Virtual quantum resource distillation: General framework and applications

Ryuji Takagi,[1,2,*] Xiao Yuan,[3,4,†] Bartosz Regula,[5,6,‡] and Mile Gu[2,7,8,§]

[1]*Department of Basic Science, The University of Tokyo, Tokyo 153-8902, Japan*
[2]*Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore*
[3]*Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China*
[4]*School of Computer Science, Peking University, Beijing 100871, China*
[5]*Mathematical Quantum Information RIKEN Hakubi Research Team, RIKEN Cluster for Pioneering Research, and RIKEN Center for Quantum Computing, Wako, Saitama 351-0198, Japan*
[6]*Department of Physics, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*
[7]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore*
[8]*CNRS-UNS-NUS-NTU International Joint Research Unit, UMI No. 3654, Singapore 117543, Singapore*

(Received 2 March 2023; accepted 17 November 2023; published 2 February 2024)

We develop the general framework of virtual resource distillation, an alternative distillation strategy proposed in Phys. Rev. Lett. **132**, 050203 (2024), which extends conventional quantum resource distillation by integrating the power of classical postprocessing. The framework presented here is applicable not only to quantum states, but also to dynamical quantum objects such as quantum channels and higher-order processes. We provide a general characterization and benchmarks for the performance of virtual resource distillation in the form of computable semidefinite programs as well as several operationally motivated quantities. We apply our general framework to various concrete settings of interest, including standard resource theories such as entanglement, coherence, and magic, as well as settings involving dynamical resources such as quantum memory, quantum communication, and non-Markovian dynamics. The framework of probabilistic distillation is also discussed.

DOI: 10.1103/PhysRevA.109.022403

I. INTRODUCTION

The advantages of quantum algorithms and information processing are enabled by the efficient use of quantum resources such as quantum entanglement [1] and superposition [2]. However, it is usually difficult to prepare such quantum resources with high quality due to inevitable noise and imperfection. The standard approach to address this issue is through resource distillation, a class of protocols to prepare high-quality quantum resources from those of lower quality.

The feasibility and performance of resource distillation are major topics of study in quantum information theory, often analyzed using the tools developed in quantum resource theories [3], which are frameworks that deal with the quantification and manipulation of physical quantities that are costly to access in a given setting. Resource distillation has been studied in various resource theories, with the ultimate goal of producing output quantum objects, such as quantum states and channels, that are as close as possible to a desired target object. Although this goal is well motivated, as it allows for versatile use of the distilled resource object, it may be too restrictive depending on the objective of the overall quantum algorithm that utilizes the processed resource object after the distillation procedure.

Here we observe that many quantum algorithms, such as variational quantum algorithms [4], ultimately aim to obtain some classical output, i.e., numerical values, retrieved by measuring the expectation values of suitable observables. Such quantum algorithms do not strictly require distilling a desired quantum resource physically, as long as we can retrieve the expected values of all observables made on the quantum objective. This motivates us to propose a variant of resource distillation. Our distillation strategy does not directly distill a better quantum object, but instead fully utilizes the potential of classical postprocessing to virtually approximate its measurement statistics, allowing us to simulate the expectation values one would obtain if one were in physical possession of the target object. We remark that several protocols known as virtual cooling [5] and virtual distillation in quantum error mitigation [6,7] share a similar idea that classical postprocessing enables us to simulate purer quantum states, although they differ from our framework: These techniques extract the measurement statistics for purer quantum objects by coherently interacting multiple copies of noisy objects, while our virtual resource distillation applies a probabilistic operation to a single copy of the noisy quantum object, which is inspired by and related to the techniques for simulating unphysical objects [8–10], and error mitigation techniques based on quasiprobability [11].

This paper provides an extension and a rigorous theoretical foundation to our other paper [12], where the notion of virtual resource distillation is introduced. Notably, although the discussion in [12] focuses on resource theories of quantum states, here we present extensive discussions of the fully

*ryuji.takagi@phys.c.u-tokyo.ac.jp
†xiaoyuan@pku.edu.cn
‡bartosz.regula@gmail.com
§mgu@quantumcomplexity.org

 ©2024 American Physical Society

general framework, which includes applications to dynamical resource theories of quantum channels and higher-order processes [13–34], and the framework of probabilistic distillation [31,32,35–40]. We give detailed comparisons between the sampling overhead needed to realize virtual distillation protocols and conventional ones. We discuss applications of virtual resource distillation in various physical settings, providing an in-depth study of the distillation performance in each case. Along the way, we include the proofs for Theorems 1 and 2 in Ref. [12], which correspond to Theorems 1 and 2 in this article.

II. RESOURCE THEORIES

The restrictions imposed by a given physical settings can usually be represented by a limited set of quantum states and operations that one has access to. For instance, when two parties are physically separated and quantum communication is hard to establish, it is reasonable to study the scenario where they only have access to local quantum operations and classical communication (LOCC). In such a scenario, they can only generate separable states, and other states are costly "resources" that cannot be created for free, where entanglement serves as the resource quantity of interest. Central questions in such a scenario include those related to (i) resource quantification, e.g., what is a good way of quantifying the underlying resource that we do not have free access to, and (ii) resource manipulation, e.g., what are the resource transformations possible by only using the freely accessible operations. In general, the manipulated resources need not be quantum states, but can be e.g. quantum channels, measurements, or higher-order quantum operations. We will approach the problem generally by considering all such resource objects in a common formalism.

Resource theories are frameworks that provide a systematic approach to study the above questions of quantum resource quantification and manipulation [3]. The basic building blocks of the resource theory framework include a set $\mathcal{F}$ of free objects, a subset of objects that can be prepared for free, and a set O of free operations, the accessible operations that are allowed to transform the resource objects in the given setting. To reflect the physical constraints and the underlying quantum resource, we impose a basic condition on free operations; no free operation can create a resourceful object from a free object, i.e., if $\Lambda \in O$, then $\Lambda(X) \in \mathcal{F}\ \forall X \in \mathcal{F}$. The maximal such set is called resource nongenerating operations, examples of which include separability-preserving operations in entanglement theory and maximally incoherent operations in coherence theory, and any arbitrary set O of free operations is then a subset of resource nongenerating operations. With these concepts, we can formalize resource quantification by considering a function R from objects to real numbers. In particular, we call a function R a resource measure or monotone if (a) it always gives the smallest value for all free objects, i.e., for some constant c, $R(X) = c\,\forall X \in \mathcal{F}$ and $R(X) \geqslant c\,\forall X$, and (b) it is monotonically nonincreasing under free operations, i.e., $R(X) \geqslant R(\Lambda(X))$ for every object X and every free operation $\Lambda \in O$.

Depending on the specific setting of interest, one can flexibly select the set of resource objects to study. When we are interested in the manipulation of quantum states and the resources contained therein, the set of all quantum states is the relevant object of study, which includes the designated set $\mathcal{F}$ of free states as a subset. In this scenario, quantum channels serve as the operations that manipulate quantum states. We call the resource theories whose resource objects are quantum states resource theories of quantum states or state theories in short.

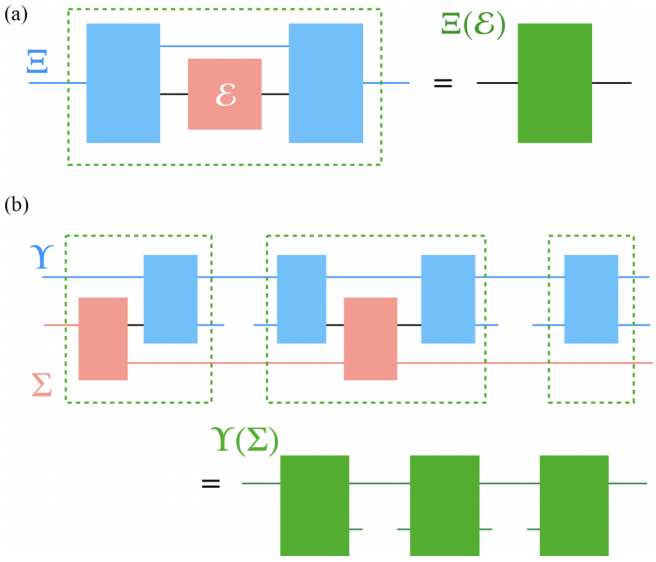

FIG. 1. (a) Channel transformation by a superchannel. A superchannel Ξ is constructed by connecting two channels (denoted by blue boxes) by the identity channel (blue wire between two boxes). Inserting a channel $\mathcal{E}$ into a slot between the two channels results in another channel $\Xi(\mathcal{E})$ (green). (b) Comb transformation by another comb. Combs Υ (blue) and Σ (red) are constructed by interlocking a series of channels, which results in another comb $\Upsilon(\Sigma)$ (green).

On the other hand, if one would like to study the resource contents belonging to quantum channels, as done, e.g., in the theory of quantum communication, then the relevant object of study becomes the set of quantum channels. Quantum channels are manipulated by quantum superchannels [41,42] that transform quantum channels to quantum channels. We call this framework resource theories of quantum channels. A quantum superchannel is constructed by a combination of two channels, between which another channel can be inserted. Inserting a channel $\mathcal{E}$ into the slot for a superchannel Ξ then results in an output channel $\Xi(\mathcal{E})$ (Fig. 1).

