Probabilistic distillation of quantum coherence

Kun Fang,[1,*] Xin Wang,[1,†] Ludovico Lami,[2,‡] Bartosz Regula,[2,§] and Gerardo Adesso[2,¶]

[1]*Centre for Quantum Software and Information, Faculty of Engineering and Information Technology, University of Technology Sydney, NSW 2007, Australia*

[2]*School of Mathematical Sciences and Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*

The ability to distill quantum coherence is pivotal for optimizing the performance of quantum technologies; however, such a task cannot always be accomplished with certainty. Here we develop a general framework of *probabilistic* distillation of quantum coherence in a one-shot setting, establishing fundamental limitations for different classes of free operations. We first provide a geometric interpretation for the maximal success probability, showing that under maximally incoherent operations (MIO) and dephasing-covariant incoherent operations (DIO) the problem can be simplified into efficiently computable semidefinite programs. Exploiting these results, we find that DIO and its subset of strictly incoherent operations (SIO) have equal power in probabilistic distillation of coherence from pure input states, while MIO are strictly stronger. We then prove a fundamental no-go result: distilling coherence from any full-rank state is impossible even probabilistically. We further find that in some conditions the maximal success probability can vanish suddenly beyond a certain threshold in the distillation fidelity. Finally, we consider probabilistic coherence distillation assisted by a catalyst and demonstrate, with specific examples, its superiority to the unassisted and deterministic cases.

arXiv:1804.09500v2 [quant-ph] 31 Jul 2018

Quantum coherence is a physical resource that is essential for various tasks in quantum computing (e.g. Deutsch-Jozsa algorithm [1]), cryptography (e.g. quantum key distribution [2]), information processing (e.g. quantum state merging [3], state redistribution [4] and channel simulation [5]), thermodynamics [6], metrology [7], and quantum biology [8]. A series of efforts have been devoted to building a resource framework of coherence in recent years [8–12], characterizing in particular the state transformations and operational uses of coherence in fundamental resource manipulation protocols [3, 13–17]. As in any physical resource theory, a central problem of the resource theory of quantum coherence is *distillation*: the process of extracting canonical units of coherence, as represented by the maximally coherent state $|\Psi_m\rangle$, from a given quantum state using a choice of free operations.

The usual asymptotic approach to studying the problem in quantum information theory is to assume that there is an unbounded number of independent and identically distributed copies of a quantum state available and that the transformation error asymptotically goes to zero [13, 18–20]. In reality, these assumptions become unphysical due to our limited access to a finite number of copies of a given state, making it necessary to look at non-asymptotic regimes [16, 17]. Furthermore, since loss and decoherence severely restrict our ability to manipulate large quantum systems, one needs to allow for a finite error in the distillation protocol. In this respect, deterministic protocols such as those studied in [17] may be insufficient to reach a target fidelity for desired applications. It is thus of importance to consider a more general *probabilistic* framework, in which the distillation will succeed only with some probability. Here, the allowed error can be characterized by two key parameters: the success probability of the one-shot distillation process, and the fidelity between the extracted state and the target state $|\Psi_m\rangle$. To have a systematic understanding of coherence distillation with finite resources and be able to implement practical schemes for this task, it is crucial to describe and optimize the interplay between these two parameters.

In this Letter, we develop the framework of probabilistic coherence distillation, characterizing the relation between the maximum success probability and the fidelity of distillation in the one-shot setting. We describe qualitative and quantitative aspects of this task under several representative choices of free operations, providing insights into their fundamental limitations and capabilities for coherence manipulation. In particular, we achieve a complete characterization of probabilistic coherence distillation with pure input states. The main results of our study are presented as Theorems in the following, with all proofs delegated to the Supplemental Material [21]. Before proceeding, we note that, previously, the framework of probabilistic state transformations has been employed in characterizing entanglement distillation [18, 26–29] as well as related settings in the resource theory of thermodynamics [30], and recently found use in the investigation of practical entanglement distillation schemes [31]. Our work fills an important gap in the literature by establishing the probabilistic toolbox for the key resource of quantum coherence.

Framework of probabilistic coherence distillation.—The free states in the resource theory of quantum coherence, so-called incoherent states $\mathcal{I}$, are the density operators which are diagonal in a given reference orthonormal basis $\{|i\rangle\}$. We will use $\Delta(\cdot) := \sum_i |i\rangle\langle i| \cdot |i\rangle\langle i|$ to denote the diagonal map (completely dephasing channel) in this basis. The resource theory of coherence is known not to admit a unique physically-motivated choice of allowed free operations [8, 13, 25, 32, 33], necessitating the investigation of operational capabilities of several different classes of maps. The relevant choices of free operations that we will focus on are: *maximally incoherent operations (MIO)* [9], defined to be all operations $\mathcal{E}$ such that $\mathcal{E}(\rho) \in \mathcal{I}$ for every $\rho \in \mathcal{I}$; *dephasing-covariant incoherent operations (DIO)* [25, 32], which are maps $\mathcal{E}$ such that $[\Delta, \mathcal{E}] = 0$, or equivalently $\mathcal{E}(|i\rangle\langle i|) \in \mathcal{I}$ and $\Delta(\mathcal{E}(|i\rangle\langle j|)) = 0, \forall i \neq j$; *incoherent operations (IO)* [12], which admit a set

of incoherent Kraus operators $\{K_l\}$ such that $\frac{K_l \rho K_l^\dagger}{\mathrm{Tr}\, K_l \rho K_l^\dagger} \in \mathcal{I}$ for all l and $\rho \in \mathcal{I}$; as well as *strictly incoherent operations (SIO)* [13], which are operations such that both $\{K_l\}$ and $\{K_l^\dagger\}$ are sets of incoherent operators. In particular, MIO is the largest possible choice of free operations in coherence theory, while SIO can be regarded as the smallest class which satisfies desirable resource theoretic criteria [8, 25], leading to the hierarchy SIO $\subsetneq$ IO $\subsetneq$ MIO, SIO $\subsetneq$ DIO $\subsetneq$ MIO [25].

The basic task of probabilistic distillation can be understood as follows. For any given quantum state ρ held by a single party A, we aim to transform this state to an m-dimensional maximally coherent state (target state) $|\Psi_m\rangle := m^{-1/2} \sum_{i=1}^m |i\rangle$ with high fidelity. A single-bit classical flag register L is used to indicate whether the transformation succeeds or not. If the flag is found in the 0 state, the distillation process has succeeded and the output state σ has fidelity at least $1-\varepsilon$ with the target state. Otherwise, the process has failed, and we discard the unwanted output state ω. Our goal is to maximize the success probability while keeping the transformation infidelity within some tolerance ε.

Formally, for any triplet (ρ, m, ε) with a given initial state ρ, target state dimension m, and infidelity tolerance ε, the maximal success probability of coherence distillation under the operation class $O \in \{\mathrm{SIO}, \mathrm{IO}, \mathrm{DIO}, \mathrm{MIO}\}$ is denoted as $P_O(\rho \to \Psi_m, \varepsilon)$, where $\Psi_m := |\Psi_m\rangle\langle\Psi_m|$. This is given by the maximal value of p such that there exists a transformation $\Pi_{A\to LB} \in O$ satisfying the constraints

$$\begin{aligned} &\Pi_{A\to LB}(\rho) = p|0\rangle\langle 0|_L \otimes \sigma + (1-p)|1\rangle\langle 1|_L \otimes \omega, \\ &F(\sigma, \Psi_m) \geq 1-\varepsilon, \end{aligned} \tag{1}$$

where $F(\rho,\sigma) := \| \sqrt{\rho}\sqrt{\sigma}\|_1^2$ is the fidelity and $\|\cdot\|_1$ is the trace norm. If the distillation fails, we can perform a free operation to make the unwanted state ω completely mixed without changing the success probability. Thus, without loss of generality, we can take $\omega = \mathbb{1}/m$. Exploiting the fact that the target state Ψ_m is invariant under the twirling operation $\mathcal{T}(\rho) = \frac{1}{d!}\sum_{i=1}^{d!} P_i \rho P_i$, where P_i are all the permutations on the input system and d is the input dimension, we can also fix the optimal output state as $\sigma = \Psi_m^\varepsilon$ where $\Psi_m^\varepsilon := (1-\varepsilon)\Psi_m + \varepsilon(\mathbb{1} - \Psi_m)/(m-1)$. Specifically, for any optimal output state σ, we can further perform the free operation $\mathcal{T}$, which gives a new output state $\mathcal{T}(\sigma)$ always in the form of $a\Psi_m + b(\mathbb{1}-\Psi_m)/(m-1)$, where we can choose $a = 1-\varepsilon$ and $b = \varepsilon$ while keeping the fidelity with the target state and the optimal success probability unchanged. This allows us to write $P_O(\rho \to \Psi_m, \varepsilon) = P_O(\rho \to \Psi_m^\varepsilon, 0)$, meaning that the maximal success probability of coherence distillation is the same as the maximal success probability of transforming the given state to the target Ψ_m^ε with fidelity one.

Computing the maximum distillation probability.— We now set out to find efficiently computable expressions for the maximal distillation probability. Consider a generalization of the set O to the class O_{sub} of subnormalized quantum operations, that is, completely positive and trace-nonincreasing maps. Using this notation, we can conveniently express the maximal success probability as follows (see also [29, 34]).

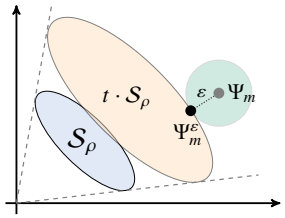


FIG. 1. Geometric interpretation of the maximal success probability of coherence distillation. See text for details.

