entanglement is bounded by

$$\mathrm{Distill}_\to^\varepsilon\left(\rho^{AB}\right) \leq \frac{1}{1-2\varepsilon}E_\to\left(\rho^{AB}\right) + \frac{1+\varepsilon}{1-2\varepsilon}h\left(\frac{\varepsilon}{1+\varepsilon}\right), \tag{298}$$

where $h(x) := -x\log x - (1-x)\log(1-x)$ is the binary Shannon entropy.

Proof: Let $m \in \mathbb{N}$ be such that $\mathrm{Distill}_\to^\varepsilon\left(\rho^{AB}\right) = \log m$, and thus

$$T\left(\rho^{AB} \xrightarrow{\mathrm{LOCC}_1} \Phi_m^{A'B'}\right) \leq \varepsilon. \tag{299}$$

This means that $\rho^{AB} \xrightarrow{\mathrm{LOCC}_1} \sigma^{A'B'}$ for some state $\sigma \in \mathfrak{D}(A'B')$ that is ε-close to $\Phi_m^{A'B'}$. Therefore, from the monotonicity of $E_\to$ under one-way LOCC we get that

$$E_\to\left(\rho^{AB}\right) \geq E_\to\left(\sigma^{A'B'}\right). \tag{300}$$

Next, we use the fact that $\sigma^{A'B'}$ is ε-close to Φ_m to show that the right-hand side of the equation above cannot be much smaller than $\log(m)$. Indeed, combining the continuity of $I(A'\rangle B')_\rho := -H(A'|B')_\rho$ (see [94, Lem. 2]), with the fact that $I(A'\rangle B')_{\Phi_m} = \log m$ (this follows directly from Eq. (266)), gives

$$\begin{aligned}
&E_\to\left(\sigma^{A'B'}\right)\\
&\geq I(A'\rangle B')_\sigma\\
&\geq I(A'\rangle B')_{\Phi_m} - 2\varepsilon\log m - (1+\varepsilon)h\left(\frac{\varepsilon}{1+\varepsilon}\right)\\
&= (1-2\varepsilon)\log(m) - (1+\varepsilon)h\left(\frac{\varepsilon}{1+\varepsilon}\right).
\end{aligned} \tag{301}$$

The proof is concluded by combining this inequality with Eq. (300).

Lemma 16: Let $\rho \in \mathfrak{D}(AB)$, and $\Phi_m \in \mathfrak{D}(A'B')$ be the maximally entangled state with $m := |A'| = |B'|$. Then,

$$T\left(\rho \xrightarrow{\text{LOCC}_1} \Phi_m\right) = P^2\left(\rho \xrightarrow{\text{LOCC}_1} \Phi_m\right) = 1 - \sup_{\mathcal{N}\in\text{LOCC}_1} \text{Tr}\left[\Phi_m \mathcal{N}(\rho)\right] , \quad (302)$$

where the supremum is over all $\mathcal{N} \in \text{LOCC}_1(AB \to A'B')$, and P is the purified distance as given in Eq. (7).

Proof: Let $\mathcal{G} \in \text{LOCC}_1(A'B' \to A'B')$ be the twirling map introduced in [95] acting on any $\omega \in \mathfrak{D}(A'B')$ as

$$\mathcal{G}(\omega) := \int_{U(m)} dU\, (U \otimes \overline{U})\omega(U \otimes \overline{U})^* \quad (303)$$

where dU denotes the uniform probability distribution on the unitary group proportional to the Haar measure. It was also shown in [95] that

$$\mathcal{G}(\omega) = (1 - \text{Tr}[\Phi_m\omega])\,\tau + \text{Tr}[\Phi_m\omega]\,\Phi_m , \quad (304)$$

where $\tau \in \mathfrak{D}(A'B')$ is given by $\tau = (I - \Phi_m)/(m^2 - 1)$. From this follows that for all $\omega \in \mathfrak{D}(A'B')$

$$\frac{1}{2}\|\mathcal{G}(\omega) - \Phi_m\|_1 = \left(1 - \text{Tr}[\Phi_m\omega]\right)\frac{1}{2}\|\tau - \Phi_m\|_1 = 1 - \text{Tr}[\Phi_m\omega] , \quad (305)$$

where the last equality follows from the fact

$$\|\tau - \Phi_m\|_1 = \left\|\frac{I}{m^2-1} - \frac{m^2}{m^2-1}\Phi_m\right\|_1 = 2. \quad (306)$$

From the data processing inequality and the fact that Φ_m is invariant under the twirling map $\mathcal{G}$, it follows that for all $\mathcal{N} \in \text{LOCC}_1(AB \to A'B')$ and all $\rho \in \mathfrak{D}(AB)$

$$\frac{1}{2}\|\mathcal{N}(\rho) - \Phi_m\|_1 \geq \frac{1}{2}\|\mathcal{G}\circ\mathcal{N}(\rho) - \Phi_m\|_1 . \quad (307)$$

Since $\mathcal{G}\circ\mathcal{N}$ is also an LOCC_1 channel ($\mathcal{G}$ can be implemented with shared randomness) it follows from the inequality above that the conversion distance can be expressed as

$$T\left(\rho \xrightarrow{\text{LOCC}_1} \Phi_m\right) = \inf_{\mathcal{N}\in\text{LOCC}_1} \frac{1}{2}\|\mathcal{G}\circ\mathcal{N}(\rho) - \Phi_m\|_1 \overset{Eq.\ (305)}{=} 1 - \sup_{\mathcal{N}\in\text{LOCC}_1} \text{Tr}\left[\Phi_m\mathcal{N}(\rho)\right] . \quad (308)$$

This completes the proof.