One can extend superchannels to ones with multiple empty slots, known as quantum combs, and take them as the main resource objects to study, which construct resource theories of quantum combs [34,41]. For instance, this framework is useful for studying noise suppression, where non-Markovian noise can be considered as a quantum comb [43]. A quantum comb consists of a set of bipartite quantum channels and has empty slots between these channels. This allows a quantum comb Υ to act on another quantum comb Σ by interlocking them as in Fig. 1, which results in another comb $\Upsilon(\Sigma)$. Therefore, quantum combs themselves serve as the operations that manipulate quantum combs, and thus any set O of free operations is a subset of quantum combs in this framework.

It is worth noting that there is a strict hierarchy among these three types of objects: Quantum states are special forms of quantum channels, and quantum channels are special forms of quantum combs. Unless stated otherwise, our results hold for any type of resource object as long as it is isomorphic to a closed convex subset of operators acting on a finite-dimensional Hilbert space. Our results also do not assume any specification of the set $\mathcal{F}$ of free objects and the set O of free operations except the assumption that they are closed convex sets. This approach that does not rely on the specific structure of free objects is due to the recently developed general resource theories [3,44], which has provided operational characterizations of general resources in terms of discrimination tasks [22,45–47], resource erasure [24,48], and resource manipulation [31,32,38–40,49–52].

III. RESOURCE DISTILLATION

Many quantum information processing protocols are designed under the assumption that we are in possession of a specific form of resource objects, e.g., maximally entangled states. However, this assumption is hard to meet in a realistic noisy scenario. Therefore, preparing the desired specific object from distorted noisy ones using only the freely accessible operations is a crucial subroutine in the realization of quantum information processing tasks. This procedure is known as resource distillation, and its performance in relation to resource quantification has been a major topic of study.

For instance, suppose two parties, Alice and Bob, would like to run the quantum teleportation protocol but only have access to noisy entangled states. Quantum teleportation can be run by first distilling a maximally entangled state from the accessible noisy entangled states by using local operations and classical communication and then using the distilled entangled state as a resource for quantum teleportation. In this protocol, the performance of the distillation process plays a crucial role in characterizing the efficiency of running quantum teleportation. In the following, we formalize the distillation performance in the general setting, including the scenario where the resource object of interest is not only a quantum state, but a quantum process.

Let T be a desired target object; again, this can be a quantum state, channel, or even a more general comb, depending on the setting of interest. Suppose now that our goal is to obtain as many copies of T as possible within a tolerable error ε. The one-shot distillation rate is defined as

$$D^{\varepsilon}(X) \coloneqq \sup_{\Lambda\in O}\{m \mid \Lambda(X) \sim_{\varepsilon} T^{\otimes m}\}, \tag{1}$$

where $A \sim_{\varepsilon} B$ means that A and B are ε-close with respect to some distance measure. In this work, we focus on the trace-norm–based distance, which has the form below depending on the type of resource theories under study.

For two quantum states ρ_1 and ρ_2, we consider the trace distance, i.e.,

$$\rho_1 \sim_{\varepsilon} \rho_2 \iff \tfrac{1}{2}\|\rho_1 - \rho_2\|_1 \leqslant \varepsilon, \tag{2}$$

where $\|\cdot\|_1$ is the trace norm. In this paper we use τ to denote a target state for state distillation and ψ to emphasize that the target is pure. A target state is commonly set as a pure state with some specific form, e.g., a Bell state in the entanglement distillation and the T state in the magic state distillation, although we do not put any restriction on the target state unless stated otherwise. For clarity, we will sometimes use τ instead of T when the target is a quantum state.

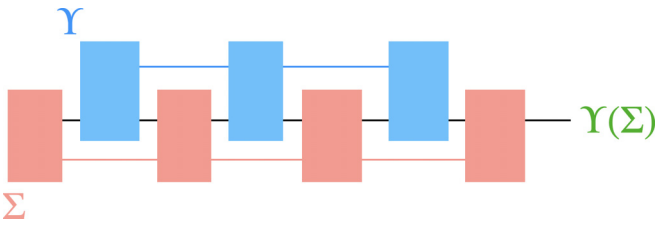

FIG. 2. For a given comb Υ, interlocking another comb Σ that fills all the slots of Υ results in a quantum state as an output $\Upsilon(\Sigma)$. Here $\frac{1}{2}\|\Upsilon_1 - \Upsilon_2\|_c$ is defined as the maximum trace distance between states $\Upsilon_1(\Sigma)$ and $\Upsilon_2(\Sigma)$.

For two quantum channels $\mathcal{E}_1$ and $\mathcal{E}_2$, the distance between two channels is described by the diamond distance, i.e.,

$$\begin{aligned}\mathcal{E}_1 \sim_{\varepsilon} \mathcal{E}_2 &\iff \max_{\rho} \tfrac{1}{2}\|\mathrm{id}\otimes\mathcal{E}_1(\rho) - \mathrm{id}\otimes\mathcal{E}_2(\rho)\|_1 \leqslant \varepsilon \\ &\iff \tfrac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_{\diamond} \leqslant \varepsilon,\end{aligned} \tag{3}$$

where $\|\cdot\|_{\diamond}$ is the diamond norm [53], with id denoting the identity channel on an ancillary space, *a priori* unbounded, but it is in fact sufficient to consider an ancillary space of the same dimension as the input space of the channel [54]. What Eq. (3) means is that the distance between two channels can be measured by the maximum trace distance between two output states obtained by the partial application to the same input state.

This idea can be extended to measuring the distance between two quantum combs. For two given combs Υ_1 and Υ_2 with the same input-output structure, we can consider applying them to another interlocking comb that outputs quantum states and takes the trace distance between these two states (Fig. 2). We define the distance between Υ_1 and Υ_2 by taking the maximization over all such interlocking combs, i.e.,

$$\tfrac{1}{2}\|\Upsilon_1 - \Upsilon_2\|_c \coloneqq \max_{\Sigma} \tfrac{1}{2}\|\Upsilon_1(\Sigma) - \Upsilon_2(\Sigma)\|_1, \tag{4}$$

where the maximization is taken over all the combs such that $\Upsilon_{1,2}(\Sigma)$ is a quantum state. We remark that this distance measure was introduced and discussed previously in Refs. [55,56]. With this distance measure, we can define the ε-closeness of two combs Υ_1 and Υ_2 as

$$\Upsilon_1 \sim_{\varepsilon} \Upsilon_2 \iff \tfrac{1}{2}\|\Upsilon_1(\Sigma) - \Upsilon_2(\Sigma)\|_c \leqslant \varepsilon. \tag{5}$$

IV. VIRTUAL RESOURCE DISTILLATION

A. Setting

As the ultimate goal of most quantum information processing tasks is to obtain classical information of interest, they typically terminate with measurements. We show that when the desired classical information is an expectation value of the final quantum state at the output of the protocol, we can extend the notion of resource distillation.

For clarity, let us begin with state theories, i.e., $\mathcal{F}$ is a set of quantum states and O is a set of quantum channels. Consider a quantum information processing protocol that applies a quantum channel Λ to a certain resource state τ to produce the final quantum state $\Lambda(\tau)$. This is followed by a measurement of an observable M. Here we assume $-I/2 \leqslant M \leqslant I/2$ because we can always normalize a bounded observable acting on a finite-dimensional system. This ensures that the measurement outcomes for M are bounded in $[-1/2, 1/2]$.

We suppose that the classical information of interest is the expectation value of the final state for the observable M, i.e., $\mathrm{Tr}[M\Lambda(\tau)]$. The expectation value can be estimated by taking the statistical average over many measurement outcomes. The Hoeffding inequality [57] ensures that $O[\log_2(1/\delta)/\beta^2]$ samples provide the estimate with accuracy β with probability $1-\delta$.

Suppose that we are not in possession of the state τ but instead have access to another state ρ together with free operations O. For simplicity, we also suppose that the accessible free operations only accommodate a single copy of ρ coherently, in the sense that they are one-shot protocols. This will allow us to provide a description particularly suitable for near-term technologies, where the size of available quantum devices is restricted; a more general characterization can be obtained by assuming that the input state is of the form $\rho^{\otimes n}$ for some number of copies of a state ρ.

One conventional way to estimate the desired expectation value is to prepare $D^\varepsilon(\rho)$ copies of the state τ from the available state ρ by a distillation protocol. The optimal distillation protocol can prepare a state $\tilde{\tau}$ that is ε-close to $\tau^{\otimes D^\varepsilon(\rho)}$. Recalling that we are interested in the expectation value of a single copy of τ, we will assume that every reduced state of $\tilde{\tau}$ is identical, which can be realized by symmetrization, and let τ' be the reduced state of $\tilde{\tau}$ with the same size of τ. Since the partial trace does not increase the trace distance, we have $\frac{1}{2}\|\tau'-\tau\|_1 \leqslant \varepsilon$. We then get

$$\begin{aligned}|\mathrm{Tr}[M\Lambda(\tau')] - \mathrm{Tr}[M\Lambda(\tau)]| &\leqslant \tfrac{1}{2}\|\Lambda(\tau')-\Lambda(\tau)\|_1 \\ &\leqslant \tfrac{1}{2}\|\tau'-\tau\|_1 \\ &\leqslant \varepsilon \end{aligned} \tag{6}$$

for an arbitrary observable M with $-I/2 \leqslant M \leqslant I/2$.