Proposition 1 *For any triplet (ρ, m, ε) and operation class O, the maximal success probability $P_O(\rho \to \Psi_m, \varepsilon)$ is given by* $\max\left\{p \in \mathbb{R}_+ \,\middle|\, \mathcal{E}(\rho) = p \cdot \Psi_m^\varepsilon,\ \mathcal{E} \in O_{\mathrm{sub}}\right\}$. *It then holds that* $P_O(\rho \to \Psi_m, \varepsilon)^{-1} = \min\left\{t \in \mathbb{R}_+ \middle| \Psi_m^\varepsilon \in t \cdot \mathcal{S}_\rho\right\}$ *where* $\mathcal{S}_\rho := \left\{\mathcal{E}(\rho) \middle| \mathcal{E} \in O_{\mathrm{sub}}\right\}$ *is the set of all the output operators of ρ under the operation class O_{sub}.*

The result simplifies the optimization of the maximal success probability via subnormalized free operations, providing a geometric interpretation for the maximal success probability as a gauge function [35], as shown in Fig. 1. This justifies our intuition that the closer the state ρ to Ψ_m^ε, the less we need to expand the set $\mathcal{S}_\rho$, and thus the larger success probability we can obtain. We note that the result in Prop. 1 can be extended also to more general convex resource theories [36–38].

By further exploiting the symmetry of Ψ_m^ε, we can compute the maximal success probability under MIO/DIO via the following semidefinite programs (SDPs).

Theorem 2 *For any triplet (ρ, m, ε), the maximal success probability of distillation under MIO and DIO are*

$$\begin{aligned} P_{\mathrm{MIO}}(\rho \to \Psi_m, \varepsilon) = \max.\ & \mathrm{Tr}\, G\rho \\ \text{s.t.}\ & \Delta(G) = m\Delta(C), && \text{(2a)} \\ & 0 \leq C \leq G \leq \mathbb{1}, && \text{(2b)} \\ & \mathrm{Tr}\, C\rho \geq (1-\varepsilon)\,\mathrm{Tr}\, G\rho, && \text{(2c)} \\ P_{\mathrm{DIO}}(\rho \to \Psi_m, \varepsilon) = \max.\ & \mathrm{Tr}\, G\rho \\ \text{s.t.}\ & \text{Eqs. (2a, 2b, 2c)},\ G = \Delta(G). \end{aligned}$$

For completeness, we give the dual forms and alternative formulations of the SDPs in the Supplemental Material [21]. These SDPs provide us with efficient ways to numerically calculate the maximal success probability for general triplets (ρ, m, ε), and allow us to obtain key results on the power of different operations for probabilistic coherence distillation.

In this respect, let us also consider the choice of IO or SIO as the free operations. It is known that these two sets of operations have the same power in pure-state transformations, completely characterized by majorization relations.

This yields [25–27, 39, 40]

$$P_{(\mathrm{S})\mathrm{IO}}(\varphi \to \Psi_m, 0) = \begin{cases} 0 & \text{if } \operatorname{rank}\Delta(\varphi) < m, \\ \min\limits_{k\in[1,m]} \dfrac{m}{k} \displaystyle\sum_{i=m-k+1}^{d} \varphi_i^2 & \text{otherwise,} \end{cases}$$

where we have assumed without loss of generality that the coefficients of φ are non-negative and arranged in non-increasing order. In a similar way, operations in the class DIO can never increase the diagonal rank of a pure state, while it is known that MIO allow for the rank to increase [25], suggesting that MIO is a much stronger class. It is thus surprising that MIO and DIO have exactly the same power in the task of deterministic coherence distillation [17], and that the two sets of operations lead to the same asymptotic transformation rates for all states [41]. In the following, we will instead show crucial differences between MIO and DIO when one goes beyond deterministic transformations, highlighting the increased capabilities of MIO in probabilistic distillation, as well as establishing limitations on coherence distillation in general.

Theorem 3 *For any triplet $(\rho, m, 0)$ with a full-rank state ρ, it holds that $P_{\mathrm{MIO}}(\rho \to \Psi_m, 0) = 0$. For any triplet $(\varphi, m, 0)$ with a coherent pure state $|\varphi\rangle = \sum_{i=1}^{n} \varphi_i |i\rangle$, $\varphi_i \neq 0$, $n \geq 2$, it holds*

$$P_{\mathrm{MIO}}(\varphi \to \Psi_m, 0) \geq \frac{n^2}{m(\sum_{i=1}^{n} |\varphi_i|^{-2})} > 0. \tag{3}$$

This result establishes a no-go theorem for coherence distillation, showing that no class of free operations preserving incoherent states can allow to distill any perfect coherence from a full-rank state, even probabilistically. Note that any generic density matrix has full rank, and so does Ψ_m^{ε} for any $\varepsilon > 0$. Thus $|P_{\mathrm{MIO}}(\Psi_m^{\varepsilon} \to \Psi_m, 0) - P_{\mathrm{MIO}}(\Psi_m \to \Psi_m, 0)| = 1$, even though Ψ_m^{ε} can be arbitrarily close to Ψ_m, implying that the maximal success probability is *not continuous* with respect to the input state. The physical implications of this result are that any noise typically resulting in full-rank states will lead in practice to an irretrievable loss of resources. For example, in a scenario where the coherent state Ψ_m is stored in a quantum memory exposed to depolarizing noise, it is impossible to recover it perfectly via free operations, even probabilistically.

However, for any *pure* coherent state, Theorem 3 shows that it is always possible to probabilistically distill a maximally coherent state of arbitrary dimension via MIO. In the Supplemental Material [21] we establish a tighter bound for the probability of distillation under MIO, which in particular gives $P_{\mathrm{MIO}}(\Psi_n \to \Psi_m, 0) \geq \frac{n-1}{m-1}$ when $m > n$. Observe that instead $P_{\mathrm{DIO}}(\Psi_n \to \Psi_m, 0) = 0$ for $m > n$. This tells us that, as the dimension n increases, there are n-dimensional density matrices ρ_n such that $P_{\mathrm{MIO}}(\rho_n \to \Psi_{n+1}, 0) \to 1$ while $P_{\mathrm{DIO}}(\rho_n \to \Psi_{n+1}, 0) = 0$ for all n, i.e., P_{MIO} and P_{DIO} can exhibit an arbitrarily large gap. This shows that in the probabilistic distillation scenario, MIO can be much more powerful than DIO in general, in a stark contrast with the case of deterministic coherence distillation [17] (see Table I).

Deterministic distillation [17]	Pure states MIO = DIO = IO = SIO	General states MIO = DIO
Probabilistic distillation [⋆]	Pure states MIO > DIO = IO = SIO	Full-rank states MIO = DIO = IO = SIO = ∅

TABLE I. Comparison of the operational power of different sets of free operations for deterministic [17] versus probabilistic [⋆] (this paper) distillation of quantum coherence. ∅ denotes the empty set.

The relation between the capabilities of different operations is made precise by the following result, characterizing the fundamental task of distilling coherence from pure input states.

Theorem 4 *For any pure state φ and any m, we have $P_{\mathrm{DIO}}(\varphi \to \Psi_m, 0) = P_{(\mathrm{S})\mathrm{IO}}(\varphi \to \Psi_m, 0)$.*

We can therefore see that DIO does not provide any operational advantage over SIO and IO in pure-state probabilistic coherence distillation, despite being a strictly larger class than SIO [25, 42]. Putting together the results of Theorems 3 and 4, we have shown that a large gap in the operational capabilities of operations in the one-shot resource theory of quantum coherence exists between MIO and DIO, but not between DIO and SIO/IO — this can be compared with the case of deterministic distillation, where all sets of operations $O \in \{\mathrm{SIO}, \mathrm{IO}, \mathrm{DIO}, \mathrm{MIO}\}$ allow for the same achievable rate of distillation from pure states [17], see Table I.

In the task of distilling maximally coherent qubit states Ψ_2, we can extend the above result and obtain analytically the maximal success probability for arbitrary infidelity ε. In this particular case, MIO provides no advantage over DIO.

Proposition 5 *For $O \in \{DIO, MIO\}$ and any pure state φ with $\varphi_1 \geq \ldots \geq \varphi_n > 0$, it holds that*

$$P_O(\varphi \to \Psi_2, \varepsilon) = \begin{cases} 1 & \text{if } \varepsilon \geq \varepsilon_0(\varphi_1), \\ 2(1-\varphi_1^2)\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2 & \text{otherwise.} \end{cases}$$

Here the function ε_0, defined as $\varepsilon_0(\varphi_1) = 0$ if $\varphi_1 \leq \frac{1}{\sqrt{2}}$ and $\varepsilon_0(\varphi_1) = \frac{1}{2} - \varphi_1\sqrt{1-\varphi_1^2}$ otherwise, can be related to the m-distillation norm [17] characterizing the fidelity of deterministic distillation. Using this analytical result, we can give a concrete example to show that the probabilistic distillation framework can outperform the deterministic one. Suppose we need to distill a maximally coherent qubit state Ψ_2 from the input state $|\varphi\rangle = (3|0\rangle + |1\rangle)/\sqrt{10}$ with acceptable fidelity at least 0.9. The input state becomes useless in the deterministic scenario, since the maximal fidelity achievable via deterministic protocols is given by 0.8. However, probabilistic operations allow us to achieve the required distillation fidelity with $1/2$ success probability, demonstrating an explicit operational advantage of probabilistic distillation. In another scenario, if the acceptable fidelity is 0.8, we can gain higher distillation fidelity by compromising some success probability even though deterministic protocols are sufficient to accomplish the task.

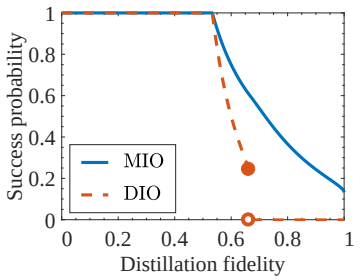


FIG. 2. Interplay between the fidelity $F = 1 - \varepsilon$ and the success probability p of coherence distillation for the example transformation $(|0\rangle + 3|1\rangle)/\sqrt{10} \to \Psi_3$. A discontinuity occurs at $F = 2/3$.

Such a setting may be dubbed "gambling with coherence", adapting terminology from [18, 26].