Theorem 28: Let $\mathcal{N} \in \text{CPTP}(A \to B)$ be a quantum channel. It then holds that

$$E_{\to}\left(\mathcal{N}^{A\to B}\right) = I(A\rangle B)_{\mathcal{N}} . \quad (309)$$

Proof: From its definition (see Def. 1) we know that for all $\rho \in \mathfrak{D}(AB)$, it holds that $E_{\to}(\rho^{AB}) \geq I(A\rangle B)_\rho$. Combining this with the definition in Eq. (88) gives

$$E_{\to}\left(\mathcal{N}^{A\to B}\right) \geq \max_{\psi\in\text{Pure}(A\tilde{A})} I(A\rangle B)_{\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})} = I(A\rangle B)_{\mathcal{N}} . \quad (310)$$

To obtain the reverse inequality, observe that

$$\begin{aligned}
\max_{\psi\in\text{Pure}(A\tilde{A})} E_{\to}\left(\mathcal{N}^{\tilde{A}\to B}\left(\psi^{A\tilde{A}}\right)\right)
&= \sup_{\substack{\mathcal{E}\in\text{CPTP}(A\to AX)\\ \psi\in\text{Pure}(A\tilde{A})}} I\big(A\rangle BX\big)_{\mathcal{E}^{A\to AX}\otimes\mathcal{N}^{\tilde{A}\to B}(\psi^{A\tilde{A}})}\\
&\underset{(i)}{\leq} \sup_{\sigma\in\mathfrak{D}(A\tilde{A}X)} I\big(A\rangle BX\big)_{\mathcal{N}^{\tilde{A}\to B}(\sigma^{A\tilde{A}X})}\\
&\underset{(ii)}{=} \sup_{\sigma\in\mathfrak{D}(A\tilde{A}X)} \sum_{x\in[n]} p_x I(A\rangle B)_{\mathcal{N}^{\tilde{A}\to B}(\sigma_x^{A\tilde{A}})}\\
&= \max_{\sigma\in\mathfrak{D}(A\tilde{A})} I(A\rangle B)_{\mathcal{N}^{\tilde{A}\to B}(\sigma^{A\tilde{A}})}\\
&\underset{(iii)}{=} I(A\rangle B)_{\mathcal{N}} ,
\end{aligned} \quad (311)$$

where (i) follows by replacing $\mathcal{E}^{A\to AX}(\psi^{A\tilde{A}})$ with arbitrary $\sigma \in \mathfrak{D}(A\tilde{A}X)$, (ii) follows from expanding $\sigma^{A\tilde{A}X}$ as $\sigma^{A\tilde{A}X} = \sum_{x\in[n]} p_x\sigma_x^{A\tilde{A}} \otimes |x\rangle\langle x|^X$, and in (iii), we used that in Eq. (89), we can replace the maximum over all pure states in $\text{Pure}(A\tilde{A})$ with a maximum over all mixed states in $\mathfrak{D}(A\tilde{A})$ because the coherent information is convex.

Thm. 13 was proven in the main text.

Appendix C
Computability

In the following, we will provide the proof of Thm. 4 presented in the main text as well as additional related results. To this end, we begin by showing that $\left\|\mathbf{p}^{\otimes n}\right\|_{(k)}$ can be computed efficiently. Due to Thms. 2 and 3 which express $\text{Distill}^{\varepsilon}\left(\psi^{AB}\right)$ and $\text{Cost}^{\varepsilon}\left(\psi^{AB}\right)$ in terms of the Ky-Fan norm, this will then allow us to derive the promised results. Moreover, we will provide explicit algorithms that can be used to compute all relevant quantities.

Lemma 17: Let $n, k, d \geq 1$ be integers and $\mathbf{p} \in \text{Prob}^{\downarrow}(d)$. Algorithm 1 can be used to efficiently compute $\left\|\mathbf{p}^{\otimes n}\right\|_{(k)}$.

Proof: Note that the n-fold tensor product $\mathbf{p}^{\otimes n}$ has $r := \binom{n+d-1}{d-1}$ different terms $p_1^{n_1}p_2^{n_2}\cdots p_d^{n_d}$, where the n_i are non-negative integers such that $n_1 + n_2 + \cdots + n_d = n$ (see multinomial coefficient), which is a polynomial in n for fixed d. Each term $p_1^{n_1}p_2^{n_2}\cdots p_d^{n_d}$ repeats $\binom{n}{n_1,n_2,\cdots,n_d} = \frac{n!}{n_1!n_2!\cdots n_d!}$ times. First, we sort these r terms in non-increasing order to obtain the ordered vector $(s_1, s_2, \cdots, s_r)$. Let v_i be the number of times that s_i repeats. From this, we get that

$$(\mathbf{p}^{\otimes n})^{\downarrow} = (\underbrace{s_1}_{v_1\text{ times}}, \cdots, \underbrace{s_i}_{v_i\text{ times}}, \cdots, \underbrace{s_r}_{v_r\text{ times}}). \quad (312)$$

Let $N_0 := 0$, $N_k := \sum_{i=1}^k v_i$, and $P_k := \sum_{i=1}^k v_i s_i$. Then we have that $\|\mathbf{p}^{\otimes n}\|_{(N_k)} = P_k$ and for any $m \in [N_k, N_{k+1}]$,

$$\begin{aligned}
\|\mathbf{p}^{\otimes n}\|_{(m)} &= s_{k+1}(m - N_k) + P_k\\
&= s_{k+1}(m - N_{k+1}) + P_{k+1}.
\end{aligned} \quad (313)$$

Therefore, a complete algorithm is given in Algorithm 1.

At this point, let us mention that in Algorithm 1, essentially, we first create an ordered vector and then search it. The search part, i.e., determining which interval m belongs to, can be accomplished with (one of the variations of) binary search, see Algorithm 2. This reduces the required search time, but as a trade-off, one needs to compute and store all values of N_k and P_k.

In summary, we showed how $\left\|\mathbf{p}^{\otimes n}\right\|_{(k)}$ can be computed efficiently. To compute $\text{Distill}^{\varepsilon}\left(\psi^{\otimes n}\right)$, we must determine the integer m satisfying $\|\mathbf{p}^{\otimes n}\|_{(m-1)} \le \varepsilon < \|\mathbf{p}^{\otimes n}\|_{(m)}$ (see Thm. 2, also for notation). That this can be done efficiently is the content of the following Proposition.

Proposition 5: For any $\mathbf{p} \in \text{Prob}^{\downarrow}(d)$, integer $n \ge 1$, and $\varepsilon \in [0, 1)$, the integer $\ell(\mathbf{p}^{\otimes n}, \varepsilon) := \min\{m : \|\mathbf{p}^{\otimes n}\|_{(m)} > \varepsilon\}$ can be computed efficiently by Algorithms 3 and 4.