Therefore, the distilled state admits the estimation of expectation value with the accuracy ε. This means that $O\{[1/D^\varepsilon(\rho)]\log_2(1/\delta)/\beta^2\}$ copies of ρ provide the estimate with accuracy $\beta+\varepsilon$ with probability $1-\delta$, where the number of copies differs by the factor $D^\varepsilon(\rho)$ compared to the case when the resource state τ is available.

B. Framework

We now introduce virtual resource distillation. The basic idea is that one can apply classical postprocessing to help estimate the desired expectation value. Suppose that a state $\tilde{\tau}$ that is ε-close to $\tau^{\otimes m}$ can be decomposed into the form

$$\tilde{\tau} = \lambda_+\Lambda_+(\rho) - \lambda_-\Lambda_-(\rho), \tag{7}$$

where $\Lambda_\pm \in O$ are free operations and $\lambda_\pm \geqslant 0$ are nonnegative numbers. Since we assume here that $\Lambda_\pm$ are trace preserving, we have $\lambda_+ - \lambda_- = 1$. Letting $\tau'_\pm$ be the reduced states of $\Lambda_\pm(\rho)$, the reduced state τ' of $\tilde{\tau}$ can be written as $\tau' = \lambda_+\tau'_+ - \lambda_-\tau'_-$. Defining $\gamma := \lambda_+ + \lambda_-$ and $p_\pm = \lambda_\pm/\gamma$, the quantity

$$\begin{aligned}&\lambda_+\mathrm{Tr}[M\Lambda(\Lambda_+(\rho))] - \lambda_-\mathrm{Tr}[M\Lambda(\Lambda_-(\rho))] \\ &\quad = p_+\gamma\mathrm{Tr}[M\Lambda(\Lambda_+(\rho))] - p_-\gamma\mathrm{Tr}[M\Lambda(\Lambda_-(\rho))]\end{aligned} \tag{8}$$

corresponds to the desired expectation value $\mathrm{Tr}[M\Lambda(\tau)]$ with error ε. This form ensures that the following procedure gives the estimator of the expectation value with bias ε.

(1) Flip the biased coin that lands heads with probability p_+ and tails with probability p_-.
(2) When we see heads, apply Λ_+ to ρ and measure m commuting observables $M\otimes I^{\otimes m-1}, I\otimes M\otimes I^{\otimes m-2}, \ldots, I^{\otimes m-1}\otimes M$ to get outcomes $o_1, \ldots, o_m$. Store the value $\gamma o_1, \ldots, \gamma o_m$ to the classical register. If we see tails, apply Λ_- to ρ, measure the same observables, and get m measurement outcomes $o_1, \ldots, o_m$. Store the value $-\gamma o_1, \ldots, -\gamma o_m$ in the classical register.
(3) Repeat the above process and take the sample average of the values stored in the classical register.

Due to the classical postprocessing, in which we multiply γ or $-\gamma$ by the measurement outcome, the possible range of each random variable changes to $[-\gamma/2, \gamma/2]$. The Hoeffding inequality ensures that this procedure allows us to estimate the desired expectation value with accuracy $\beta+\varepsilon$ with probability $1-\delta$ with $O[(\gamma^2/m)\log_2(1/\delta)/\beta^2]$ samples. Comparing to the way that the conventional distillation rate $D^\varepsilon(\rho)$ is involved in the sample number of ρ motivates us to introduce the virtual resource distillation rate as

$$V^\varepsilon(\rho) := \sup_m \frac{m}{C^\varepsilon(\rho, m)^2}, \tag{9}$$

where $C^\varepsilon(\rho, m)$ is the virtual resource distillation overhead defined by

$$\begin{aligned}C^\varepsilon(\rho, m) := \inf\{\lambda_+ + \lambda_- \mid &\tfrac{1}{2}\|\tau^{\otimes m} - [\lambda_+\Lambda_+(\rho) \\ &- \lambda_-\Lambda_-(\rho)]\|_1 \leqslant \varepsilon,\ \lambda_\pm \geqslant 0, \\ &\lambda_+ - \lambda_- = 1,\ \Lambda_\pm \in O\}.\end{aligned} \tag{10}$$

In the above, we used inf rather than min to implicitly allow for pathological situations where a feasible decomposition of $\tilde{\tau}$ does not exist, and hence $C^\varepsilon(\rho, m) = \infty$. As long as the optimization is feasible, as is the case in most of the practically encountered cases, the optimum will always be achieved for all closed sets of operations O.

We note that the above procedure can be easily adapted to virtually simulate the expectation value of any observable M' acting on the many-copy state $\tau^{\otimes m}$; above, we only studied single-copy measurements that provide a natural justification to our definition of the distillation rate V^ε, but the approach itself is much more general.

The above argument can also be extended to resource theories of quantum channels and combs. For channel theories, consider a quantum information processing protocol described by a superchannel Ξ applying to a certain resource channel $\mathcal{A}$ such that $\Xi(\mathcal{A})$ is a quantum state corresponding

to the final state in the algorithm right before the terminating measurement M. Then

$$|\mathrm{Tr}[M\,\Xi(\mathcal{A})] - \mathrm{Tr}[M\,\Xi(\mathcal{A}')]| \leqslant \tfrac{1}{2}\|\Xi(\mathcal{A}) - \Xi(\mathcal{A}')\|_1 \leqslant \tfrac{1}{2}\|\mathcal{A} - \mathcal{A}'\|_\diamond, \tag{11}$$

where in the last inequality we used that the diamond norm does not increase under superchannels and that the diamond distance for quantum states reduces to the trace distance. This allows us to approximate $\mathcal{A}^{\otimes m}$ from a given channel $\mathcal{E}$ by $\lambda_+\Lambda_+(\mathcal{E}) - \lambda_-\Lambda_-(\mathcal{E})$, where $\Lambda_\pm \in \boldsymbol{O}$ are free superchannels.

The case for resource theories of quantum combs goes similarly. A quantum information processing protocol can now be considered as a comb Υ applied to a certain resource comb Θ. Then the error in expectation value for using another comb Θ' instead of Θ is bounded by

$$|\mathrm{Tr}[M\,\Upsilon(\Theta)] - \mathrm{Tr}[M\,\Upsilon(\Theta')]| \leqslant \tfrac{1}{2}\|\Upsilon(\Theta) - \Upsilon(\Theta')\|_1 \leqslant \tfrac{1}{2}\|\Theta - \Theta'\|_c, \tag{12}$$

where the fact that the comb distance does not increase under the application of another comb is apparent from the definition of the comb distance. This similarly allows us to approximate copies of the $\Theta^{\otimes m}$ from a given comb Υ by $\lambda_+\Lambda_+(\Upsilon) - \lambda_-\Lambda_-(\Upsilon)$, where $\Lambda_\pm \in \boldsymbol{O}$ are free combs.

These observations can be summarized as the following definition of virtual resource distillation rate and overhead that can be applied to general types of resource objects.

Definition 1. For a resource theory with a set $\boldsymbol{O}$ of free operations, the virtual resource distillation rate of a given object X with respect to the target object T is

$$V^\varepsilon(X) := \sup_m \frac{m}{C^\varepsilon(X,m)^2}, \tag{13}$$

with the virtual resource distillation overhead $C^\varepsilon(X,m)$ defined by

$$C^\varepsilon(X,m) := \inf\{\lambda_+ + \lambda_- \mid T^{\otimes m} \sim_\varepsilon \lambda_+\Lambda_+(X) - \lambda_-\Lambda_-(X),\ \lambda_\pm \geqslant 0,\ \lambda_+ - \lambda_- = 1,\ \Lambda_\pm \in \boldsymbol{O}\}. \tag{14}$$

We remark that the overhead is equivalently written as

$$C^\varepsilon(X,m) = \inf\left\{ \sum_i |\lambda_i| \,\middle|\, T^{\otimes m} \sim_\varepsilon \sum_i \lambda_i\Lambda_i(X),\ \lambda_i \in \mathbb{R}\ \forall\, i,\ \sum_i \lambda_i = 1,\ \Lambda_i \in \boldsymbol{O}\ \forall\, i \right\}. \tag{15}$$

The form in Definition 1 is recovered by letting $\lambda_+ := \sum_{i:\lambda_i\geqslant 0}\lambda_i$ and $\lambda_- := \sum_{i:\lambda_i<0}(-\lambda_i)$, as well as $\Lambda_+ := \lambda_+^{-1}\sum_{i:\lambda_i\geqslant 0}\lambda_i\Lambda_i$ and $\Lambda_- := \lambda_-^{-1}\sum_{i:\lambda_i<0}(-\lambda_i)\Lambda_i$, and noting that $\Lambda_\pm \in \boldsymbol{O}$ follows due to the convexity of $\boldsymbol{O}$.

Since the conventional distillation can be reconstructed with a restriction $\lambda_- = 0$, we always have $D^\varepsilon(X) \leqslant V^\varepsilon(X)$. We will see later that this inequality is strict in many cases.