Relation between distillation fidelity and probability.—For any given input state ρ and target state dimension m, the maximal success probability is only dependent on the transformation fidelity. The higher the fidelity we require, the lower the probability that we will succeed. Intuitively, one would expect the success probability to decrease smoothly as the fidelity increases; however, we will now show that the success probability can vanish discontinuously beyond some fidelity threshold. This phenomenon is analogous to the strong converse theorem in channel coding theory [43–45], which says that the coding success probability goes to zero if the coding rate exceeds the capacity of the channel. While this phenomenon cannot occur in distillation from pure input states under MIO due to Theorem 3, in the following result we completely characterize this "sudden death" property for pure input states under DIO.

Proposition 6 *For any pure state* $|\varphi\rangle = \sum_{i=1}^{n} \varphi_i |i\rangle$ *with nonzero coefficients* φ_i*, it holds that*

$$P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) \begin{cases} > 0 & \text{if } n \geq m \text{ or if } n < m \text{ and } \varepsilon \geq 1 - \frac{n}{m}, \\ = 0 & \text{if } n < m \text{ and } \varepsilon < 1 - \frac{n}{m}. \end{cases}$$

In the particular case of the transformation $\Psi_n \to \Psi_m$ with $n \leq m$, the probability equals 1 as long as $\varepsilon \geq 1 - \frac{n}{m}$. The result shows in particular that, if the output dimension is larger than the input dimension, any trade-off between the maximal success probability and the transformation fidelity will always be truncated at the fidelity threshold $\frac{n}{m}$. Specifically, at the point $\varepsilon = 1 - \frac{n}{m}$, demanding a slightly higher fidelity will make the probabilistic distillation impossible, as shown in Fig. 2.

Probabilistic distillation with catalytic assistance.— A more general coherence distillation setting is to consider the scenario with catalytic assistance [28], where the input to the protocol consists of the resource state ρ together with another state γ (catalyst). As suggested by its name, we need to reproduce γ untouched in the output regardless of whether the distillation process succeeds or not. In [46], the authors studied catalytic coherence transformations without enforcing the preservation of the catalyst when the transformation fails — it is then not surprising that catalytic assistance improves the success probability, since we take the risk to sacrifice our catalyst. However, we can show that using catalysts can enhance probabilistic distillation even when we require them to be reproduced regardless of the outcome.

Formally, we denote the catalysis-assisted maximal success probability of coherence distillation under the operation class O as $P_O\big(\rho \xrightarrow{\gamma} \Psi_m, \varepsilon\big)$, which is given by the maximal value of p subject to the constraints

$$\begin{aligned} \Pi(\rho \otimes \gamma) &= \big(p|0\rangle\langle 0| \otimes \sigma + (1-p)|1\rangle\langle 1| \otimes \omega\big) \otimes \gamma, \\ F(\sigma, \Psi_m) &\geq 1 - \varepsilon, \quad \Pi \in O. \end{aligned} \tag{4}$$

Since we can always choose not to interact with the catalyst, it is clear that $P_O\big(\rho \xrightarrow{\gamma} \Psi_m, \varepsilon\big) \geq P_O(\rho \to \Psi_m, \varepsilon)$.

Taking as an example the two-qubit state $\rho = \frac{1}{2}(\nu_1 + \nu_2)$ with $|\nu_1\rangle = \frac{1}{2}(|00\rangle - |01\rangle - |10\rangle + |11\rangle)$ and $|\nu_2\rangle = \frac{1}{5\sqrt{2}}(2|00\rangle + 6|01\rangle - 3|10\rangle + |11\rangle)$, it turns out that the catalytic assistance of $\gamma = \Psi_2$ can enhance the success probability (by at least 12%) of distilling one coherent bit via DIO reliably ($\varepsilon \leq 0.01$) [21]. This example shows that the maximally coherent state can be used as a catalyst, manifesting a difference with the case of deterministic state transformation, where no transformation can be catalyzed by a maximally coherent state [28, 40]. We further note that if we allow a small perturbation of the catalyst to be returned in the protocol, one may obtain an even higher success probability as shown in the Supplemental Material [21] — such a setting has been considered e.g. in [47] for the resource theory of thermodynamics.

Conclusions.— We have developed a general framework of probabilistic coherence distillation. We interpreted the fundamental relations between the distillation fidelity and the maximal success probability via a gauge function construction, and showed that the maximal success probability under MIO and DIO can be efficiently computed via semidefinite programming. We proved that distilling perfect coherence from any full-rank state is *impossible* even probabilistically, while any pure coherent state can always be perfectly distilled with MIO into a maximally coherent state of arbitrary dimension with a non-zero probability, highlighting an operational advantage of MIO over other sets of operations, in contrast with the deterministic case. On the other hand, we found that DIO provides no operational advantage over SIO in pure-state distillation, with the maximal achievable distillation probability being equal under the two classes of operations. We provided an analytical characterization of distillation with pure input states and in particular described the distillation of qubit maximally coherent states under MIO and DIO. We further explored novel phenomena of coherence distillation such as the breakdown of the trade-off between maximal success probability and fidelity under a certain threshold as well as the catalyst-assisted enhancement by maximally coherent states.

Our work establishes fundamental limitations to the processing of quantum coherence in realistic settings, opening new perspectives for its investigation and exploitation as a resource in quantum information processing and quantum technology [1–8]. It would be of interest to analyze as well

the task of probabilistic coherence dilution under different free operations, complementing the deterministic case studied in [16]. Another interesting perspective for future work may be to apply the framework of probabilistic distillation developed here to the study of other important resource theories, such as those of asymmetry, magic, and thermodynamics [38].

Acknowledgments.— We would like to thank Gilad Gour, Filip Rozpędek, Alexander Streltsov, Marco Tomamichel, and Yunlong Xiao for fruitful discussions. We acknowledge financial support from the European Research Council (ERC) under the Starting Grant GQCOP (Grant No. 637352).

* kun.fang-1@student.uts.edu.au
† xin.wang-8@student.uts.edu.au
‡ ludovico.lami@gmail.com
§ bartosz.regula@gmail.com
¶ gerardo.adesso@nottingham.ac.uk

[1] M. Hillery, Phys. Rev. A **93**, 012111 (2016).
[2] P. J. Coles, E. M. Metodiev, and N. Lütkenhaus, Nat. Commun. **7**, 11712 (2016).
[3] A. Streltsov, E. Chitambar, S. Rana, M. Bera, A. Winter, and M. Lewenstein, Phys. Rev. Lett. **116**, 240405 (2016).
[4] A. Anshu, R. Jain, and A. Streltsov, arXiv:1804.04915 (2018).
[5] M. G. Díaz, K. Fang, X. Wang, M. Rosati, M. Skotiniotis, J. Calsamiglia, and A. Winter, arXiv:1805.04045 (2018).
[6] M. Lostaglio, D. Jennings, and T. Rudolph, Nat. Commun. **6**, 6383 (2015).
[7] F. Fröwis and W. Dür, Phys. Rev. Lett. **106**, 110402 (2011), 1012.4472.
[8] A. Streltsov, G. Adesso, and M. B. Plenio, Rev. Mod. Phys. **89**, 041003 (2017).
[9] J. Aberg, (2006), arXiv:quant-ph/0612146.
[10] G. Gour and R. W. Spekkens, New J. Phys. **10**, 033023 (2008).
[11] F. Levi and F. Mintert, New J. Phys. **16**, 033007 (2014).
[12] T. Baumgratz, M. Cramer, and M. B. Plenio, Phys. Rev. Lett. **113**, 140401 (2014).
[13] A. Winter and D. Yang, Phys. Rev. Lett. **116**, 120404 (2016).
[14] E. Chitambar, A. Streltsov, S. Rana, M. Bera, G. Adesso, and M. Lewenstein, Phys. Rev. Lett. **116**, 070402 (2016).
[15] E. Chitambar and M.-H. Hsieh, Phys. Rev. Lett. **117**, 020402 (2016).
[16] Q. Zhao, Y. Liu, X. Yuan, E. Chitambar, and X. Ma, Phys. Rev. Lett. **120**, 070403 (2018).
[17] B. Regula, K. Fang, X. Wang, and G. Adesso, Phys. Rev. Lett. **121**, 010401 (2018).
[18] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Phys. Rev. A **53**, 2046 (1996).
[19] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Phys. Rev. Lett. **76**, 722 (1996).
[20] E. M. Rains, Phys. Rev. A **60**, 173 (1999).
[21] See the Supplemental Material for technical derivations and proofs of the results in the Letter, which includes Refs. [22-25].
[22] F. Zhang, ed., *The Schur Complement and Its Applications* (Springer, 2005).
[23] R. Bhatia, *Positive Definite Matrices* (Princeton University Press, 2007).
[24] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed., CMS Books in Mathematics (Springer-Verlag, New York, 2003).
[25] E. Chitambar and G. Gour, Phys. Rev. Lett. **117**, 030401 (2016).
[26] H.-K. Lo and S. Popescu, Phys. Rev. A **63**, 022301 (2001).
[27] G. Vidal, Phys. Rev. Lett. **83**, 1046 (1999).
[28] D. Jonathan and M. B. Plenio, Phys. Rev. Lett. **83**, 3566 (1999).
[29] S. Ishizaka and M. B. Plenio, Phys. Rev. A **71**, 052303 (2005).
[30] A. M. Alhambra, J. Oppenheim, and C. Perry, Phys. Rev. X **6**, 041016 (2016).
[31] F. Rozpędek, T. Schiet, L. P. Thinh, D. Elkouss, A. C. Doherty, and S. Wehner, Phys. Rev. A **97**, 062333 (2018).
[32] I. Marvian and R. W. Spekkens, Phys. Rev. A **94**, 052324 (2016).
[33] J. I. de Vicente and A. Streltsov, J. Phys. A: Math. Theor. **50**, 045301 (2017).
[34] F. Buscemi and G. Gour, Phys. Rev. A **95**, 012110 (2017).
[35] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970).
[36] F. G. S. L. Brandão and G. Gour, Phys. Rev. Lett. **115**, 070503 (2015).
[37] B. Regula, J. Phys. A: Math. Theor. **51**, 045303 (2018).
[38] E. Chitambar and G. Gour, (2018), arXiv:1806.06107.
[39] H. Zhu, Z. Ma, Z. Cao, S.-M. Fei, and V. Vedral, Phys. Rev. A **96**, 032316 (2017).
[40] S. Du, Z. Bai, and Y. Guo, Phys. Rev. A **91**, 052120 (2015).
[41] E. Chitambar, Phys. Rev. A **97**, 050301 (2018).
[42] K. Bu and C. Xiong, Quant Inf Comput **13**, 1206 (2017), arXiv:1604.06524.
[43] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge (Springer-Verlag, Berlin Heidelberg, 1978).
[44] T. Ogawa and H. Nagaoka, IEEE Trans. Inf. Theory **45**, 2486 (1999).
[45] A. Winter, IEEE Trans. Inf. Theory **45**, 2481 (1999).
[46] K. Bu, U. Singh, and J. Wu, Phys. Rev. A **93**, 042326 (2016).
[47] F. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, PNAS **112**, 3275 (2015).