Algorithm 3 Efficient Evaluation of the Threshold

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \text{Prob}^{\downarrow}(d)$, integer $n \ge 1$, $\varepsilon \in [0, 1)$

Output: $\ell(\mathbf{p}^{\otimes n}, \varepsilon)$

Compute the $r := \binom{n+d-1}{d-1}$ different terms $p_1^{n_1} p_2^{n_2} \cdots p_d^{n_d}$ where $n_1 + n_2 + \cdots + n_d = n$;
Sort the r terms in non-increasing order resulting in the vector $(s_1, s_2, \cdots, s_r)$. Let v_i be the number of times that s_i is repeated;
Let $N := 0$, $P := 0$;
foreach $k \in [r]$ **do**
    Let $N \leftarrow N + v_k$, and $P \leftarrow P + v_k s_k$;
    if $\varepsilon < P$ **then**
        return $\left\lfloor \frac{\varepsilon - P}{s_k} + N \right\rfloor + 1$;
    end
end

Algorithm 4 Efficient Evaluation of the Threshold via Binary Search

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \text{Prob}^{\downarrow}(d)$, integer $n \ge 1$, $\varepsilon \in [0, 1)$

Output: $\ell(\mathbf{p}^{\otimes n}, \varepsilon)$

Compute the $r := \binom{n+d-1}{d-1}$ different terms $p_1^{n_1} p_2^{n_2} \cdots p_d^{n_d}$ where $n_1 + n_2 + \cdots + n_d = n$;
Sort the r terms in non-increasing order resulting in the vector $(s_1, s_2, \cdots, s_r)$. Let v_i be the number of times that s_i is repeated;
Let $N_0 := 0$, $P_0 := 0$, and $\forall k \in [r]$, $N_k := \sum_{i=1}^k v_i$ and $P_k := \sum_{i=1}^k v_i s_i$;
Let $a = 0$, $b = r$;
while *True* **do**
    Let $c = \lfloor (a + b)/2 \rfloor$;
    if $P_c \le \varepsilon < P_{c+1}$ **then**
        return $\left\lfloor \frac{\varepsilon - P_c}{s_{c+1}} + N_c \right\rfloor + 1$;
    else if $\varepsilon < P_c$ **then**
        $b \leftarrow c$;
    else
        $a \leftarrow c + 1$;
    end
end

Proof: The Proposition/Algorithm 3/Algorithm 4 are essentially variations of Lem. 17/Algorithm 1/Algorithm 2 and we will thus employ the notation used there. The principal idea is again that the elements of $\mathbf{p}^{\otimes n}$ repeat many times. After calculating again the ordered vector $(s_1, s_2, \cdots, s_r)$ and the number of times v_i that s_i repeats, we do a "rough" search and determine the integer j such that $P_j \le \varepsilon < P_{j+1}$, that is, determine which interval $\ell := \ell(\mathbf{p}^{\otimes n}, \varepsilon)$ falls into (remember that $\|\mathbf{p}^{\otimes n}\|_{(N_k)} = P_k$). This can be done efficiently and implies that $N_j < \ell \le N_{j+1}$. Moreover, according to Eq. (313), we know that for any m in this interval,

$$\begin{aligned} \|\mathbf{p}^{\otimes n}\|_{(m)} &= s_{j+1}(m - N_j) + P_j \\ &= s_{j+1}(m - N_{j+1}) + P_{j+1}, \end{aligned} \tag{314}$$

and thus

$$\ell = \left\lfloor \frac{\varepsilon - P_{j+1}}{s_{j+1}} + N_{j+1} \right\rfloor + 1. \tag{315}$$

This is exactly what Algorithm 3 returns. Algorithm 4 is again the variant employing binary search.

After we showed how to determine $\ell = \ell(\mathbf{p}^{\otimes n}, \varepsilon)$, according to Thm. 2, what remains to do in order to compute $\text{Distill}^{\varepsilon}\left(\psi^{\otimes n}\right)$ is to solve the optimization problem

$$\min_{k \in \{\ell, \ldots, d^n\}} \log \left\lfloor \frac{k}{\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon} \right\rfloor. \tag{316}$$

That this can be done efficiently is a consequence of the following Proposition.

Proposition 6: For any $\mathbf{p} \in \text{Prob}^{\downarrow}(d)$, integer $n \ge 1$, and $\varepsilon \in [0, 1)$, let $\ell = \ell(\mathbf{p}^{\otimes n}, \varepsilon)$. Then

$$f_{\min} := \min_{k \in \{\ell, \cdots, d^n\}} f(k) \quad \text{with} \quad f(k) := \frac{k}{\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon} \tag{317}$$

can be efficiently computed by Algorithm 5.

Proof: If $\ell = d^n$, the minimum is taken at ℓ. Otherwise, for any $k \in \{\ell, \cdots, d^n - 1\}$,

$$\begin{aligned} f(k+1) - f(k) &= \frac{k+1}{\|\mathbf{p}^{\otimes n}\|_{(k+1)} - \varepsilon} - \frac{k}{\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon} \\ &= \frac{\|\mathbf{p}^{\otimes n}\|_{(k)} - k(\mathbf{p}^{\otimes n})^{\downarrow}_{k+1} - \varepsilon}{(\|\mathbf{p}^{\otimes n}\|_{(k+1)} - \varepsilon)(\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon)}. \end{aligned} \tag{318}$$

Let $g(k) := \|\mathbf{p}^{\otimes n}\|_{(k)} - k(\mathbf{p}^{\otimes n})^{\downarrow}_{k+1} - \varepsilon$ and thus

$$g(k+1) - g(k) = (k+1)((\mathbf{p}^{\otimes n})^{\downarrow}_{k+1} - (\mathbf{p}^{\otimes n})^{\downarrow}_{k+2}) \ge 0. \tag{319}$$

This means that $g(k)$ is non-decreasing in k. Now consider three cases: If $g(\ell) \ge 0$, then $g(k) \ge 0$ for all $k \in \{\ell, \cdots, d^n - 1\}$. This implies that $f(k)$ is non-decreasing in $k \in \{\ell, \cdots, d^n - 1\}$ (remember that by definition of ℓ, $\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon > 0$). The minimum of $f(k)$ is taken at ℓ. If $g(d^n - 1) \le 0$, then $f(k)$ is non-increasing and the minimum of $f(k)$ is taken at d^n. It remains to consider the case that there is a $k^* \in \{\ell + 1, \cdots, d^n - 2\}$ such that $g(k^* - 1) < 0 \le g(k^*)$. In this case, it holds that for any $k \le k^* - 1$, $g(k) < 0$, and thus $f(k+1) < f(k)$. Moreover, for any $k \ge k^*$, $g(k) \ge 0$ and thus $f(k+1) \ge f(k)$. That is, $f(k)$ is decreasing in k for $k \in \{\ell, \cdots, k^*\}$ and non-decreasing in k for $k \in \{k^*, \cdots, d^n\}$. So the minimum of $f(k)$ is taken at k^*, which can be located via the bisection method. This takes at most $\log(d^n) = n \log d$ steps.