We stress here that this framework is conceptually very different from many-copy distillation protocols, which are often encountered in practical applications of conventional distillation. Specifically, at all stages of the virtual distillation process, only single-copy operations are used and no joint channels acting on $\rho^{\otimes n}$ are needed.

We note also a superficial conceptual similarity to a recent approach of [58], where distillation under non-completely-positive resource manipulation protocols was considered; however, that framework does not require the operations to be implementable through a classical postprocessing of physical (completely positive) free operations, yielding a setting that may be difficult to directly compare with ours.

We further remark that several protocols known as virtual cooling [5] and virtual distillation in quantum error mitigation [6,7] share a similar idea that classical postprocessing enables us to simulate purer quantum states, although they differ from our framework: These techniques extract the measurement statistics for purer quantum objects by coherently interacting multiple copies of noisy objects, while our virtual resource distillation applies a probabilistic operation to a single copy of the noisy quantum object, which is inspired by and related to the techniques for simulating unphysical objects [8–10], and error mitigation technique based on quasiprobability [11].

C. Probabilistic distillation

A more general form of distillation protocols is one that can succeed only with some probability. Here we describe the basic setting, compare it with virtual distillation, and discuss the possibility of extending virtual resource distillation to probabilistic protocols.

Let us begin with conventional distillation in state theories. General probabilistic operations are represented by subchannels (completely positive trace-nonincreasing maps). Any such map can be thought of as being part of a quantum instrument, that is, a collection of probabilistic operations $\{\Lambda_i\}_i$ such that the overall transformation $\sum_i \Lambda_i$ is trace preserving. The outcome i, obtained with probability $p_i := \mathrm{Tr}\Lambda_i(\rho)$, then corresponds to the final state $\Lambda_i(\rho)/p_i$. Among such probabilistic operations, we define a subset $\boldsymbol{O}_{\leqslant 1}$ of subchannels and call it free subchannels. Then probabilistic distillation is a process that transforms a given state to a desired target state with $\boldsymbol{O}_{\leqslant 1}$ with some probability.

The distillation process should, upon success, output a state close to the target state, which can then be measured to estimate the expectation value of interest. The experimenters, knowing if the distillation did or did not succeed, can postselect only the successful outcomes of the process. To collect a sufficient number of samples to estimate the expectation value with the desired accuracy, one needs to use a number of samples that is inversely proportional to the success probability. This motivates the definition of a probabilistic one-shot distillation rate with respect to the target state τ as

$$D_p^\varepsilon(\rho) := \sup_{\Lambda^p\in\boldsymbol{O}_{\leqslant 1}} \left\{ m\mathrm{Tr}[\Lambda^p(\rho)] \,\middle|\, \frac{\Lambda^p(\rho)}{\mathrm{Tr}[\Lambda^p(\rho)]} \sim_\varepsilon \tau^{\otimes m} \right\}. \tag{16}$$

Here the superscript p denotes the probabilistic nature of subchannels.

Such an approach is seemingly very similar to virtual distillation: Multiple samples are taken by applying free operations to a single copy of ρ, and a protocol can only succeed by collecting a sufficient number of them. As we showed in [12], virtual distillation can offer strict improvements over proba-

bilistic distillation, and in particular there exist cases when $D_p^\varepsilon(\rho) = 0 < V^\varepsilon(\rho)$. However, the fact that free subchannels $O_{\leqslant 1}$ are employed in probabilistic distillation, which are, in general, a strictly larger class of maps than O, means that a direct comparison between the probabilistic one-shot rate D_p^ε and the virtual rate V^ε may not be possible in general.

Let us then formalize an explicit extension of virtual distillation to subchannels, which we will allow for a more direct comparison with conventional probabilistic approaches. Suppose that the target state can be written as

$$\tau^{\otimes m} \sim_\varepsilon \lambda_+ \frac{\Lambda^p_{+(\rho)}}{\mathrm{Tr}[\Lambda^p_+(\rho)]} - \lambda_- \frac{\Lambda^p_{-(\rho)}}{\mathrm{Tr}[\Lambda^p_{-(\rho)}]} \tag{17}$$

for $\lambda_\pm \geqslant 0$ and $\Lambda^p_\pm \in O_{\leqslant 1}$. Then extending the deterministic virtual resource distillation introduced above, the expectation value can be estimated in the following manner.

(1) Flip a biased coin that lands on heads with probability $p_+ := \frac{\lambda_+}{\lambda_+ + \lambda_-}$ and on tails with probability $p_- := \frac{\lambda_-}{\lambda_+ + \lambda_-}$.

(2) When we see heads, apply Λ^p_+ to ρ. If failure is reported, start over from step 1. If successful, measure m commuting observables $M \otimes I^{\otimes m-1}, I \otimes M \otimes I^{\otimes m-2}, \ldots, I^{\otimes m-1} \otimes M$ to get outcomes $o_1, \ldots, o_m$. Store the value $\gamma o_1, \ldots, \gamma o_m$ with $\gamma := \lambda_+ + \lambda_-$ to the classical register. If we see tails, apply Λ^p_- to ρ and follow the same procedure.

(3) Repeat the above process and take the sample average of the values stored in the classical register.

Note that postselection is involved in the second step, which makes the protocol probabilistic. This process provides an estimator with bias ε. The number of samples to use scales with γ^2 by the same mechanism for the deterministic case, as well as the average success probability in step 2, $\frac{\lambda_+ \mathrm{Tr}[\Lambda^p_+(X)] + \lambda_- \mathrm{Tr}[\Lambda^p_-(X)]}{\lambda_+ + \lambda_-}$, which linearly contributes to the sampling cost. This motivates us to introduce the probabilistic virtual distillation rate with respect to a target state τ defined as

$$V_p^\varepsilon(\rho) := \sup_{\Lambda^p_\pm \in O_{\leqslant 1}} \left\{ \frac{m\{\lambda_+ \mathrm{Tr}[\Lambda^p_+(\rho)] + \lambda_- \mathrm{Tr}[\Lambda^p_-(\rho)]\}}{(\lambda_+ + \lambda_-)^3} \,\middle|\, \lambda_\pm \geqslant 0,\ \tau^{\otimes m} \sim_\varepsilon \lambda_+ \frac{\Lambda^p_{+(\rho)}}{\mathrm{Tr}[\Lambda^p_+(\rho)]} - \lambda_- \frac{\Lambda^p_-(\rho)}{\mathrm{Tr}[\Lambda^p_-(\rho)]},\ \lambda_+ - \lambda_- = 1 \right\}. \tag{18}$$

Since probabilistic distillation could outperform deterministic distillation, we expect that probabilistic virtual distillation also outperforms deterministic virtual distillation. We leave a detailed study of this advantage to future work.

Alternatively, the expectation value can be estimated without postselection. Let us rewrite (17) with $\lambda_\pm \to \lambda_\pm \mathrm{Tr}[\Lambda^p_\pm(\rho)]$, which gives $\tau^{\otimes m} = \lambda_+ \Lambda^p_+(\rho) - \lambda_- \Lambda^p_-(\rho)$. With this $\lambda_\pm$, we follow the same procedure as above, except that instead of postselecting on the successful events in step 2, we store the value 0 upon failure. This gives the virtual distillation rate without postselection as

$$\tilde{V}_p^\varepsilon(\rho) := \sup_{\Lambda^p_\pm \in O_{\leqslant 1}} \left\{ \frac{m}{(\lambda_+ + \lambda_-)^2} \,\middle|\, \tau^{\otimes m} \sim_\varepsilon \lambda_+ \Lambda^p_+(\rho) - \lambda_- \Lambda^p_-(\rho),\ \lambda_\pm \geqslant 0,\ \lambda_+ \mathrm{Tr}[\Lambda^p_+(\rho)] - \lambda_- \mathrm{Tr}[\Lambda^p_-(\rho)] = 1 \right\}. \tag{19}$$

We remark that the values of $\lambda_\pm$ in (18) do not explicitly depend on $\mathrm{Tr}[\Lambda^p_\pm(\rho)]$ as $\lambda_+ - \lambda_- = 1$. On the other hand, the values for $\lambda_\pm$ in (19) are larger than those in (18) by a factor of $1/\mathrm{Tr}[\Lambda^p_\pm(\rho)]$. This makes V_p^ε scale with $\mathrm{Tr}[\Lambda^p_\pm(\rho)]$ while $\tilde{V}_p^\varepsilon$ scales with $\{\mathrm{Tr}[\Lambda^p_\pm(\rho)]\}^2$, reflecting the absence of postselection. Therefore, $\tilde{V}_p^\varepsilon$ becomes significantly smaller than V_p^ε when the success probability of the free subchannels is small.

The discussion becomes more involved for channel theories. The probabilistic channel transformation can be formalized by subsuperchannels, which transform channels to subchannels even when acting only on a part of a larger system [59]. The difference from the case of state theories is that the success probability of the protocol depends on not only the description of the subchannel but also input states. Therefore, to ensure that the resultant channel is close to the target channel upon success, we need to make sure that all output states are close to the desired final states upon success [31].