Supplemental Material: Probabilistic distillation of quantum coherence

This supplemental material provides more detailed analysis and proofs of the results in the main text. We may reiterate some of the steps in the main text to make the supplemental material more explicit and self-contained.

PROOF OF PROPOSITION 1

For any triplet (ρ, m, ε), the definition of the maximal success probability of coherence distillation is defined by

$$\begin{aligned} P_O(\rho \to \Psi_m, \varepsilon) = \max\ & p && \text{(S1a)}\\ \text{s.t. } & \Pi_{A\to LB}(\rho) = p|0\rangle\langle 0|_L \otimes \sigma + (1-p)|1\rangle\langle 1|_L \otimes \omega, && \text{(S1b)}\\ & F(\sigma, \Psi_m) \geq 1-\varepsilon,\ \Pi_{A\to LB} \in O,\ 0 \leq p \leq 1, && \text{(S1c)} \end{aligned}$$

where $F(\rho,\sigma) := \| \sqrt{\rho}\sqrt{\sigma}\|_1^2$ is the fidelity and $\|\cdot\|_1$ is the trace norm.

Remark 1 *The optimal solution in the optimization (S1) can always achieve the equality in the condition $F(\rho, \Psi_m) \geq 1-\varepsilon$. That is $P_O(\rho \to \Psi_m, \varepsilon) = \widetilde{P}_O(\rho \to \Psi_m, \varepsilon)$ with*

$$\begin{aligned} \widetilde{P}_O(\rho \to \Psi_m, \varepsilon) = \max\ & p && \text{(S2a)}\\ \text{s.t. } & \Pi_{A\to LB}(\rho) = p|0\rangle\langle 0|_L \otimes \sigma + (1-p)|1\rangle\langle 1|_L \otimes \omega, && \text{(S2b)}\\ & F(\sigma, \Psi_m) = 1-\varepsilon,\ \Pi_{A\to LB} \in O,\ 0 \leq p \leq 1. && \text{(S2c)} \end{aligned}$$

It is clear that $P_O(\rho \to \Psi_m, \varepsilon) \geq \widetilde{P}_O(\rho \to \Psi_m, \varepsilon)$ since $P_O(\rho \to \Psi_m, \varepsilon)$ is maximizing over a larger feasible set. However, for any optimal solution in (S1) such that $F(\sigma, \Psi_m) \geq 1-\varepsilon$, we can further apply a depolarizing noise (free operation) to make the fidelity between ρ and Ψ_m decrease exactly to $1-\varepsilon$, which gives a feasible solution for (S2). Thus it also holds $P_O(\rho \to \Psi_m, \varepsilon) \leq \widetilde{P}_O(\rho \to \Psi_m, \varepsilon)$. In the following, we will equivalently use both optimizations (S1) and (S2).

Remark 2 *Denote the twirling operation $\mathcal{T}(\rho) = \frac{1}{d!}\sum_{i=1}^{d!} P_i \rho P_i$, where P_i are all the permutations on the input system and d is the input dimension. If the distillation fails, we can perform a free operation $\mathcal{T} \circ \Delta \in \mathrm{SIO}$ to make the unwanted state ω completely mixed without changing the success probability. Thus, without loss of generality, we can take $\omega = \mathbb{1}/m$. On the other hand, if the distillation process succeeds and we obtain an output state σ such that $F(\sigma, \Psi_m) = 1-\varepsilon$, we can further perform the free operation $\mathcal{T}$ which gives a new output state $\mathcal{T}(\sigma)$ always in the form of $a\Psi_m + b(\mathbb{1}-\Psi_m)/(m-1)$. Since $F(\sigma, \Psi_m) = 1-\varepsilon$, we have $\mathcal{T}(\sigma) = \Psi_m^\varepsilon$ with $\Psi_m^\varepsilon := (1-\varepsilon)\Psi_m + \varepsilon\frac{\mathbb{1}-\Psi_m}{m-1}$. Thus, without loss of generality, we can take $\sigma = \Psi_m^\varepsilon$. Finally, the maximal success probability of coherence distillation can be given by*

$$\begin{aligned} P_O(\rho \to \Psi_m, \varepsilon) = \max\ & p && \text{(S3a)}\\ \text{s.t. } & \Pi_{A\to LB}(\rho) = p|0\rangle\langle 0|_L \otimes \Psi_m^\varepsilon + (1-p)|1\rangle\langle 1|_L \otimes \mathbb{1}/m, && \text{(S3b)}\\ & \Pi_{A\to LB} \in O,\ 0 \leq p \leq 1, && \text{(S3c)} \end{aligned}$$

Proposition 1 *For any triplet (ρ, m, ε) and operation class O, the maximal success probability $P_O(\rho \to \Psi_m, \varepsilon)$ is given by*

$$\max\left\{ p \in \mathbb{R}_+ \,\middle|\, \mathcal{E}(\rho) = p \cdot \Psi_m^\varepsilon,\ \mathcal{E} \in O_{\mathrm{sub}} \right\}. \tag{S4}$$

As a consequence, it holds that $P_O(\rho \to \Psi_m, \varepsilon)^{-1} = \min\left\{ t \in \mathbb{R}_+ \,\middle|\, \Psi_m^\varepsilon \in t \cdot \mathcal{S}_\rho \right\}$ where $\mathcal{S}_\rho := \left\{ \mathcal{E}(\rho) \,\middle|\, \mathcal{E} \in O_{\mathrm{sub}} \right\}$ is the set of all the output operators of ρ under the operation class O_{sub}.

Proof. For any quantum operation $\Pi_{A\to LB}$ such that $\Pi_{A\to LB}(\rho) = |0\rangle\langle 0|_L \otimes \mathcal{E}_0(\rho) + |1\rangle\langle 1|_L \otimes \mathcal{E}_1(\rho)$ where $\mathcal{E}_0$ and $\mathcal{E}_1$ are two subnormalized operations, we can check that $\Pi_{A\to LB} \in O$ if and only if $\mathcal{E}_0, \mathcal{E}_1 \in O_{\mathrm{sub}}$ and $\mathcal{E}_0 + \mathcal{E}_1$ is trace preserving. Thus finding the optimal solution in the optimization (S3) is equivalent to find the optimal subnormalized operations $\mathcal{E}_0$ and $\mathcal{E}_1$ such that $\mathcal{E}_0(\rho) = p \cdot \Psi_m^\varepsilon$, $\mathcal{E}_1(\rho) = (1-p)\mathbb{1}/m$ and $\mathcal{E}_0 + \mathcal{E}_1$ trace-preserving. Since we can always take $\mathcal{E}_1(\rho) = (\mathrm{Tr}\,\rho - \mathrm{Tr}\,\mathcal{E}_0(\rho))\mathbb{1}/m$ without compromising the success probability, the maximal success probability of coherence distillation is only dependent on $\mathcal{E}_0$, and the result follows. ∎

PROOF OF THEOREM 2

Theorem 2 *For any triplet (ρ, m, ε), the maximal success probability of distillation under MIO and DIO are given by*

$$\begin{aligned} P_{\mathrm{MIO}}(\rho \to \Psi_m, \varepsilon) = \underset{G,C}{\text{maximize}} \quad & \operatorname{Tr} G\rho && \text{(S5a)}\\ \text{subject to} \quad & \Delta(G) = m\Delta(C), && \text{(S5b)}\\ & 0 \le C \le G \le \mathbb{1}, && \text{(S5c)}\\ & \operatorname{Tr} C\rho \ge (1-\varepsilon)\operatorname{Tr} G\rho, && \text{(S5d)} \end{aligned}$$

and

$$\begin{aligned} P_{\mathrm{DIO}}(\rho \to \Psi_m, \varepsilon) = \underset{G,C}{\text{maximize}} \quad & \operatorname{Tr} G\rho && \text{(S6a)}\\ \text{subject to} \quad & \Delta(G) = m\Delta(C), && \text{(S6b)}\\ & 0 \le C \le G \le \mathbb{1}, && \text{(S6c)}\\ & \operatorname{Tr} C\rho \ge (1-\varepsilon)\operatorname{Tr} G\rho, && \text{(S6d)}\\ & G = \Delta(G). && \text{(S6e)} \end{aligned}$$

Proof. We show the proof for MIO first. Denote $J_{\mathcal{N}}$ as the Choi-Jamiołkowski matrix of operation $\mathcal{N}$. Due to the Choi-Jamiołkowski isomorphism, we can write the optimization (S4) in terms of Choi-Jamiołkowski matrix,

$$\begin{aligned} P_{\mathrm{MIO}}(\rho \to \Psi_m, \varepsilon) = \max\ & p && \text{(S7a)}\\ \text{s.t.}\ & \operatorname{Tr}_A J_{\mathcal{E}}(\rho^T \otimes \mathbb{1}_B) = p \cdot \Psi_m^{\varepsilon}, && \text{(S7b)}\\ & J_{\mathcal{E}} \ge 0, \operatorname{Tr}_B J_{\mathcal{E}} \le \mathbb{1}_A, && \text{(S7c)}\\ & \operatorname{Tr}_A J_{\mathcal{E}}(|i\rangle\langle i|^T \otimes \mathbb{1}_B) = \Delta(\operatorname{Tr}_A J_{\mathcal{E}}(|i\rangle\langle i|^T \otimes \mathbb{1}_B)),\ \forall i. && \text{(S7d)} \end{aligned}$$