Theorem 4: Let $n \in \mathbb{N}$, $\varepsilon \in [0, 1)$, and $\psi \in \text{Pure}(AB)$. This implies that both $\text{Distill}^{\varepsilon}\left(\psi^{\otimes n}\right)$ and $\text{Cost}^{\varepsilon}\left(\psi^{\otimes n}\right)$ can be computed efficiently.

Algorithm 5 Efficient Evaluation of the Distillable Entanglement

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \mathrm{Prob}^{\downarrow}(d)$, integer $n \geq 1$, $\varepsilon \in [0, 1)$
Output: $f_{\min}$

Use Algorithm 3 to compute ℓ;
if $\ell = d^n$ **then**
  return $f_{\min} = \frac{d^n}{1-\varepsilon}$;
end
Use Algorithm 1 to compute $g(\ell)$;
if $g(\ell) \geq 0$ **then**
  return $f_{\min} = f(\ell)$ using Algorithm 1;
end
Use Algorithm 1 to compute $g(d^n - 1)$;
if $g(d^n - 1) \leq 0$ **then**
  return $f_{\min} = f(d^n) = \frac{d^n}{1-\varepsilon}$;
end
Let $a = \ell$ and $b = d^n - 1$;
while $b > a + 1$ **do**
  Let $c = \lfloor (a + b)/2 \rfloor$;
  if $g(c) \geq 0$ **then**
    $b \leftarrow c$;
  else
    $a \leftarrow c$;
  end
end
return $f_{\min} = f(b)$ using Algorithm 1;

Proof: That $\mathrm{Distill}^{\varepsilon}\left(\psi^{\otimes n}\right)$ can be computed efficiently follows from Thm. 2, Prop. 6, and the observation that

$$\min_{k \in \{\ell, \ldots, d^n\}} \log \left\lfloor \frac{k}{\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon} \right\rfloor = \log \left\lfloor \min_{k \in \{\ell, \ldots, d^n\}} \frac{k}{\|\mathbf{p}^{\otimes n}\|_{(k)} - \varepsilon} \right\rfloor . \quad (320)$$

Determining $\mathrm{Cost}^{\varepsilon}\left(\psi^{\otimes n}\right)$ is even easier: According to Thm. 3, we only need to determine the integer $m \in [d]$ satisfying $\|\mathbf{p}\|_{(m-1)} < 1 - \varepsilon \leq \|\mathbf{p}\|_{(m)}$. It is straightforward to see that this can be achieved by running a slight variation of Algorithm 3 which is presented in Algorithm 6. Of course, a similar adaptation is possible for Algorithm 4.

Appendix D
Closed-Form Formula for Ref. [14]

In this Appendix, we provide an efficient method to evaluate the fidelity of distillation defined in Eq. (71) as well as $E_D^{(1),\varepsilon}(\psi^{\otimes n})$. For an arbitrary pure state $|\psi\rangle \in \mathrm{Pure}(AB)$, let $\mathbf{p} = (p_1, \cdots, p_{|A|})$ be its Schmidt vector (ordered non-increasingly according to our convention). Moreover, let $\sqrt{\mathbf{p}} = (\sqrt{p_1}, \cdots, \sqrt{p_{|A|}})$. The fidelity of distillation is then given by [14]

$$F(\psi^{\otimes n}, m) = \frac{1}{m} \left\| \sqrt{\mathbf{p}^{\otimes n}} \right\|_{[m]}^2 , \quad (321)$$

where the distillation norm can be expressed as [62]

$$\| \sqrt{\mathbf{p}^{\otimes n}} \|_{[m]} := \|(\sqrt{\mathbf{p}})^{\otimes n}\|_{(m-k^*)} + \sqrt{k^*(1 - \|\mathbf{p}^{\otimes n}\|_{(m-k^*)})} \quad (322)$$

Algorithm 6 Efficient Evaluation of the Entanglement Cost

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \mathrm{Prob}^{\downarrow}(d)$, integer $n \geq 1$, $\varepsilon \in [0, 1)$
Output: m

Compute the $r := \binom{n+d-1}{d-1}$ different terms $p_1^{n_1} p_2^{n_2} \cdots p_d^{n_d}$ where $n_1 + n_2 + \cdots + n_d = n$;
Sort the r terms in non-increasing order resulting in the vector $(s_1, s_2, \cdots, s_r)$. Let v_i be the number of times that s_i is repeated;
Let $N := 0$, $P := 0$;
foreach $k \in [r]$ **do**
  Let $N \leftarrow N + v_k$, and $P \leftarrow P + v_k s_k$;
  if $1 - \varepsilon \leq P$ **then**
    return $\log \left\lceil \frac{1-\varepsilon-P}{s_k} + N \right\rceil$;
  end
end

with

$$k^* = \arg\min_{1 \leq k \leq m} \frac{1 - \|\mathbf{p}^{\otimes n}\|_{(m-k)}}{k} \quad (323)$$

and $\|\mathbf{p}^{\otimes n}\|_{(0)} := 0$. To determine the fidelity of distillation, we thus need to find k^* first.