The number of samples to ensure the estimation of expectation values with the desired accuracy also depends on the success probability, which depends on input states. It is therefore reasonable to take the worst-case scenario and define the probabilistic distillation rate of a channel $\mathcal{E}$ with respect to the target channel $\mathcal{A}$ as

$$D_p^\varepsilon(\mathcal{E}) := \sup_{\Lambda \in O_{\leqslant 1}} \min_\rho \left\{ m\mathrm{Tr}[\mathrm{id} \otimes \Lambda^p(\mathcal{E})(\rho)] \,\middle|\, \frac{\mathrm{id} \otimes \Lambda^p(\mathcal{E})(\sigma)}{\mathrm{Tr}[\mathrm{id} \otimes \Lambda^p(\mathcal{E})(\sigma)]} \sim_\varepsilon \mathrm{id} \otimes \mathcal{A}^{\otimes m}(\sigma)\, \forall\, \sigma \right\}, \tag{20}$$

with the minimization being over all possible input states ρ.

Analogously, the probabilistic virtual resource distillation rate with postselection can be written as

$$V_p^\varepsilon(\mathcal{E}) := \sup_{\Lambda^p_\pm \in O_{\leqslant 1}} \min_\rho \left\{ \frac{m\{\lambda_+ \mathrm{Tr}[\mathrm{id} \otimes \Lambda^p_+(\mathcal{E})(\rho)] + \lambda_- \mathrm{Tr}[\mathrm{id} \otimes \Lambda^p_-(\mathcal{E})(\rho)]\}}{(\lambda_+ + \lambda_-)^3} \,\middle|\, \mathrm{id} \otimes \mathcal{A}^{\otimes m} \sim_\varepsilon \lambda_+ \frac{\mathrm{id} \otimes \Lambda^p_+(\mathcal{E})(\sigma)}{\mathrm{Tr}[\mathrm{id} \otimes \Lambda^p_+(\mathcal{E})(\sigma)]} - \lambda_- \frac{\mathrm{id} \otimes \Lambda^p_-(\mathcal{E})(\sigma)}{\mathrm{Tr}[\mathrm{id} \otimes \Lambda^p_-(\mathcal{E})(\sigma)]}\, \forall\, \sigma,\ \lambda_+ - \lambda_- = 1,\ \lambda_\pm \geqslant 0 \right\}. \tag{21}$$

Unlike the state case, the probabilistic virtual resource distillation rate without postselection does not work in general, as

$$\mathrm{id}\otimes\mathcal{A}^{\otimes m}(\sigma)\sim_{\varepsilon}\lambda_+\mathrm{id}\otimes\Lambda_+^p(\mathcal{E})(\sigma)-\lambda_-\mathrm{id}\otimes\Lambda_-^p(\mathcal{E})(\sigma)\,\forall\,\sigma \tag{22}$$

is not generally satisfied when $\mathrm{Tr}[\mathrm{id}\otimes\Lambda_\pm^p(\mathcal{E})(\sigma)]$ differs depending on σ. Indeed, if we take the trace on both sides, the left-hand side always gives $\mathrm{Tr}[\mathrm{id}\otimes\mathcal{A}^{\otimes m}(\sigma)]=1\,\forall\,\sigma$ while the right-hand side can vary for different states σ.

An analogous extension can be made to the resource theories of combs, where a similar subtlety about the success probability depending on the input channels and states remains.

In the rest of this paper, we focus on the deterministic virtual rates defined through V^ε and C^ε, which are more easily characterizable than the probabilistic virtual rates while at the same time being sufficiently general to allow for considerable advantages over conventional distillation.

D. Estimation of probability distribution

The above discussion shows that when there exist $\Lambda_\pm\in O$ and $\lambda_\pm\geqslant 0$ such that $T\sim_\varepsilon\lambda_+\Lambda_+(X)-\lambda_-\Lambda_-(X)$, any expectation value of a target object T can be obtained by measuring $\Lambda_\pm(X)$ instead. Here we apply this observation to the estimation of probability distributions of T.

Let η be an arbitrary output state resulting from a target object T. For state theories, η coincides with T, while for channel and comb theories η is an output from T with an arbitrary input state. Suppose that we measure η in the computational basis. For each measurement, we get a one-shot measurement outcome o_j with probability $p(j)=\mathrm{Tr}(\eta|j\rangle\langle j|)$. After N independent measurements, we will get $n(j)$ counts for outcome o_j with $\sum_j n(j)=N$, and according to the Hoeffding inequality [57], with failure probability $\delta\in(0,1)$, we have

$$|p(j)-\tilde{p}(j)|=O\left(\sqrt{\frac{\log_2\delta}{N}}\right), \tag{23}$$

where $\tilde{p}(j)=n(j)/N$.

Next we consider how to simulate this measurement process with $\Lambda_\pm(X)$. Since the projector $|j\rangle\langle j|$ is also an observable, we can directly apply the virtual resource distillation framework developed above. Let $\eta_\pm$ be output states from $\Lambda_\pm(X)$. To ensure the same accuracy, we use $(\lambda_++\lambda_-)^2N$ copies of X. With probability $\lambda_\pm/(\lambda_++\lambda_-)$, we measure $\eta_\pm$ in the computational basis, multiplying the outcomes $+1$ (click) or 0 (no click) by $\pm(\lambda_++\lambda_-)$ for each j and take the sample average. This is equivalent to measuring $\eta_\pm$ for $N_\pm:=(\lambda_++\lambda_-)^2\frac{\lambda_\pm}{\lambda_++\lambda_-}N=\lambda_\pm(\lambda_++\lambda_-)N$ times, where our estimate for the probability is

$$p'(j):=(\lambda_++\lambda_-)\left(\frac{n_+(j)}{N_+}-\frac{n_-(j)}{N_-}\right), \tag{24}$$

with $n_\pm(j)$ standing for the number of times the outcome o_j is observed among $N_\pm$ measurements. The general framework of virtual resource distillation developed above ensures that

$$|p(j)-p'(j)|=O\left(\sqrt{\frac{\log_2\delta}{N}}\right)+\varepsilon. \tag{25}$$

This also implies that the measurement counts $n(j)$ can be approximated as

$$n(j)\approx Np'(j)=\frac{n_+}{\lambda_+}-\frac{n_-}{\lambda_-}, \tag{26}$$

with negligible error and failure probability.

E. Virtual resource monotones

The virtual distillation rate is an operationally motivated quantity and may be hard to evaluate exactly for some settings. Therefore, it will be useful to establish other quantities that can help evaluate the virtual distillation rate. Here we introduce a notion of a monotone that can always be used to bound the virtual distillation overhead with zero error.

Proposition 1. Let M be a function that obeys the following properties: (a) $M(X)\geqslant M(\mu_+\Lambda_+(X)-\mu_-\Lambda_+(X))\,\forall\,\Lambda_+,\Lambda_-\in O,\ \mu_++\mu_-=1$, and (b) $M(\mu X)=\mu M(X)\,\forall\,\mu>0$. Then

$$C^0(X,m)\geqslant\frac{M(T^{\otimes m})}{M(X)}. \tag{27}$$

Proof. For any operations $\Lambda_\pm\in O$ and any $\lambda_\pm>0$, we can write

$$\begin{aligned}&M(\lambda_+\Lambda_+(X)-\lambda_-\Lambda_-(X))\\&=M\left((\lambda_++\lambda_-)\frac{\lambda_+\Lambda_+(X)-\lambda_-\Lambda_-(X)}{\lambda_++\lambda_-}\right)\\&=(\lambda_++\lambda_-)M\left(\frac{\lambda_+\Lambda_+(X)-\lambda_-\Lambda_-(X)}{\lambda_++\lambda_-}\right)\\&\leqslant(\lambda_++\lambda_-)M(X),\end{aligned} \tag{28}$$

where the last two lines are by definition of a virtual monotone. Optimizing over all virtual operations yields the stated result. ■

The following are examples of virtual resource monotones. (i) In the resource theory of entanglement, the base norm $\|\rho\|_{\mathcal{S}}=\min\{\mu_++\mu_-\mid\rho=\mu_+\sigma_+-\mu_-\sigma_-,\ \sigma_\pm\in\mathcal{S}\}$, where $\mathcal{S}$ is the set of separable states, is a virtual monotone under all separability-preserving operations. This quantity is directly related to the (standard) robustness of entanglement [60]. (ii) Also in the resource theory of entanglement, the negativity $\|\rho^\Gamma\|_1$ [61] is a virtual monotone under all positive partial transpose (PPT) operations. (iii) In the resource theory of coherence, the ℓ_1 norm of coherence $\|\rho\|_{\ell_1}$ [62] is a virtual monotone under all incoherent operations. One special virtual resource monotone is the inverse virtual distillation overhead

$$\tilde{M}(X,m):=\frac{1}{C^0(X,m)}. \tag{29}$$

To prove the first property, we define $X' = \lambda_+\Lambda_+(X) - \lambda_-\Lambda_+(X)$ and have

$$\begin{aligned} &C^0(X', m) \\ &\quad = \inf\{\tilde{\lambda}_+ + \tilde{\lambda}_- \mid \tilde{\lambda}_+\tilde{\Lambda}_+(X') - \tilde{\lambda}_-\tilde{\Lambda}_+(X') = T^{\otimes m}\}, \\ &\quad = \inf\{\tilde{\lambda}_+ + \tilde{\lambda}_- \mid (\tilde{\lambda}_+\lambda_+\tilde{\Lambda}_+ \circ \Lambda_+ + \tilde{\lambda}_-\lambda_-\tilde{\Lambda}_- \circ \Lambda_-)(X) \\ &\qquad - (\tilde{\lambda}_+\lambda_-\tilde{\Lambda}_+ \circ \Lambda_- + \tilde{\lambda}_-\lambda_+\tilde{\Lambda}_- \circ \Lambda_+)(X) = T^{\otimes m}\}. \end{aligned} \tag{30}$$