For any optimal subnormalized quantum operation $\mathcal{E}$ in optimization (S4), the operation $\widetilde{\mathcal{E}} = \mathcal{T} \circ \mathcal{E}$ is also optimal since Ψ_m^{ε} is invariant under the twirling operation $\mathcal{T}$. Thus $J_{\widetilde{\mathcal{E}}}$ admits the structure of $J_{\widetilde{\mathcal{E}}} = C^T \otimes \Psi_m + D^T \otimes (\mathbb{1} - \Psi_m)$ for some operators C and D. Taking this form of Choi-Jamiołkowski matrix into the conditions of the optimization (S7), we have

$$\begin{aligned} P_{\mathrm{MIO}}(\rho \to \Psi_m, \varepsilon) = \max\ & p && \text{(S8a)}\\ \text{s.t.}\ & \operatorname{Tr} C\rho = p(1-\varepsilon),\ \operatorname{Tr}(m-1)D\rho = p\varepsilon, && \text{(S8b)}\\ & C \ge 0,\ D \ge 0,\ C + (m-1)D \le \mathbb{1}, && \text{(S8c)}\\ & \Delta(C) = \Delta(D). && \text{(S8d)} \end{aligned}$$

Denoting $G := C + (m-1)D$ and eliminating variable D, we will obtain the result (S5). Note that the last condition in (S5) can be taken as equality due to the argument in Remark 1. Following the similar steps, we can obtain the optimization for DIO. ∎

Remark 3 *From the conditions $G = \Delta(G) = m\Delta(C)$, we can eliminate the variable G in the optimization (S6) and obtain*

$$\begin{aligned} P_{\mathrm{DIO}}(\rho \to \Psi_m, \varepsilon) = \underset{C}{\text{maximize}} \quad & \operatorname{Tr} m\Delta(C)\rho && \text{(S9a)}\\ \text{subject to} \quad & 0 \le C \le m\Delta(C) \le \mathbb{1}, && \text{(S9b)}\\ & \operatorname{Tr} C\rho \ge m(1-\varepsilon)\operatorname{Tr}\Delta(C)\rho. && \text{(S9c)} \end{aligned}$$

Remark 4 *We will make repeated use of the semidefinite programs given in Theorem 2 and their Lagrange duals. We report the dual forms below. The dual of (S5) is given by*

$$\begin{aligned} P_{\mathrm{MIO}}(\rho \to \Psi_m, \varepsilon) = \underset{X,Y,Z,\lambda}{\text{minimize}} \quad & Tr\, X\\ \text{subject to} \quad & [1 - \lambda(1-\varepsilon)]\,\rho + Y - X + \Delta(Z) \le 0,\\ & \lambda\rho - Y - m\Delta(Z) \le 0,\\ & X, Y \ge 0,\\ & \lambda \in \mathbb{R}_+, \end{aligned} \qquad \text{(S10)}$$

The dual of (S9) *is given by*

$$\begin{aligned} P_{\mathrm{DIO}}(\rho\to\Psi_m,\varepsilon) = \underset{X,Y,\lambda}{\text{minimize}} \quad & Tr\, X \\ \text{subject to} \quad & m\Delta(\rho)+m\Delta(Y)-Y-m\Delta(X)+\lambda\rho-m(1-\varepsilon)\lambda\Delta(\rho)\le 0, \\ & X,Y\ge 0, \\ & \lambda\in\mathbb{R}_+. \end{aligned} \tag{S11}$$

To see that strong duality holds, it suffices to note the existence of strictly feasible solutions: in (S10)*, take* $\lambda=\frac{1}{1-\varepsilon}, X=3\lambda\mathbb{1}, Y=2\lambda\mathbb{1}, Z=0$*, and in* (S11)*, take* $\lambda=\frac{1}{1-\varepsilon}, X=3\lambda\mathbb{1}, Y=\mathbb{1}$*.*

Remark 5 *In the case* $\varepsilon=0$*, the primal problems can alternatively be expressed as*

$$\begin{aligned} P_{\mathrm{MIO}}(\rho\to\Psi_m,0) = \underset{G,C}{\text{maximize}} \quad & Tr\, G\rho \\ \text{subject to} \quad & C\ge 0, \\ & G\le\mathbb{1}, \\ & \Delta(G)=m\Delta(C), \\ & G-C\in\left\{A\,\middle|\,\mathrm{supp}(A)\subseteq\ker(\rho)\right\}\cap\mathbb{H}_+. \end{aligned} \tag{S12}$$

$$\begin{aligned} P_{\mathrm{DIO}}(\rho\to\Psi_m,0) = \underset{C}{\text{maximize}} \quad & Tr\, m\Delta(C)\rho \\ \text{subject to} \quad & C\ge 0, \\ & m\Delta(C)\le\mathbb{1}, \\ & m\Delta(C)-C\in\left\{A\,\middle|\,\mathrm{supp}(A)\subseteq\ker(\rho)\right\}\cap\mathbb{H}_+, \end{aligned} \tag{S13}$$

where we have used $\mathbb{H}_+$ *to denote the set of positive semidefinite matrices. This gives the duals*

$$\begin{aligned} P_{\mathrm{MIO}}(\rho\to\Psi_m,0) = \underset{W,Z}{\text{minimize}} \quad & Tr\,(W+Z)+1 \\ \text{subject to} \quad & \rho+W+\Delta(Z)\ge 0, \\ & W+m\Delta(Z)\ge 0, \\ & W\in\left\{A\,\middle|\,\mathrm{supp}(A)\subseteq\mathrm{supp}(\rho)\right\}+\mathbb{H}_+, \end{aligned} \tag{S14}$$

$$\begin{aligned} P_{\mathrm{DIO}}(\rho\to\Psi_m,0) = \underset{X,W}{\text{minimize}} \quad & Tr\, X \\ \text{subject to} \quad & m\Delta(\rho)+W-m\Delta(W)-m\Delta(X)\le 0, \\ & X\ge 0, \\ & W\in\left\{A\,\middle|\,\mathrm{supp}(A)\subseteq\mathrm{supp}(\rho)\right\}+\mathbb{H}_+. \end{aligned} \tag{S15}$$

PROOF OF THEOREM 3

To prove Theorem 3 in the main text, we will establish a stronger version of the result as follows.

Theorem S1 *For any triplet* $(\rho,m,0)$ *with full-rank state* ρ*, it holds that* $P_{\mathrm{MIO}}(\rho\to\Psi_m,0)=0$*. For any triplet* $(\varphi,m,0)$ *with coherent pure state* $|\varphi\rangle=\sum_{i=1}^n\varphi_i|i\rangle$ *and* $\varphi_i\neq 0$*,* $n\ge 2$*, it holds that*

$$P_{\mathrm{MIO}}(\varphi\to\Psi_m,0)\ge\frac{n^2}{\sum_{i=1}^n|\varphi_i|^{-2}}\left\|\frac{n-m}{n-1}\widetilde{\varphi}+\frac{n(m-1)}{n-1}\Delta(\widetilde{\varphi})\right\|_\infty^{-1}\ge\frac{n^2}{m(\sum_{i=1}^n|\varphi_i|^{-2})}>0 \tag{S16}$$

where

$$|\widetilde{\varphi}\rangle:=\frac{1}{\sqrt{s}}\sum_{i=1}^n\frac{\varphi_i}{|\varphi_i|^2}|i\rangle\quad\text{with}\quad s=\sum_{j=1}^n|\varphi_j|^{-2}. \tag{S17}$$

Proof. Recall the SDP

$$P_{\mathrm{MIO}}(\rho \to \Psi_m, \varepsilon) = \max \left\{ \operatorname{Tr} G\rho \;\middle|\; \Delta(G) = m\Delta(C),\ 0 \le C \le G \le \mathbb{1}, \operatorname{Tr} C\rho = (1-\varepsilon)\operatorname{Tr} G\rho \right\}. \tag{S18}$$

For the first argument, we know that $G - C \ge 0$ and $\operatorname{Tr}(G-C)\rho = 0$. Since ρ is full-rank, we have $G = C$. Together with $\Delta(G) = m\Delta(C)$, we have $G = C = 0$ and $P_{\mathrm{MIO}}(\rho \to \Psi_m, 0) = 0$.

As for the second argument, let us choose

$$C = c\widetilde{\varphi}, \qquad G = c\widetilde{\varphi} + \frac{(m-1)c}{n-1}\left(n\Delta(\widetilde{\varphi}) - \widetilde{\varphi}\right), \tag{S19}$$

where

$$c = \left\| \frac{n-m}{n-1}\widetilde{\varphi} + \frac{n(m-1)}{n-1}\Delta(\widetilde{\varphi}) \right\|_\infty^{-1}. \tag{S20}$$

We check the constraints in (S18) one by one. The first condition trivially holds by the construction. The last condition holds since $\langle\varphi|n\Delta(\widetilde{\varphi}) - \widetilde{\varphi}|\varphi\rangle = 0$, which implies that $\langle\varphi|C|\varphi\rangle = \langle\varphi|G|\varphi\rangle$. We now move on to the second condition. Clearly $C \ge 0$ and furthermore $G \ge C$ as follows from $\varphi \le n\Delta(\varphi)$. To show that $G \le \mathbb{1}$, just observe that

$$\|G\|_\infty = c\left\| \widetilde{\varphi} + \frac{m-1}{n-1}\left(n\Delta(\widetilde{\varphi}) - \widetilde{\varphi}\right) \right\|_\infty = c\left\| \frac{n-m}{n-1}\widetilde{\varphi} + \frac{n(m-1)}{n-1}\Delta(\widetilde{\varphi}) \right\|_\infty = 1\,. \tag{S21}$$