Let

$$h(k) = \frac{1 - \|\mathbf{p}^{\otimes n}\|_{(m-k)}}{k}. \quad (324)$$

Then for any $k \in [m-1]$,

$$h(k+1) - h(k) = \frac{\|\mathbf{p}^{\otimes n}\|_{(m-k)} + k(\mathbf{p}^{\otimes n})^{\downarrow}_{m-k} - 1}{k(k+1)}. \quad (325)$$

Defining

$$\begin{aligned} t(k) &= \|\mathbf{p}^{\otimes n}\|_{(m-k)} + k(\mathbf{p}^{\otimes n})^{\downarrow}_{m-k} - 1 \\ &= (k+1)\|\mathbf{p}^{\otimes n}\|_{(m-k)} - k\|\mathbf{p}^{\otimes n}\|_{(m-k-1)} - 1, \end{aligned} \quad (326)$$

we have

$$t(k+1) - t(k) = (k+1)((\mathbf{p}^{\otimes n})^{\downarrow}_{m-k-1} - (\mathbf{p}^{\otimes n})^{\downarrow}_{m-k}) \geq 0. \quad (327)$$

So $t(k)$ is non-decreasing in $k \in [m-1]$. Analogously to the proof of Prop. 6, we can now consider three cases: If $t(1) \geq 0$, this implies that $h(k)$ is non-decreasing in $\{1, \cdots, m\}$ and $k^* = 1$. If $t(m-1) \leq 0$, $h(k)$ is non-increasing and we get $k^* = m$. Otherwise, there exists a $k' \in \{2, \cdots, m-2\}$ such that $t(k'-1) < 0 \leq t(k')$ and $h(k)$ is again decreasing in k for $k \in \{1, \cdots, k'\}$ and non-increasing for $k \in \{k', \cdots, m\}$. This implies that $k^* = k'$ which can be located via the bisection method. This takes at most $\log(d^n) = n \log d$ steps. The corresponding Algorithm is provided as Algorithm 7. Once k^* is determined, we can use Algorithm 1 to compute

$$\begin{aligned} &F(\psi^{\otimes n}, m) \\ &= \frac{1}{m} \left\| \sqrt{\mathbf{p}^{\otimes n}} \right\|_{[m]}^2 \\ &= \frac{1}{m} \left(\|(\sqrt{\mathbf{p}})^{\otimes n}\|_{(m-k^*)} + \sqrt{k^* \left(1 - \|\mathbf{p}^{\otimes n}\|_{(m-k^*)}\right)} \right)^2 . \end{aligned} \quad (328)$$

Algorithm 7 Efficient Determination of k^*

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \text{Prob}^{\downarrow}(d)$, integer $n \geq 1, m \geq 2$
Output: k^* necessary to determine $F(\psi^{\otimes n}, m)$

Use Algorithm 1 to compute $t(1)$;
if $t(1) \geq 0$ **then**
  return $k^* = 1$;
end
Use Algorithm 1 to compute $t(m-1)$;
if $t(m-1) \leq 0$ **then**
  return $k^* = m$;
end
Let $a = 1$ and $b = m - 1$;
while $b > a + 1$ **do**
  Let $c = \lfloor (a+b)/2 \rfloor$;
  if $t(c) \geq 0$ **then**
    $b \leftarrow c$;
  else
    $a \leftarrow c$;
  end
end
return $k^* = b$;

Algorithm 8 Efficient Evaluation of the Distillable Entanglement of [14]

Input: $\mathbf{p} = (p_1, \cdots, p_d) \in \text{Prob}^{\downarrow}(d)$, integer $n \geq 1$, $\varepsilon \in [0, 1)$
Output: $E_D^{(1),\varepsilon}(\psi^{\otimes n})$

if $F(\psi^{\otimes n}, 2) < 1 - \varepsilon$ **then**
  return 0;
end
if $F(\psi^{\otimes n}, \log \left\lfloor \frac{|A|^n}{(1-\varepsilon)^2} \right\rfloor) \geq 1 - \varepsilon$ **then**
  return $\log \log \left\lfloor \frac{|A|^n}{(1-\varepsilon)^2} \right\rfloor$;
end
Let $a = 2$ and $b = \log \left\lfloor \frac{|A|^n}{(1-\varepsilon)^2} \right\rfloor$;
while $b > a + 1$ **do**
  Let $c = \lfloor (a+b)/2 \rfloor$;
  if $F(\psi^{\otimes n}, c) < 1 - \varepsilon$ **then**
    $b \leftarrow c$;
  else
    $a \leftarrow c$;
  end
end
return $\log a$

According to [61, Lem. 1], the maximal achievable fidelity between a state obtained by applying LOCC to any initial state $\rho \in \mathfrak{D}(AB)$ with a maximally entangled state of dimension $m > |A|$ is given by $\sqrt{|A|/m}$. This implies that

$$\sqrt{\frac{|A|^n}{m}} \geq F(\psi^{\otimes n}, m), \tag{329}$$

and thus $E_D^{(1),\varepsilon}(\psi^{\otimes n}) \leq \log \left\lfloor \frac{|A|^n}{(1-\varepsilon)^2} \right\rfloor$. From this follows that Algorithm 8 efficiently computes $E_D^{(1),\varepsilon}(\psi^{\otimes n})$.

Acknowledgment

The authors thank an anonymous Referee for a simplification of the proof of Lem. 7. Thomas Theurer and Gilad Gour acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC). Thomas Theurer acknowledges support from the Postdoctoral Scholarship on Quantum Algorithms or Quantum Software from the Danish e-infrastructure Consortium (DeiC) and the Pacific Institute for the Mathematical Sciences (PIMS). The research and findings may not reflect those of the institute.