Note that

$$\tilde{\lambda}_+ + \tilde{\lambda}_- = \tilde{\lambda}_+\lambda_+ + \tilde{\lambda}_-\lambda_- + \tilde{\lambda}_+\lambda_- + \tilde{\lambda}_-\lambda_+; \tag{31}$$

thus we have

$$C^0(X', m) \geqslant C^0(X, m) \tag{32}$$

and hence

$$\tilde{M}(X, m) \geqslant \tilde{M}(X', m). \tag{33}$$

It is obvious that $\tilde{M}(\lambda X, m) = \lambda\tilde{M}(X, m)$. According to Eq. (27), we have

$$M(X) \geqslant M(T^{\otimes m}) \cdot \tilde{M}(X, m). \tag{34}$$

Considering normalized virtual monotones with $M(T^{\otimes m}) = 1$, we thus have

$$M(X) \geqslant \tilde{M}(X, m), \tag{35}$$

that is, the specific virtual monotone $\tilde{M}(X, m)$ lower bounds all normalized virtual monotones.

V. EVALUATION OF VIRTUAL RESOURCE DISTILLATION OVERHEAD

Evaluating the virtual resource distillation rate is generally a formidable task, mainly due to the optimization over the number m of copies of the target object. Here we show that a closely related quantity, namely, the virtual resource distillation overhead $C^\varepsilon(\rho, m)$ for fixed m, can be efficiently characterized in general state theories.

We first present a useful alternative form of the distillation overhead. The assumption that O is convex ensures that the overhead $C^\varepsilon(\rho, m)$ is a solution of a convex optimization program. Therefore, taking the convex dual (see the Appendix), we obtain an alternative expression for state theories as

$$\begin{aligned} C^\varepsilon(\rho, m) = \inf_{\substack{\tilde{\tau} \sim_\varepsilon \tau^{\otimes m} \\ \mathrm{Tr}(\tilde{\tau})=1}} \sup_{W \in \mathrm{Herm}} &\{2\mathrm{Tr}(W\tilde{\tau}) - 1 \mid 0 \\ &\leqslant \mathrm{Tr}[W\Lambda(\rho)] \leqslant 1 \,\forall\, \Lambda \in O\}, \end{aligned} \tag{36}$$

where Herm is the set of Hermitian operators. This form can be extended to channel and combs theories by considering Choi operators in place of quantum states.

The overhead with $\varepsilon = 0$ can be computed by linear or semidefinite programming if the structure of the free operations is sufficiently simple. For example, if the set of free objects is the convex hull of a finite number of objects $\{f_i\}_i$, an operation Λ is resource nongenerating if and only if

$$\Lambda(f_i) = \sum_j p_j f_j \,\forall i, \quad p_j \geqslant 0, \ \sum_j p_j = 1, \ f_j \in \mathcal{F}. \tag{37}$$

Then the requirement of a resource nongenerating operation is characterized by a linear constraint, making the overhead computable by linear programming. This includes resource theories such as purity, thermodynamics, coherence, and magic.

There also exist resource theories with the infinite number of extreme free resource objects, such as the theory of entanglement or its dynamical counterpart, the theory of quantum memory. We then need to relax the requirement of free operations by considering free objects either induced from a finite number of extreme objects or characterizable via a semidefinite constraint, making the overhead computable by semidefinite programming (SDP). Both ways will give an upper bound to the overhead and the bound can be tight with better approximations.

As an example, we write down the SDP for entanglement under PPT operations, i.e., the set of bipartite channels $AB \to A'B'$ whose Choi operators are PPT across the bipartition $AA' : BB'$. If we suppose the unnormalized Choi operator of a PPT channel $\mathcal{N}$ is $J_{\mathcal{N}}^{A'ABB'}$, it should satisfy

$$J_{\mathcal{N}}^{A'ABB'} \geqslant 0, \quad \mathrm{Tr}_{AB}\big(J_{\mathcal{N}}^{A'ABB'}\big) = I_{A'B'}, \quad \big(J_{\mathcal{N}}^{A'ABB'}\big)^{T_{BB'}} \geqslant 0. \tag{38}$$

For any input state ρ_{AB}, the output state is

$$\mathcal{N}(\rho_{AB}) = \mathrm{Tr}_{A'B'}\big(\rho_{A'B'}^T \cdot J_{\mathcal{N}}^{A'ABB'}\big). \tag{39}$$

Then the virtual distillation overhead of ρ_{AB} with respect to the Bell state $\Phi := |\Phi\rangle\langle\Phi|$ with $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is

$$\begin{aligned} C^\varepsilon(\rho_{AB}) = \min \big\{ &\lambda_+ + \lambda_- \mid \mathrm{Tr}_{A'B'}\big(\rho_{A'B'}^T \cdot J_{\Lambda_+}^{A'ABB'}\big) \\ &- \mathrm{Tr}_{A'B'}\big(\rho_{A'B'}^T \cdot J_{\Lambda_-}^{A'ABB'}\big) = \Phi, \\ &J_{\Lambda_+}^{A'ABB'}, J_{\Lambda_-}^{A'ABB'} \geqslant 0, \mathrm{Tr}_{AB}\big(J_{\Lambda_+}^{A'ABB'}\big) \\ &= \lambda_+ I_{A'B'}, \ \mathrm{Tr}_{AB}\big(J_{\Lambda_-}^{A'ABB'}\big) = \lambda_- I_{A'B'} \\ &\big(J_{\Lambda_+}^{A'ABB'}\big)^{T_{BB'}}, \ \big(J_{\Lambda_-}^{A'ABB'}\big)^{T_{BB'}} \geqslant 0\big\}. \end{aligned} \tag{40}$$

A. Tight bounds for general resources of quantum states

We will now show that much simpler bounds based on convex and semidefinite programming can be obtained for state theories. This will remove the need to optimize over all free operations and apply also to resource theories such as quantum entanglement. We introduce general upper and lower bounds on the distillation overhead in general quantum resource theories. They depend on several resource measures, whose definitions we now recall.

For a set $\mathcal{F}$ of free states, define the generalized robustness $R_{\mathcal{F}}^g$ [60,63,64], the standard robustness $R_{\mathcal{F}}^s$ [60], and the resource fidelity $F_{\mathcal{F}}$ as

$$\begin{aligned} R_{\mathcal{F}}^g(\rho) &:= \inf\left\{\lambda \,\middle|\, \frac{\rho + \lambda\omega}{1+\lambda} \in \mathcal{F}, \ \omega \in \mathcal{D}\right\}, \\ R_{\mathcal{F}}^s(\rho) &:= \inf\left\{\lambda \,\middle|\, \frac{\rho + \lambda\sigma}{1+\lambda} \in \mathcal{F}, \ \sigma \in \mathcal{F}\right\}, \\ F_{\mathcal{F}}(\rho) &:= \max_{\sigma \in \mathcal{F}} F(\rho, \sigma), \end{aligned} \tag{41}$$

respectively, where $\mathcal{D}$ is the set of quantum states and F is the fidelity. We remark that when ρ is a pure state, it holds

that $F_{\mathcal{F}}(\rho) = \max_{\sigma\in\mathcal{F}} \mathrm{Tr}(\rho\sigma)$. Define now the optimization problem

$$\zeta_\varepsilon^s(\rho,k) := \min\Big\{\mu_+ + \mu_- \,\Big|\, 0 \leqslant Q_+ \leqslant \mu_+ I,\ 0 \leqslant Q_- \leqslant \mu_- I, \mathrm{Tr}Q_+\sigma \leqslant \frac{\mu_+}{k}\,\forall\,\sigma\in\mathcal{F},\ \mathrm{Tr}Q_-\sigma \leqslant \frac{\mu_-}{k}\,\forall\,\sigma\in\mathcal{F},$$
$$\mu_+ - \mu_- = 1,\ \mathrm{Tr}\rho(Q_+ - Q_-) \geqslant 1-\varepsilon\Big\}, \tag{42}$$

where k is some parameter to be fixed. We also define $\zeta_\varepsilon^g(\rho,k)$ to be the same optimization except that the inequality constraints $\mathrm{Tr}(Q_+\sigma) \leqslant \mu_+/k$ and $\mathrm{Tr}(Q_-\sigma) \leqslant \mu_-/k\,\forall\,\sigma\in\mathcal{F}$ become equality constraints. Then we obtain the following general lower and upper bounds.