Hence, C, G as defined above form a valid ansatz for the semidefinite program (S18) and

$$P_{\mathrm{MIO}}(\varphi, m, 0) \ge \operatorname{Tr} G\varphi = \frac{n^2 c}{s}, \tag{S22}$$

which yields the first lower bound in (S16). As for the second bound, it suffices to show that $c \ge 1/m$, i.e. that $c^{-1} \le m$. This can be done thanks to the triangle inequality:

$$c^{-1} = \left\| \widetilde{\varphi} + \frac{m-1}{n-1}\left(n\Delta(\widetilde{\varphi}) - \widetilde{\varphi}\right) \right\|_\infty \le \|\widetilde{\varphi}\|_\infty + (m-1)\left\| \frac{n\Delta(\widetilde{\varphi}) - \widetilde{\varphi}}{n-1} \right\|_\infty \le 1 + (m-1) = m, \tag{S23}$$

where we have used the fact that $\frac{n\Delta(\widetilde{\varphi})-\widetilde{\varphi}}{n-1}$ is a valid density matrix. ∎

PROOF OF THEOREM 4

Theorem 4 *For any pure state φ and any m, it holds that*

$$P_{\mathrm{DIO}}(\varphi \to \Psi_m, 0) = P_{\mathrm{SIO}}(\varphi \to \Psi_m, 0). \tag{S24}$$

Proof. We will assume without loss of generality that the coefficients of $|\varphi\rangle$ are non-negative and arranged in non-increasing order. Let

$$k := \operatorname*{arg\,min}_{1 \le j \le m} \frac{1}{j} \sum_{i=m-j+1}^{d} \varphi_i^2 \tag{S25}$$

such that $P_{\mathrm{SIO}}(\varphi \to \Psi_m, 0) = \frac{m}{k}\sum_{i=m-k+1}^{d}\varphi_i^2$. Note that if $k = m$, then $1 \ge P_{\mathrm{DIO}}(\varphi \to \Psi_m, 0) \ge P_{\mathrm{SIO}}(\varphi \to \Psi_m, 0) = 1$. In the following, we will therefore assume that $1 \le k \le m-1$. Further, let us begin by considering strictly positive ε.

Define

$$\begin{aligned} \varphi_A &:= \sum_{i,j=1}^{m-k} |i\rangle\langle i|\varphi|j\rangle\langle j| \\ \varphi_B &:= \sum_{i,j=m-k+1}^{d} |i\rangle\langle i|\varphi|j\rangle\langle j| \end{aligned} \tag{S26}$$

and notice that $P_{\mathrm{SIO}}(\varphi \to \Psi_m, 0) = \frac{m}{k} \mathrm{Tr}\, \varphi_B$.

Recall that the SDP for the maximum distillation probability under DIO is given by

$$\begin{aligned} P_{\mathrm{DIO}}(\rho \to \Psi_m, \varepsilon) = \underset{X,P,\lambda}{\text{minimize}} \quad & \mathrm{Tr}\, X \\ \text{subject to} \quad & m\Delta(\rho) + m\Delta(P) - P - m\Delta(X) + \lambda\rho - m(1-\varepsilon)\lambda\Delta(\rho) \leq 0 && \text{(S27a)} \\ & P, X \geq 0 && \text{(S27b)} \\ & \lambda \in \mathbb{R}_+. && \text{(S27c)} \end{aligned}$$

We will take the ansatz

$$\begin{aligned} P &= \mu_1 \varphi_B - \mu_2 \varphi_A + [(1-\varepsilon)\lambda - 1]\, \varphi \\ X &= \mu_1 \Delta(\varphi_B) \end{aligned} \tag{S28}$$

for some coefficients $\mu_1, \mu_2 \in \mathbb{R}_+$. Our aim will now be to show that there exists a suitable choice of λ, μ_1, and μ_2 such that (X, P, λ) is feasible for the SDP (S27).

To this end, denote by Θ the all-ones matrix of appropriate size, and notice that the condition (S27a) reduces to

$$\begin{aligned} 0 &\geq m\Delta(\varphi) + m\Delta(P) - P - m\Delta(X) + \lambda\varphi - m(1-\varepsilon)\lambda\Delta(\varphi) \\ &= \mu_2 \left(\varphi_A - m\Delta(\varphi_A)\right) - \mu_1 \varphi_B + (1+\lambda\varepsilon)\varphi \\ &= \begin{pmatrix} [\mu_2 + 1 + \lambda\varepsilon]\, \Theta_A - m\mu_2 \mathbb{1}_A & [1+\lambda\varepsilon]\, \Theta_O \\ [1+\lambda\varepsilon]\, \Theta_O^T & [1+\lambda\varepsilon-\mu_1]\, \Theta_B \end{pmatrix} \circ \varphi \\ &=: Q \circ \varphi \end{aligned} \tag{S29}$$

where $\circ$ denotes the Schur product. Showing that $Q \leq 0$ will then imply the desired relation $Q \circ \varphi \leq 0$ by the Schur product theorem. From the generalized Schur complement condition, $Q \leq 0$ if and only if the following all hold [**? ?**]:

$$\begin{aligned} &\text{(i)} && (\mu_2 + 1 + \lambda\varepsilon)\, \Theta_A - m\mu_2 \mathbb{1}_A \leq 0, \\ &\text{(ii)} && (1 + \lambda\varepsilon - \mu_1)\, \Theta_B \leq 0, \\ &\text{(iii)} && (\mu_2 + 1 + \lambda\varepsilon)\, \Theta_A - m\mu_2 \mathbb{1}_A - (1+\lambda\varepsilon)^2 \Theta_O \left([1+\lambda\varepsilon-\mu_1]\, \Theta_B\right)^{-1} \Theta_O^T \leq 0 \end{aligned} \tag{S30}$$

where M^{-1} denotes the Moore-Penrose pseudoinverse, which in particular satisfies $M^{-1} = M/\mathrm{Tr}\,(M)^2$ for any rank-one Hermitian M [24]. From the first two conditions, we have

$$\begin{aligned} &\text{(i)} && \mu_2 \geq (1+\lambda\varepsilon)\frac{m-k}{k}, \\ &\text{(ii)} && \mu_1 \geq 1 + \lambda\varepsilon, \end{aligned} \tag{S31}$$

and the third condition reduces to

$$\begin{aligned} 0 &\geq (\mu_2 + 1 + \lambda\varepsilon)\, \Theta_A - m\mu_2 \mathbb{1}_A - \frac{(1+\lambda\varepsilon)^2}{[1+\lambda\varepsilon-\mu_1]\, \mathrm{Tr}\, \Theta_B^2} \Theta_O \Theta_B \Theta_O^T \\ &= (\mu_2 + 1 + \lambda\varepsilon)\, \Theta_A - m\mu_2 \mathbb{1}_A - \frac{(1+\lambda\varepsilon)^2}{1+\lambda\varepsilon-\mu_1} \Theta_A \\ &= \left[\mu_2 - \frac{(1+\lambda\varepsilon)\mu_1}{1+\lambda\varepsilon-\mu_1}\right] \Theta_A - m\mu_2 \mathbb{1}_A. \end{aligned} \tag{S32}$$

Noting that the coefficient $\mu_2 - \frac{(1+\lambda\varepsilon)\mu_1}{1+\lambda\varepsilon-\mu_1}$ can never be negative when conditions (i) and (ii) are satisfied, we have

$$\text{(iii)} \quad m\mu_2 \geq (m-k)\left[\mu_2 - \frac{(1+\lambda\varepsilon)\mu_1}{1+\lambda\varepsilon-\mu_1}\right]. \tag{S33}$$

To ensure that $P \geq 0$, notice that since

$$P = \begin{pmatrix} [-\mu_2 + (1-\varepsilon)\lambda - 1]\, \Theta_A & [(1-\varepsilon)\lambda - 1]\, \Theta_O \\ [(1-\varepsilon)\lambda - 1]\, \Theta_O^T & [\mu_1 + (1-\varepsilon)\lambda - 1]\, \Theta_B \end{pmatrix} \circ \varphi, \tag{S34}$$

the positivity of P is equivalent to the positivity of the matrix

$$P' \coloneqq \begin{pmatrix} -\mu_2 + (1-\varepsilon)\lambda - 1 & (1-\varepsilon)\lambda - 1 \\ (1-\varepsilon)\lambda - 1 & \mu_1 + (1-\varepsilon)\lambda - 1 \end{pmatrix}. \tag{S35}$$

We therefore have two additional conditions:

$$\begin{aligned} &\text{(iv)} \quad 0 \leq \operatorname{Tr} P' = 2(1-\varepsilon)\lambda + \mu_1 - \mu_2 - 2 \\ &\text{(v)} \quad 0 \leq \operatorname{Det} P' = \mu_2 - (1-\varepsilon)\lambda\mu_2 - \mu_1\left[1 - (1-\varepsilon)\lambda + \mu_2\right]. \end{aligned} \tag{S36}$$

Let us stress that conditions (i)–(v) together with $\mu_1 \geq 0$ constitute a set of sufficient conditions for a given choice of (X, P, λ) of the form (S28) to satisfy the feasibility conditions of SDP (S27).