References

[1] M. B. Plenio and S. Virmani, "An introduction to entanglement measures," *Quantum Inf. Comput.*, vol. 7, no. 1, pp. 1–51, Jan. 2007.
[2] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Rev. Mod. Phys.*, vol. 81, no. 2, pp. 865–942, 2009.
[3] 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," *Phys. Rev. Lett.*, vol. 70, no. 13, pp. 1895–1899, Mar. 1993.
[4] C. H. Bennett and S. J. Wiesner, "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states," *Phys. Rev. Lett.*, vol. 69, no. 20, pp. 2881–2884, Nov. 1992.
[5] A. K. Ekert, "Quantum cryptography based on Bell's theorem," *Phys. Rev. Lett.*, vol. 67, no. 6, pp. 661–663, Aug. 1991.
[6] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," *Phys. Rev. A, Gen. Phys.*, vol. 54, no. 5, pp. 3824–3851, Nov. 1996.
[7] E. Chitambar and G. Gour, "Quantum resource theories," *Rev. Mod. Phys.*, vol. 91, no. 2, Apr. 2019, Art. no. 025001.
[8] M. A. Nielsen, "Conditions for a class of entanglement transformations," *Phys. Rev. Lett.*, vol. 83, no. 2, pp. 436–439, Jul. 1999.
[9] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A, Gen. Phys.*, vol. 53, no. 4, pp. 2046–2052, Apr. 1996.
[10] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, "Purification of noisy entanglement and faithful teleportation via noisy channels," *Phys. Rev. Lett.*, vol. 76, no. 5, pp. 722–725, Jan. 1996.
[11] E. M. Rains, "Bound on distillable entanglement," *Phys. Rev. A, Gen. Phys.*, vol. 60, no. 1, pp. 179–184, Jul. 1999.
[12] P. M. Hayden, M. Horodecki, and B. M. Terhal, "The asymptotic entanglement cost of preparing a quantum state," *J. Phys. A: Math. Gen.*, vol. 34, no. 35, pp. 6891–6898, Aug. 2001.
[13] K. Fang, X. Wang, M. Tomamichel, and R. Duan, "Non-asymptotic entanglement distillation," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6454–6465, Oct. 2019.
[14] B. Regula, K. Fang, X. Wang, and M. Gu, "One-shot entanglement distillation beyond local operations and classical communication," *New J. Phys.*, vol. 21, no. 10, Oct. 2019, Art. no. 103017.
[15] E. Zanoni, T. Theurer, and G. Gour, "Complete characterization of entanglement embezzlement," *Quantum*, vol. 8, p. 1368, Jun. 2024.
[16] W. Kumagai and M. Hayashi, "Second-order asymptotics of conversions of distributions and entangled states based on Rayleigh-normal probability distributions," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1829–1857, Mar. 2017.
[17] F. Buscemi and N. Datta, "Entanglement cost in practical scenarios," *Phys. Rev. Lett.*, vol. 106, no. 13, Mar. 2011, Art. no. 130503.
[18] N. Datta, "Min- and max-relative entropies and a new entanglement monotone," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2816–2826, Jun. 2009.
[19] M. Tomamichel, "A framework for non-asymptotic quantum information theory," 2012, *arXiv:1203.2142*.
[20] T. Theurer, D. Egloff, L. Zhang, and M. B. Plenio, "Quantifying operations with an application to coherence," *Phys. Rev. Lett.*, vol. 122, no. 19, May 2019, Art. no. 190405.

[21] Q. Zhuang, P. W. Shor, and J. H. Shapiro, "Resource theory of non-Gaussian operations," *Phys. Rev. A, Gen. Phys.*, vol. 97, no. 5, May 2018, Art. no. 052317.

[22] S. Bäuml, S. Das, X. Wang, and M. M. Wilde, "Resource theory of entanglement for bipartite quantum channels," 2019, *arXiv:1907.04181*.

[23] K. Fang, X. Wang, M. Tomamichel, and M. Berta, "Quantum channel simulation and the Channel's smooth max-information," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2129–2140, Apr. 2020.

[24] Y. Liu and X. Yuan, "Operational resource theory of quantum channels," *Phys. Rev. Res.*, vol. 2, no. 1, Feb. 2020, Art. no. 012035.

[25] X. Wang, M. M. Wilde, and Y. Su, "Quantifying the magic of quantum channels," *New J. Phys.*, vol. 21, no. 10, Oct. 2019, Art. no. 103002.

[26] Z.-W. Liu and A. Winter, "Resource theories of quantum channels and the universal role of resource erasure," 2019, *arXiv:1904.04201*.

[27] G. Gour and A. Winter, "How to quantify a dynamical quantum resource," *Phys. Rev. Lett.*, vol. 123, no. 15, Oct. 2019, Art. no. 150401.

[28] R. Takagi and B. Regula, "General resource theories in quantum mechanics and beyond: Operational characterization via discrimination tasks," *Phys. Rev. X*, vol. 9, no. 3, Sep. 2019, Art. no. 031053.

[29] G. Gour and C. M. Scandolo, "Dynamical entanglement," *Phys. Rev. Lett.*, vol. 125, no. 18, Oct. 2020, Art. no. 180505.

[30] G. Gour and C. M. Scandolo, "Entanglement of a bipartite channel," *Phys. Rev. A, Gen. Phys.*, vol. 103, no. 6, Jun. 2021, Art. no. 062422.

[31] K. Fang and Z.-W. Liu, "No-go theorems for quantum resource purification: New approach and channel theory," *PRX Quantum*, vol. 3, no. 1, Mar. 2022, Art. no. 010337.

[32] C. Li, B.-H. Wang, B. Wu, and X. Yuan, "Detecting entanglement of quantum channels," *Commun. Theor. Phys.*, vol. 73, no. 11, Sep. 2021, Art. no. 115101.

[33] H. Zhou, T. Gao, and F. Yan, "Quantifying the entanglement of quantum channel," *Phys. Rev. Res.*, vol. 4, no. 1, Mar. 2022, Art. no. 013200.

[34] E. M. Rains, "A semidefinite program for distillable entanglement," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2921–2933, Nov. 2001.

[35] K. Audenaert, M. B. Plenio, and J. Eisert, "Entanglement cost under Positive-Partial-Transpose-Preserving operations," *Phys. Rev. Lett.*, vol. 90, no. 2, Jan. 2003, Art. no. 027901.

[36] S. Das, S. Bäuml, and M. M. Wilde, "Entanglement and secret-key-agreement capacities of bipartite quantum interactions and read-only memory devices," *Phys. Rev. A, Gen. Phys.*, vol. 101, no. 1, Jan. 2020, Art. no. 012344.

[37] X. Wang and M. M. Wilde, "Exact entanglement cost of quantum states and channels under positive-partial-transpose-preserving operations," *Phys. Rev. A, Gen. Phys.*, vol. 107, no. 1, Jan. 2023, Art. no. 012429.

[38] H.-J. Kim, S. Lee, L. Lami, and M. B. Plenio, "One-shot manipulation of entanglement for quantum channels," *IEEE Trans. Inf. Theory*, vol. 67, no. 8, pp. 5339–5351, Aug. 2021.

[39] H.-J. Kim and S. Lee, "One-shot static entanglement cost of bipartite quantum channels," *Phys. Rev. A, Gen. Phys.*, vol. 103, no. 6, Jun. 2021, Art. no. 062415.