Theorem 1 (Theorem 1 in [12]). Consider a convex resource theory and a target pure resource state ψ. Let O be the class of resource nongenerating operations. If $R_{\mathcal{F}}^s(\psi) < \infty$, then

$$\zeta_\varepsilon^s(\rho, F_{\mathcal{F}}(\psi^{\otimes m})^{-1}) \leqslant C^\varepsilon(\rho,m) \leqslant \zeta_\varepsilon^s\big(\rho, R_{\mathcal{F}}^s(\psi^{\otimes m})+1\big). \tag{43}$$

Furthermore, if it holds that $\langle\psi|\sigma|\psi\rangle$ is constant for all $\sigma\in\mathcal{F}$, then

$$\zeta_\varepsilon^g(\rho, F_{\mathcal{F}}(\psi^{\otimes m})^{-1}) \leqslant C^\varepsilon(\rho,m) \leqslant \zeta_\varepsilon^g\big(\rho, R_{\mathcal{F}}^g(\psi^{\otimes m})+1\big). \tag{44}$$

The crucial property of the bounds is that whenever $R_{\mathcal{F}}^s(\psi^{\otimes m})+1 = F_{\mathcal{F}}(\psi^{\otimes m})^{-1}$, which is true in resource theories such as bi- and multipartite entanglement [50,60] or multilevel quantum coherence [65], or if $R_{\mathcal{F}}^g(\psi^{\otimes m})+1 = F_{\mathcal{F}}(\psi^{\otimes m})^{-1}$ and the overlap $\langle\psi|\sigma|\psi\rangle$ is constant, which is true in resource theories such as coherence or athermality, then the upper and lower bounds coincide, yielding an exact expression for the overhead $C^\varepsilon(\rho,m)$.

We will later consider specific examples of resource theories, showing how the result can be applied in different contexts and in some cases improving on and extending the statement of Theorem 1.

Proof. We first prove (43). Consider any feasible distillation protocol such that $\Lambda_\pm \in O$ and $\frac{1}{2}\|\lambda_+\Lambda_+(\rho) - \lambda_-\Lambda_-(\rho) - \psi^{\otimes m}\|_1 \leqslant \varepsilon$. Define $Q_\pm = \lambda_\pm \Lambda_\pm^\dagger(\psi^{\otimes m})$ and $\mu_\pm = \lambda_\pm$. Since $\Lambda_\pm$ are free operations, it holds that $\Lambda_\pm(\sigma) \in \mathcal{F}$ for any $\sigma\in\mathcal{F}$ and hence

$$\begin{aligned}\max_{\sigma\in\mathcal{F}} \mathrm{Tr}Q_\pm\sigma &= \mu_\pm \max_{\sigma\in\mathcal{F}} \mathrm{Tr}\psi^{\otimes m}\Lambda_\pm(\sigma)\\ &\leqslant \mu_\pm \max_{\sigma'\in\mathcal{F}} \mathrm{Tr}\psi^{\otimes m}\sigma'\\ &\leqslant \mu_\pm F_{\mathcal{F}}(\psi^{\otimes m}).\end{aligned} \tag{45}$$

Due to the fact that $\Lambda_\pm$ are completely positive and trace-preserving (CPTP) maps, we also get $0 \leqslant Q_\pm \leqslant \mu_\pm I$, and the condition $\mathrm{Tr}\rho(Q_+ - Q_-) \geqslant 1-\varepsilon$ is ensured by the fact that

$$\begin{aligned}\varepsilon &\geqslant \tfrac{1}{2}\|\psi^{\otimes m} - \lambda_+\Lambda_+(\rho) + \lambda_-\Lambda_-(\rho)\|_1\\ &= \max\{\mathrm{Tr}[(\psi^{\otimes m} - \lambda_+\Lambda_+(\rho) + \lambda_-\Lambda_-(\rho))X]\,|\\ &\qquad 0 \leqslant X \leqslant I\}\\ &\geqslant \mathrm{Tr}[(\psi^{\otimes m} - \lambda_+\Lambda_+(\rho) + \lambda_-\Lambda_-(\rho))\psi^{\otimes m}]\\ &= 1 - \mathrm{Tr}\rho(Q_+ - Q_-).\end{aligned} \tag{46}$$

Therefore, $Q_\pm$ give a valid feasible solution to $\zeta_\varepsilon^s(\rho, F_{\mathcal{F}}(\psi^{\otimes m}))$ with optimal value $\mu_+ + \mu_- = \lambda_+ + \lambda_-$, which concludes the first part of the proof.

Conversely, let $Q_\pm$ be feasible solutions to $\zeta_\varepsilon^s(\rho, R_{\mathcal{F}}^s(\psi^{\otimes m})+1)$. Note that we can always take $Q_\pm$ such that $\mathrm{Tr}[(Q_+ - Q_-)\rho] = 1-\varepsilon$, since for any feasible $Q_\pm$ with $\mathrm{Tr}[(Q_+ - Q_-)\rho] = t(1-\varepsilon)$ for some $t > 1$, $\frac{1}{t}Q_\pm$ are also feasible with the same optimal value. Now define the quantum channels

$$\Lambda_\pm(\omega) = \mathrm{Tr}\left(\frac{Q_\pm}{\mu_+}\omega\right)\psi^{\otimes m} + \mathrm{Tr}\left[\left(I - \frac{Q_\pm}{\mu_\pm}\right)\omega\right]\sigma_\psi, \tag{47}$$

where $\sigma_\psi \in \mathcal{F}$ is a state such that

$$\frac{\psi^{\otimes m} + R_{\mathcal{F}}^s(\psi^{\otimes m})\sigma_\psi}{1 + R_{\mathcal{F}}^s(\psi^{\otimes m})} \in \mathcal{F}. \tag{48}$$

Note that

$$\begin{aligned}\Lambda_\pm(\sigma) &\propto \psi^{\otimes m} + \frac{\mathrm{Tr}\big[\big(I - \frac{Q_\pm}{\mu_\pm}\big)\sigma\big]}{\mathrm{Tr}\big(\frac{Q_\pm}{\mu_\pm}\sigma\big)}\sigma_\psi\\ &= \psi^{\otimes m} + \left(\frac{\mu_+}{\mathrm{Tr}Q_\pm\sigma} - 1\right)\sigma_\psi,\end{aligned} \tag{49}$$

which entails that, since $\mathrm{Tr}Q_\pm\sigma \leqslant \frac{\mu_+}{R_{\mathcal{F}}^s(\psi^{\otimes m})+1}$ for any $\sigma\in\mathcal{F}$, we necessarily have $\Lambda_\pm(\sigma)\in\mathcal{F}$ and thus $\Lambda_\pm$ are resource nongenerating operations. Now, since

$$\begin{aligned}&\mu_+\Lambda_+(\rho) - \mu_-\Lambda_-(\rho)\\ &\quad= \mathrm{Tr}[(Q_+ - Q_-)\rho]\psi^{\otimes m} + \{\mu_+ - \mu_-\\ &\qquad - \mathrm{Tr}[(Q_+ - Q_-)\rho]\}\sigma_\psi\\ &\quad= \mathrm{Tr}[(Q_+ - Q_-)\rho]\psi^{\otimes m} + \{1 - \mathrm{Tr}[(Q_+ - Q_-)\rho]\}\sigma_\psi,\end{aligned} \tag{50}$$

we get

$$\begin{aligned}&\|\mu_+\Lambda_+(\rho) - \mu_-\Lambda_-(\rho) - \psi^{\otimes m}\|_1\\ &\quad\leqslant 2|1 - \mathrm{Tr}[(Q_+ - Q_-)\rho]|\\ &\quad= 2\varepsilon,\end{aligned} \tag{51}$$

and thus we see that the maps realize the desired transformation with error $\frac{1}{2}\|\psi^{\otimes m} - \lambda_+\Lambda_+(\rho) + \lambda_-\Lambda_-(\rho)\|_1 \leqslant \varepsilon$, yielding $C^\varepsilon(\rho,m) \leqslant \zeta_\varepsilon^s(\rho, R_{\mathcal{F}}^s(\psi^{\otimes m})+1)$.

The proof for (44) proceeds analogously. To show the lower bounds, note that the inequalities in (45) become equalities due to the assumption that $\langle\psi|\sigma|\psi\rangle$ is constant for all $\sigma\in\mathcal{F}$. To show the upper bound, we choose in (47) a state ω_ψ which satisfies

$$\frac{\psi^{\otimes m} + R_{\mathcal{F}}^g(\psi^{\otimes m})\omega_\psi}{1 + R_{\mathcal{F}}^g(\psi^{\otimes m})} \in \mathcal{F} \tag{52}$$

instead of σ_ψ. ∎

B. Bounds in terms of the maximum overlap

In the case of state theories, we can give alternative expressions for general lower bounds for the virtual distillation overhead. They can be formulated in relation to how close the given resource state can be brought to the target state via free operations. To formalize this, we define the maximum overlap with a target pure state ψ as

$$f_O(\rho, m) := \max_{\Lambda \in O} \mathrm{Tr}[\Lambda(\rho)\psi^{\otimes m}]. \tag{53}$$

Then we obtain the following general bound in terms of the maximum overlap.

Proposition 2. Let ψ denote a pure target resource state and let O be an arbitrary convex and closed set of free operations. Then for every state ρ, positive integer m, and $\varepsilon \in [0, 1]$,

$$C^\varepsilon(\rho, m) \geqslant \max\left\{\frac{2(1-\varepsilon)}{f_O(\rho, m)} - 1, 1\right\} \tag{54}$$

holds.