We will now make the ansatz

$$\begin{aligned} \lambda &= \frac{\mu_1 - \mu_2}{\varepsilon\mu_2} \\ \mu_1 &= \frac{k\mu_2(\mu_2 - 1)}{m - k} \end{aligned} \tag{S37}$$

leaving μ_2 as a free variable for now. With this choice, conditions (i)–(iii) are always satisfied for any $\mu_2 \in \mathbb{R}$, while the other conditions reduce to

$$\begin{aligned} &\text{(iv)} \quad 2(k\mu_2 - m) + \varepsilon(2k - (2k+m)\mu_2 + k\mu_2^2) \geq 0 \\ &\text{(v)} \quad \mu_2\left(\varepsilon k m(1-\mu_2)^2 - (m - k\mu_2)^2\right) \leq 0 \end{aligned} \tag{S38}$$

We will now make a choice of μ_2 which can be verified to satisfy the above the inequalities for any $\varepsilon < \frac{k}{m}$ as

$$\mu_2 = \frac{m + m^2\sqrt{\varepsilon}}{k - \varepsilon m} \tag{S39}$$

where we note that $\mu_2 \geq 1$ and $\lim_{\varepsilon\to 0}\mu_1 = \lim_{\varepsilon\to 0}\mu_2 = \frac{m}{k}$. Since all conditions (i)–(v) are satisfied for our choice of (X, P, λ) with the given μ_1 and μ_2 for any $\varepsilon < \frac{k}{m}$, the triple (X, P, λ) is a valid feasible solution for the SDP (S27). This means in particular that for any $0 < \varepsilon < \frac{k}{m}$ we have

$$P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) \leq \operatorname{Tr} X = \mu_1 \operatorname{Tr}\varphi_B. \tag{S40}$$

Using the fact that SIO $\subset$ DIO as well as that $P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) \geq P_{\mathrm{DIO}}(\varphi \to \Psi_m, 0)$ for any ε, this then gives

$$\begin{aligned} P_{\mathrm{SIO}}(\varphi \to \Psi_m, 0) &\leq P_{\mathrm{DIO}}(\varphi \to \Psi_m, 0) \\ &\leq \lim_{\varepsilon\to 0} P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) \\ &\leq \frac{m}{k}\operatorname{Tr}\varphi_B \\ &= P_{\mathrm{SIO}}(\varphi \to \Psi_m, 0) \end{aligned} \tag{S41}$$

which completes the proof. ∎

PROOF OF PROPOSITION 5 AND 6

Proposition 5 *For $O \in \{DIO, MIO\}$ and any pure state φ with $\varphi_1 \geq \ldots \geq \varphi_n > 0$, it holds that*

$$P_O(\varphi \to \Psi_2, \varepsilon) = \begin{cases} 1 & \text{if } \varepsilon \geq \varepsilon_0(\varphi_1), \\ 2(1-\varphi_1^2)\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2 & \text{otherwise,} \end{cases} \tag{S42}$$

where

$$\varepsilon_0(\varphi_1) \coloneqq \begin{cases} 0 & \text{if } \varphi_1 \leq \frac{1}{\sqrt{2}}, \\ \frac{1}{2} - \varphi_1\sqrt{1-\varphi_1^2} & \text{otherwise.} \end{cases} \tag{S43}$$

Proof. Since the case $\varepsilon = 0$ follows from Theorem 4, we will assume $\varepsilon > 0$. For the case of $\varphi_1 \leq \frac{1}{\sqrt{2}}$, we have

$$P_O(\varphi \to \Psi_2, \varepsilon) \geq P_{\mathrm{SIO}}(\varphi \to \Psi_2, 0) = 1. \tag{S44}$$

From the result in [**?**], we know that if $\varepsilon \geq \frac{1}{2} - \varphi_1\sqrt{1-\varphi_1^2}$, then $P_{\mathrm{DIO}}(\varphi \to \Psi_2, \varepsilon) = 1$. In the following, we therefore only consider the case $\varepsilon < \frac{1}{2} - \varphi_1\sqrt{1-\varphi_1^2}$. We prove this result by explicit constructing feasible solutions in both primal and dual SDPs. The primal SDP under DIO is given by

$$P_{\mathrm{DIO}}(\varphi \to \Psi_2, \varepsilon) = \max\left\{\mathrm{Tr}\, G\varphi \;\middle|\; \mathrm{Tr}\, C\varphi \geq (1-\varepsilon)\,\mathrm{Tr}\, G\varphi,\ 0 \leq C \leq G \leq \mathbb{1},\ \Delta(G) = 2\Delta(C),\ G = \Delta(G)\right\}. \tag{S45}$$

We take the ansatz

$$G = \mathbb{1} - x|0\rangle\langle 0|, \quad C = \frac{1}{2}G + y\sum_{i=2}^{n}\varphi_1\varphi_n(|1\rangle\langle i| + |i\rangle\langle 1|). \tag{S46}$$

Then we have

$$\mathrm{Tr}\, G\rho = 1 - x\varphi_1^2, \quad \mathrm{Tr}\, C\rho = \frac{1}{2}(1 - x\varphi_1^2) + 2y\varphi_1^2(1-\varphi_1^2), \tag{S47}$$

and

$$\mathrm{spec}(C) = \mathrm{spec}(G - C) = \left\{\underbrace{\frac{1}{2}, \cdots, \frac{1}{2}}_{n-2 \text{ fold}}, \frac{1}{2} - \frac{x}{4} - \sqrt{y^2\varphi_1^2(1-\varphi_1^2) + \frac{x^2}{16}}, \frac{1}{2} - \frac{x}{4} + \sqrt{y^2\varphi_1^2(1-\varphi_1^2) + \frac{x^2}{16}}\right\}. \tag{S48}$$

Then we have the relaxation

$$\begin{aligned} P_{\mathrm{DIO}}(\varphi \to \Psi_2, \varepsilon) \geq \text{maximize} \quad & 1 - x\varphi_1^2 \\ \text{subject to} \quad & 4y\varphi_1^2(1-\varphi_1^2) = (1-2\varepsilon)(1 - x\varphi_1^2), \\ & 1 - x \geq 4y^2\varphi_1^2(1-\varphi_1^2), \\ & 0 \leq x \leq 1. \end{aligned} \tag{S49}$$

By choosing

$$x = \frac{1 - 2(1-\varphi_1^2)(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon})^2}{\varphi_1^2}, \quad y = \frac{(\sqrt{1-\varepsilon} + \sqrt{\varepsilon})^2}{2\varphi_1^2(1-2\varepsilon)}, \tag{S50}$$

we can verify that this is a feasible solution. Thus we have

$$P_{\mathrm{DIO}}(\varphi \to \Psi_2, \varepsilon) \geq 2(1-\varphi_1^2)\left(\frac{\sqrt{1-\varepsilon} + \sqrt{\varepsilon}}{1-2\varepsilon}\right)^2. \tag{S51}$$

As for the dual problem, we consider the dual SDP under MIO,

$$\begin{aligned} P_{\mathrm{MIO}}(\varphi \to \Psi_2, \varepsilon) = \text{minimize} \quad & \mathrm{Tr}\, Y \\ \text{subject to} \quad & (1 - (1-\varepsilon)x)\varphi + X + \Delta(Z) \leq Y, \\ & x\varphi - X - 2\Delta(Z) \leq 0, \\ & x \geq 0, X \geq 0, Y \geq 0 \end{aligned} \tag{S52}$$

Taking

$$x = \frac{(\sqrt{1-\varepsilon} + \sqrt{\varepsilon})^2}{\sqrt{1-\varepsilon}\sqrt{\varepsilon}(1-2\varepsilon)}, \quad Y = 2\left(\frac{\sqrt{1-\varepsilon} + \sqrt{\varepsilon}}{1-2\varepsilon}\right)^2(\mathbb{1} - |0\rangle\langle 0|)\varphi(\mathbb{1} - |0\rangle\langle 0|), \tag{S53}$$

$$Z = \frac{2}{1-2\varepsilon}|0\rangle\langle 0|\varphi|0\rangle\langle 0|, \quad X = Y - \Delta(Z) - (1-(1-\varepsilon)x)\varphi, \tag{S54}$$

we can verify that $\{x, X, Y, Z\}$ is a valid feasible solution. Thus

$$P_{\mathrm{MIO}}(\varphi \to \Psi_2, \varepsilon) \le 2(1-\varphi_1^2)\left(\frac{\sqrt{1-\varepsilon}+\sqrt{\varepsilon}}{1-2\varepsilon}\right)^2. \tag{S55}$$

Combining Eqs. (S51) and (S55), we have the desired result. ∎

Recall that IO and SIO have the same power in pure-state transformations and it holds that

$$P_{(\mathrm{S})\mathrm{IO}}(\varphi \to \Psi_m, 0) = \begin{cases} 0 & \text{if } \operatorname{rank}\Delta(\varphi) < m, \\ \min\limits_{k\in[1,m]} \dfrac{m}{k} \displaystyle\sum_{i=m-k+1}^{d} \varphi_i^2 & \text{otherwise.} \end{cases} \tag{S56}$$

Proposition 6 *For any pure state $|\varphi\rangle = \sum_{i=1}^{n} \varphi_i |i\rangle$ with $\varphi_i > 0$, it holds that*

- *If $n \ge m$, $P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) > 0$.*
- *If $n < m$, $P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) \begin{cases} > 0, & \varepsilon \ge 1 - \frac{n}{m}, \\ = 0, & \varepsilon < 1 - \frac{n}{m}. \end{cases}$*

Proof. For the first argument, if $n \ge m$, we know that

$$P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) \ge P_{\mathrm{SIO}}(\varphi \to \Psi_m, 0) > 0, \tag{S57}$$

where the second inequality follows from Eq. (S56).