[40] B. Regula and R. Takagi, "Fundamental limitations on distillation of quantum channel resources," *Nature Commun.*, vol. 12, no. 1, p. 4411, Jul. 2021.

[41] B. Regula and R. Takagi, "One-shot manipulation of dynamical quantum resources," *Phys. Rev. Lett.*, vol. 127, no. 6, Aug. 2021, Art. no. 060402.

[42] X. Yuan, P. Zeng, M. Gao, and Q. Zhao, "One-shot dynamical resource theory," 2020, *arXiv:2012.02781*.

[43] R. Takagi, B. Regula, and M. M. Wilde, "One-shot yield-cost relations in general quantum resource theories," *PRX Quantum*, vol. 3, no. 1, Mar. 2022, Art. no. 010348.

[44] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Transforming quantum operations: Quantum supermaps," *EPL (Europhys. Lett.)*, vol. 83, no. 3, p. 30004, Aug. 2008.

[45] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications* (Springer Series in Statistics). Cham, Switzerland: Springer, 2011.

[46] M. Horodecki, J. Oppenheim, and C. Sparaciari, "Extremal distributions under approximate majorization," *J. Phys. A: Math. Theor.*, vol. 51, no. 30, Jun. 2018, Art. no. 305301.

[47] A. Gilchrist, N. K. Langford, and M. A. Nielsen, "Distance measures to compare real and ideal quantum processes," *Phys. Rev. A, Gen. Phys.*, vol. 71, no. 6, Jun. 2005, Art. no. 062310.

[48] M. Tomamichel, R. Colbeck, and R. Renner, "Duality between smooth min- and max-entropies," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4674–4681, Sep. 2010.

[49] G. Gour and M. Tomamichel, "Optimal extensions of resource measures and their applications," *Phys. Rev. A, Gen. Phys.*, vol. 102, no. 6, Dec. 2020, Art. no. 062401.

[50] C. T. Chubb, M. Tomamichel, and K. Korzekwa, "Beyond the thermodynamic limit: finite-size corrections to state interconversion rates," *Quantum*, vol. 2, p. 108, Nov. 2018.

[51] D. Jonathan and M. B. Plenio, "Entanglement-assisted local manipulation of pure quantum states," *Phys. Rev. Lett.*, vol. 83, no. 17, pp. 3566–3569, Oct. 1999.

[52] C. Datta, T. Varun Kondra, M. Miller, and A. Streltsov, "Catalysis of entanglement and other quantum resources," *Rep. Prog. Phys.*, vol. 86, no. 11, Oct. 2023, Art. no. 116002.

[53] T. V. Kondra, C. Datta, and A. Streltsov, "Catalytic transformations of pure entangled states," *Phys. Rev. Lett.*, vol. 127, no. 15, Oct. 2021, Art. no. 150503.

[54] S. Turgut, "Catalytic transformations for bipartite pure states," *J. Phys. A, Math. Theor.*, vol. 40, no. 40, pp. 12185–12212, Sep. 2007.

[55] M. Klimesh, "Inequalities that collectively completely characterize the catalytic majorization relation," 2007, *arXiv:0709.3680*.

[56] B. M. Terhal and P. Horodecki, "Schmidt number for density matrices," *Phys. Rev. A, Gen. Phys.*, vol. 61, no. 4, Mar. 2000, Art. no. 040301.

[57] M. Hayashi, *Quantum Information*. Cham, Switzerland: Springer, 2006.

[58] Q. Yue and E. Chitambar, "The zero-error entanglement cost is highly non-additive," *J. Math. Phys.*, vol. 60, no. 11, Nov. 2019, Art. no. 112204.

[59] G. Vidal, "Entanglement of pure states for a single copy," *Phys. Rev. Lett.*, vol. 83, no. 5, pp. 1046–1049, Aug. 1999.

[60] G. Bowen and N. Datta, "Entanglement cost for sequences of arbitrary quantum states," *J. Phys. A, Math. Theor.*, vol. 44, no. 4, Dec. 2010, Art. no. 045302.

[61] G. Vidal, D. Jonathan, and M. A. Nielsen, "Approximate transformations and robust manipulation of bipartite pure-state entanglement," *Phys. Rev. A, Gen. Phys.*, vol. 62, no. 1, Jun. 2000, Art. no. 012304.

[62] B. Regula, K. Fang, X. Wang, and G. Adesso, "One-shot coherence distillation," *Phys. Rev. Lett.*, vol. 121, no. 1, Jul. 2018, Art. no. 010401.

[63] X. Wang, K. Fang, and M. Tomamichel, "On converse bounds for classical communication over quantum channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4609–4619, Jul. 2019.

[64] F. Leditzky, E. Kaur, N. Datta, and M. M. Wilde, "Approaches for approximate additivity of the Holevo information of quantum channels," *Phys. Rev. A, Gen. Phys.*, vol. 97, no. 1, Jan. 2018, Art. no. 012332.

[65] M. Christandl, R. König, and R. Renner, "Postselection technique for quantum channels with applications to quantum cryptography," *Phys. Rev. Lett.*, vol. 102, no. 2, Jan. 2009, Art. no. 020504.

[66] M. Berta, F. G. S. L. Brandão, M. Christandl, and S. Wehner, "Entanglement cost of quantum channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6779–6795, Oct. 2013.

[67] E. Chitambar, D. Leung, L. Mančinska, M. Ozols, and A. Winter, "Everything you always wanted to know about LOCC (but were afraid to ask)," *Commun. Math. Phys.*, vol. 328, no. 1, pp. 303–326, May 2014.

[68] B. Schumacher and M. A. Nielsen, "Quantum data processing and error correction," *Phys. Rev. A, Gen. Phys.*, vol. 54, no. 4, pp. 2629–2635, 1996.

[69] F. Buscemi and N. Datta, "Distilling entanglement from arbitrary resources," *J. Math. Phys.*, vol. 51, no. 10, Oct. 2010, Art. no. 102201.

[70] I. Devetak and A. Winter, "Distillation of secret key and entanglement from quantum states," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 461, no. 2053, pp. 207–235, Jan. 2005.

[71] M. M. Wilde, M. Tomamichel, and M. Berta, "Converse bounds for private communication over quantum channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1792–1817, Mar. 2017.