Note that if $P_O(\sigma_1 \to \sigma_2, 0) = 1$, then $P_O(\rho \to \sigma_2, 0) \ge P_O(\rho \to \sigma_1, 0)$ since we can first transform ρ to σ_1 perfectly and then to get σ_2. For the second argument, if $\varepsilon \ge 1 - \frac{n}{m}$, we have

$$P_{\mathrm{DIO}}(\Psi_n \to \Psi_m^{\varepsilon}, 0) = P_{\mathrm{DIO}}(\Psi_n \to \Psi_m, \varepsilon) = 1. \tag{S58}$$

The first equality follows from the fact that $P_O(\rho \to \Psi_m, \varepsilon) = P_O(\rho \to \Psi_m^{\varepsilon}, 0)$. The second equality follows from Lemma S2 below. Then

$$\begin{aligned} P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) &= P_{\mathrm{DIO}}(\varphi \to \Psi_m^{\varepsilon}, 0) \\ &\ge P_{\mathrm{DIO}}(\varphi \to \Psi_n, 0) \\ &\ge P_{\mathrm{SIO}}(\varphi \to \Psi_n, 0) \\ &> 0. \end{aligned} \tag{S59}$$

The first inequality follows from Eq. (S58). The last inequality follows from Eq. (S56). If $\varepsilon \le 1 - \frac{n}{m}$, we have $P_{\mathrm{DIO}}(\varphi \to \Psi_m, \varepsilon) \le P_{\mathrm{DIO}}(\Psi_n \to \Psi_m, \varepsilon) = 0$, where the second equality follows from Lemma S2. ∎

Lemma S2 *For any integer $n \le m$, it holds that*

$$P_{\mathrm{DIO}}(\Psi_n \to \Psi_m, \varepsilon) = \begin{cases} 1, & \varepsilon \ge 1 - \frac{n}{m}, \\ 0, & \varepsilon < 1 - \frac{n}{m}. \end{cases} \tag{S60}$$

Proof. For $\varepsilon \ge 1 - \frac{n}{m}$, we can take feasible solution $G = \mathbb{1}$, $C = \frac{n}{m}\psi_n$, which gives feasible value 1 in the primal problem. For $\varepsilon < 1 - \frac{n}{m}$, we can take feasible solution $x = \frac{1}{1-\frac{n}{m}-\varepsilon}$, $X = Y = 0$, $Z = \frac{1}{m-n-m\varepsilon}\mathbb{1}$, $W = \frac{1}{m-n-m\varepsilon}(n\psi_n - \mathbb{1})$, which gives feasible value 0 in the dual problem. ∎

EXAMPLES FOR CATALYST-ASSISTED ENHANCEMENT

For the catalysis scenario, we require that the catalyst is returned no matter the distillation process succeeds or not. In a more general case than the setting presented in the main text, we may accept imperfect catalyst to be returned. Denote δ as the infidelity tolerance for the returning catalyst. We define the catalysis-assisted maximal success probability as

$$P_O(\rho \xrightarrow{\gamma,\,\delta} \Psi_m, \varepsilon) := \max\ p \tag{S61a}$$

$$\text{s.t. } \Pi(\rho\otimes\gamma) = p|0\rangle\langle 0| \otimes \sigma \otimes \gamma_0 + (1-p)|1\rangle\langle 1| \otimes \omega \otimes \gamma_1, \tag{S61b}$$

$$F(\sigma, \Psi_m) \ge 1-\varepsilon,\ F(\gamma, \gamma_0) \ge 1-\delta,\ F(\gamma, \gamma_1) \ge 1-\delta, \tag{S61c}$$

$$\Pi \in O,\ 0 \le p \le 1. \tag{S61d}$$

Using the same argument as Remark 1 and Remark 2, we can fix the output state $\sigma = \Psi_m^{\varepsilon}$ and $\omega = \mathbb{1}/m$. Then the maximal success probability can be written as

$$P_O(\rho \xrightarrow{\gamma,\,\delta} \Psi_m, \varepsilon) = \max\ p \tag{S62a}$$

$$\text{s.t. } \Pi(\rho\otimes\gamma) = p|0\rangle\langle 0|\otimes\Psi_m^{\varepsilon}\otimes\gamma_0 + (1-p)|1\rangle\langle 1|\otimes\mathbb{1}/m\otimes\gamma_1, \tag{S62b}$$

$$F(\gamma,\gamma_0)\geq 1-\delta,\ F(\gamma,\gamma_1)\geq 1-\delta, \tag{S62c}$$

$$\Pi\in O,\ 0\leq p\leq 1. \tag{S62d}$$

If the catalyst γ we considered is the maximally coherent state Ψ_k, we can also fix the output catalyst $\gamma_0 = \gamma_1 = \Psi_k^{\delta}$. It gives

$$P_O(\rho \xrightarrow{\Psi_k,\,\delta} \Psi_m, \varepsilon) = \max\ p \tag{S63a}$$

$$\text{s.t. } \Pi(\rho\otimes\Psi_n) = p|0\rangle\langle 0|\otimes\Psi_m^{\varepsilon}\otimes\Psi_k^{\delta} + (1-p)|1\rangle\langle 1|\otimes\mathbb{1}/m\otimes\Psi_k^{\delta}, \tag{S63b}$$

$$\Pi\in O,\ 0\leq p\leq 1. \tag{S63c}$$

In the above optimization, the only variables we need to optimize over are Π and p. It is thus clear that optimization (S63) is an SDP for DIO.

Consider the class of states $\rho = q\cdot v_1+(1-q)\cdot v_2$ with $|v_1\rangle = (|00\rangle-|01\rangle-|10\rangle+|11\rangle)/2$, $|v_2\rangle = (2|00\rangle+6|01\rangle-3|10\rangle+|11\rangle)/5\sqrt{2}$ and state parameter $q\in[0.1,0.5]$. We show the difference between the catalysis-assisted success probability $P_{\text{DIO}}\big(\rho \xrightarrow{\Psi_2,\,\delta} \Psi_2, 0.01\big)$ and the unassisted success probability $P_{\text{DIO}}\big(\rho\to\Psi_2, 0.01\big)$ in the following Fig. 1. It shows that catalyst indeed enhances the maximal success probability even when requiring perfect catalyst to be returned ($\delta = 0$), while slight infidelity tolerance of the returning catalyst will boost the maximal success probability even more. On the right hand side of Fig. 1, the enhancement ratio is given by $\big[P_{\text{DIO}}\big(\rho \xrightarrow{\Psi_2,\,\delta} \Psi_2, 0.01\big) - P_{\text{DIO}}\big(\rho\to\Psi_2, 0.01\big)\big]/P_{\text{DIO}}\big(\rho\to\Psi_2, 0.01\big)$. The enhancement ratio can achieve up to 12% when $\delta = 0$.

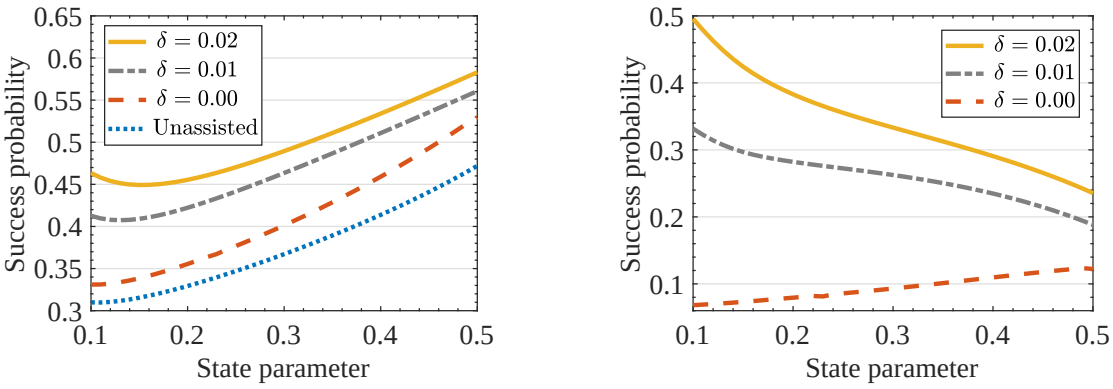


FIG. 1. Examples of catalyst-assisted probabilistic coherence distillation.

Similarly, we can also give another class of states showing the enhancement of catalyst. Consider the class of states $\rho = q\cdot u_1 + (1-q)\cdot u_2$ with $|u_1\rangle = (|00\rangle + |01\rangle + |10\rangle + |11\rangle)/2$ and $|u_2\rangle = (3|00\rangle - 2|01\rangle + |10\rangle + 2|11\rangle)/3\sqrt{2}$ and state parameter $q\in[0.2,0.7]$. We also compare $P_{\text{DIO}}\big(\rho \xrightarrow{\Psi_2,\,\delta} \Psi_2, 0.01\big)$ with $P_{\text{DIO}}\big(\rho\to\Psi_2, 0.01\big)$ and the enhancement is shown in Fig 2.

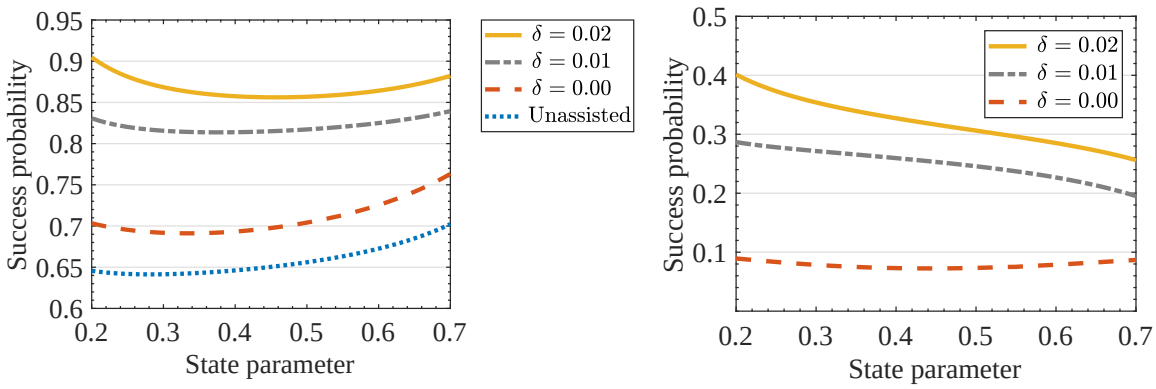


FIG. 2. Examples of catalyst-assisted probabilistic coherence distillation.

Remark 6 *We can denote $V = p\gamma_0$ and $W = (1-p)\gamma_1$ in the optimization (S62). Then for the case that catalyst γ is a pure state, the fidelity conditions in (S62c) is equivalent to (S64c). Combining with the semidefinite conditions for MIO/DIO, the maximal success probability under MIO/DIO can be written as SDPs:*

$$P_O(\rho \xrightarrow{\gamma,\delta} \Psi_m, \varepsilon) = \max \operatorname{Tr} V \tag{S64a}$$

$$\text{s.t. } \Pi(\rho\otimes\gamma) = |0\rangle\langle 0| \otimes \Psi_m^{\varepsilon} \otimes V + |1\rangle\langle 1| \otimes \mathbb{1}/m \otimes W, \tag{S64b}$$

$$\operatorname{Tr}\gamma V \geq (1-\delta)\operatorname{Tr} V,\ \operatorname{Tr}\gamma W \geq (1-\delta)\operatorname{Tr} W \tag{S64c}$$

$$\Pi \in O,\ 0 \leq p \leq 1. \tag{S64d}$$