[72] K. Fang and Z.-W. Liu, "No-go theorems for quantum resource purification," *Phys. Rev. Lett.*, vol. 125, no. 6, Aug. 2020, Art. no. 060405.

[73] T. Baumgratz, M. Cramer, and M. B. Plenio, "Quantifying coherence," *Phys. Rev. Lett.*, vol. 113, no. 14, Sep. 2014, Art. no. 140401.

[74] A. Streltsov, G. Adesso, and M. B. Plenio, "Colloquium: Quantum coherence as a resource," *Rev. Modern Phys.*, vol. 89, no. 4, Oct. 2017, Art. no. 041003.

[75] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens, and N. Yunger Halpern, "The resource theory of informational nonequilibrium in thermodynamics," *Phys. Rep.*, vol. 583, pp. 1–58, Jul. 2015.

[76] D. Janzing, P. Wocjan, R. Zeier, R. Geiss, and T. Beth, "Thermodynamic cost of reliability and low temperatures: Tightening Landauer's principle and the second law," *Int. J. Theor. Phys.*, vol. 39, no. 12, pp. 2717–2753, 2000.

[77] D. Egloff, O. C. O. Dahlsten, R. Renner, and V. Vedral, "A measure of majorization emerging from single-shot statistical mechanics," *New J. Phys.*, vol. 17, no. 7, Jul. 2015, Art. no. 073001.

[78] J. Aberg, “Truly work-like work extraction via a single-shot analysis,” *Nature Commun.*, vol. 4, no. 1, p. 1925, Jun. 2013.
[79] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, “Resource theory of quantum states out of thermal equilibrium,” *Phys. Rev. Lett.*, vol. 111, no. 25, Dec. 2013, Art. no. 250404.
[80] M. Lostaglio, “An introductory review of the resource theory approach to thermodynamics,” *Rep. Prog. Phys.*, vol. 82, no. 11, Oct. 2019, Art. no. 114001.
[81] Z. Zhang, “Estimating mutual information via Kolmogorov distance,” *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3280–3282, Sep. 2007.
[82] K. M. R. Audenaert, “A sharp continuity estimate for the von Neumann entropy,” *J. Phys. A, Math. Theoretical*, vol. 40, no. 28, p. 8127, 2007.
[83] H.-K. Lo and S. Popescu, “Concentrating entanglement by local actions: Beyond mean values,” *Phys. Rev. A, Gen. Phys.*, vol. 63, no. 2, Jan. 2001, Art. no. 022301.
[84] M. M. Wilde, A. Winter, and D. Yang, “Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy,” *Commun. Math. Phys.*, vol. 331, no. 2, pp. 593–622, Oct. 2014.
[85] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, “On quantum Rényi entropies: A new generalization and some properties,” *J. Math. Phys.*, vol. 54, no. 12, Dec. 2013, Art. no. 122203.
[86] M. Sion, “On general minimax theorems,” *Pacific J. Math.*, vol. 8, no. 1, pp. 171–176, Mar. 1958.
[87] J. Watrous, *The Theory of Quantum Information*. Cambridge, U.K.: Cambridge Univ. Press, 2018.
[88] L. P. Hughston, R. Jozsa, and W. K. Wootters, “A complete classification of quantum ensembles having a given density matrix,” *Phys. Lett. A*, vol. 183, no. 1, pp. 14–18, Nov. 1993.
[89] E. H. Lieb and M. B. Ruskai, “Proof of the strong subadditivity of quantum-mechanical entropy,” *J. Math. Phys.*, vol. 14, no. 12, pp. 1938–1941, Dec. 1973.
[90] E. H. Lieb and M. B. Ruskai, “A fundamental property of quantum-mechanical entropy,” *Phys. Rev. Lett.*, vol. 30, no. 10, pp. 434–436, Mar. 1973.
[91] A. Wehrl, “General properties of entropy,” *Rev. Modern Phys.*, vol. 50, no. 2, pp. 221–260, Apr. 1978.
[92] N. J. Cerf and C. Adami, “Negative entropy and information in quantum mechanics,” *Phys. Rev. Lett.*, vol. 79, no. 26, pp. 5194–5197, Dec. 1997.
[93] M. Horodecki, J. Oppenheim, and A. Winter, “Partial quantum information,” *Nature*, vol. 436, no. 7051, pp. 673–676, Aug. 2005.
[94] A. Winter, “Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints,” *Commun. Math. Phys.*, vol. 347, no. 1, pp. 291–313, Oct. 2016.
[95] M. Horodecki and P. Horodecki, “Reduction criterion of separability and limits for a class of distillation protocols,” *Phys. Rev. A, Gen. Phys.*, vol. 59, no. 6, pp. 4206–4216, Jun. 1999.

Thomas Theurer received the B.Sc., M.Sc., and Ph.D. degrees in physics from the University of Ulm, Ulm, Germany, in 2015, 2016, and 2021, respectively. From 2021 to 2024, he was a Post-Doctoral Fellow with the Department of Mathematics and Statistics, University of Calgary. Since 2024, he has been a Post-Doctoral Fellow with the Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark. His research interests include quantum information and quantum communication.

Kun Fang (Member, IEEE) received the bachelor’s degree in mathematics from Wuhan University in 2015 and the Ph.D. degree in quantum information from the University of Technology Sydney in 2018. He is currently an Assistant Professor with the School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen). Before joining CUHK-Shenzhen, he was a Senior Researcher and a Tech Lead with the Institute for Quantum Computing, Baidu, from 2020 to 2023. Prior to that, he was a Post-Doctoral Researcher with the University of Cambridge and the University of Waterloo from 2018 to 2020. His research interests include the mathematical foundations and optimization techniques in quantum information theory.

Gilad Gour received the Ph.D. degree in Physics from The Hebrew University of Jerusalem in 2002 under the supervision of Jacob Bekenstein. He is currently a Professor with the Faculty of Mathematics, Technion—Israel Institute of Technology. Prior to joining the Technion in 2024, he was a Professor with the Department of Mathematics and Statistics, University of Calgary, where he was a member of the Institute for Quantum Science and Technology. He then held post-doctoral positions at the University of Alberta and at the University of California at San Diego, working on quantum gravity and quantum information. His research focuses on the mathematical foundations of quantum information theory, with particular emphasis on quantum resource theories, quantum entropy, and operational aspects of quantum communication and thermodynamics.