Proof. Using the max-min inequality, the dual form of the overhead (36) can be lower bounded as

$$C^\varepsilon(\rho, m) \geqslant \sup_{W \in \mathrm{Herm}} \inf_{\substack{\mathrm{Tr}[\eta]=1 \\ \frac{1}{2}\|\eta - \psi^{\otimes m}\|_1 \leqslant \varepsilon}} \{2\,\mathrm{Tr}(\eta W) - 1 \mid 0 \leqslant \mathrm{Tr}[\Lambda(\rho)W] \leqslant 1 \,\forall\, \Lambda \in O\}. \tag{55}$$

We note that $W = f_O(\rho, m)^{-1}\psi^{\otimes m}$ is a feasible solution for the last optimization problem because $0 \leqslant f_O(\rho, m)^{-1}\mathrm{Tr}[\Lambda(\rho)\psi^{\otimes m}] \leqslant 1 \,\forall\, \Lambda \in O$. This ensures

$$C^\varepsilon(\rho, m) \geqslant \inf_{\substack{\mathrm{Tr}[\eta]=1 \\ \frac{1}{2}\|\eta - \psi^{\otimes m}\|_1 \leqslant \varepsilon}} [2 f_O(\rho, m)^{-1}\mathrm{Tr}(\eta\psi^{\otimes m}) - 1]. \tag{56}$$

Since every η such that $\mathrm{Tr}(\eta) = 1$ satisfies $\frac{1}{2}\|\eta - \psi^{\otimes m}\|_1 = \max_{0 \leqslant E \leqslant I} \mathrm{Tr}[(\eta - \psi^{\otimes m})E]$, choosing $E = \psi^{\otimes m}$ specifically results in

$$\tfrac{1}{2}\|\eta - \psi^{\otimes m}\|_1 \geqslant \mathrm{Tr}[(\psi^{\otimes m} - \eta)\psi^{\otimes m}] \quad = 1 - \mathrm{Tr}(\eta\psi^{\otimes m}). \tag{57}$$

Therefore, $\frac{1}{2}\|\eta - \psi^{\otimes m}\| \leqslant \varepsilon$ implies $\mathrm{Tr}(\eta\psi^{\otimes m}) \geqslant 1 - \varepsilon$. This allows us to lower bound the right-hand side of (56) to get

$$C^\varepsilon(\rho, m) \geqslant \frac{2(1-\varepsilon)}{f_O(\rho, m)} - 1. \tag{58}$$

The lower bound of 1 can be obtained by choosing $W = I$ in (55). ∎

Although the maximum overlap is operationally intuitive, it could be hard to compute or it might obscure the relation with the resourcefulness contained in ρ. To address this, we can employ recent results that connect the operational distillation performance and fundamental resource measures that are equipped with geometric viewpoints. To this end, we recall the weight resource measure defined for an arbitrary convex resource theory

$$W_\mathcal{F}(\rho) := \max\{w \mid \rho = w\sigma + (1-w)\tau,\ \sigma \in \mathcal{F},\ \tau \in \mathcal{D}\}. \tag{59}$$

Then the maximum overlap can be upper bounded using the generalized robustness and the weight measure as follows.

Lemma 1 (from [31,32]). For every convex and closed set $\mathcal{F}$ of free states and every set O of free operations,

$$f_O(\rho, m) \leqslant F_\mathcal{F}(\psi^{\otimes m})\left[R^g_\mathcal{F}(\rho) + 1\right] \tag{60}$$

and

$$f_O(\rho, m) \leqslant 1 - [1 - F_\mathcal{F}(\psi^{\otimes m})]W_\mathcal{F}(\rho). \tag{61}$$

Combining Proposition 2 and Lemma 1 immediately gives the following alternative lower bound for C^ε.

Corollary 1. Let ψ denote the pure target state and let O be an arbitrary convex and closed set of free operations. Then for every state ρ, positive integer m, and $\varepsilon \in [0, 1]$,

$$C^\varepsilon(\rho, m) \geqslant \frac{2(1-\varepsilon)}{F_\mathcal{F}(\psi^{\otimes m})\left[R^g_\mathcal{F}(\rho) + 1\right]} - 1 \tag{62}$$

and

$$C^\varepsilon(\rho, m) \geqslant \frac{2(1-\varepsilon)}{1 - [1 - F_\mathcal{F}(\psi^{\otimes m})]W_\mathcal{F}(\rho)} - 1. \tag{63}$$

The following result provides a sufficient condition when the bound in Proposition 2 is saturated.

Theorem 2 (Theorem 2 in [12]). For a pure target state ψ, suppose that there exists a free generalized twirling operation [52] $\mathcal{T} \in O$ of the form

$$\mathcal{T}(\cdot) = \mathrm{Tr}(\psi^{\otimes m}\cdot)\psi^{\otimes m} + \mathrm{Tr}[(I - \psi^{\otimes m})\cdot]\sigma^\star \tag{64}$$

for some $\sigma^\star \in \mathcal{F}$. Then

$$C^\varepsilon(\rho, m) = \max\left\{\frac{2(1-\varepsilon)}{f_O(\rho, m)} - 1, 1\right\} \tag{65}$$

for every $\varepsilon \in [0, 1]$.

Proof. We already showed $C^\varepsilon(\rho, m) \geqslant \max\{\frac{2(1-\varepsilon)}{f_O(\rho,m)} - 1, 1\}$ in (54) for the general case. To show the opposite inequality, let $\Lambda^\star \in O$ be the one that realizes $\mathrm{Tr}[\Lambda^\star(\rho)\psi^{\otimes m}] = f_O(\rho, m)$. Then the free twirling operation $\mathcal{T}$ maps $\Lambda^\star(\rho)$ to the generalized isotropic state with the same overlap with the target state as

$$\mathcal{T} \circ \Lambda^\star(\rho) = f_O(\rho, m)\psi^{\otimes m} + [1 - f_O(\rho, m)]\sigma^\star. \tag{66}$$

When $\frac{2(1-\varepsilon)}{f_O(\rho,m)} - 1 < 1 \iff 1 - f_O(\rho, m) < \varepsilon$, we have $\frac{1}{2}\|\psi^{\otimes m} - \mathcal{T} \circ \Lambda^\star(\rho)\|_1 \leqslant 1 - f_O(\rho, m) < \varepsilon$, implying $C^\varepsilon(\rho, m) \leqslant 1$. On the other hand, when $\frac{2(1-\varepsilon)}{f_O(\rho,m)} - 1 \geqslant 1$, we define

$$\Lambda_+ = \mathrm{id}, \quad \Lambda_-(\cdot) = \sigma^\star, \tag{67}$$

both of which are free operations. Then

$$\frac{1-\varepsilon}{f_O(\rho, m)}\Lambda_+ \circ \mathcal{T} \circ \Lambda^\star(\rho) - \left(\frac{1-\varepsilon}{f_O(\rho, m)} - 1\right) \times \Lambda_- \circ \mathcal{T} \circ \Lambda^\star(\rho) = (1-\varepsilon)\psi^{\otimes m} + \varepsilon\sigma^\star. \tag{68}$$

Since

$$\tfrac{1}{2}\|(1-\varepsilon)\psi^{\otimes m} + \varepsilon\sigma^\star - \psi^{\otimes m}\|_1 = \tfrac{1}{2}\|-\varepsilon\psi^{\otimes m} + \varepsilon\sigma^\star\|_1 \leqslant \tfrac{\varepsilon}{2}(\|\psi^{\otimes m}\|_1 + \|\sigma^\star\|_1) = \varepsilon, \tag{69}$$

where we used the triangle inequality, we get

$$C^{\varepsilon}(\rho,m) \leqslant \frac{1-\varepsilon}{f_O(\rho,m)} + \frac{1-\varepsilon}{f_O(\rho,m)} - 1 = \frac{2(1-\varepsilon)}{f_O(\rho,m)} - 1, \tag{70}$$

concluding the proof. ■

Although Theorem 2 provides a tight characterization of distillation overhead, its assumption may appear somewhat contrived, as it could be difficult to determine whether such a free generalized twirling operation exists for a given setting. First, we stress that the conditions are satisfied in a number of the most practically relevant resource theories; we will see this explicitly in Sec. VI.

We can also give a useful sufficient condition for a free twirling operation of Eq. (64) to exist: This is true whenever the fidelity-based measure and the standard robustness of the target state coincide (see Lemma 5 in [52]). This gives the following characterization, which is usually easier to verify directly.

Corollary 2. For a given set $\mathcal{F}$ of free states, consider the class O of resource nongenerating operations. If $F_{\mathcal{F}}(\psi^{\otimes m})^{-1} = 1 + R^s_{\mathcal{F}}(\psi^{\otimes m})$ holds, we have $C^{\varepsilon}(\rho,m) = \max\{\frac{2(1-\varepsilon)}{f_O(\rho,m)} - 1, 1\}$.

VI. EXAMPLES

Here we apply the general framework developed above to specific physical settings of interest. We primarily focus on evaluating virtual resource distillation overhead $C^{\varepsilon}(X,m)$ for a fixed m, which provides a lower bound for the virtual distillation rate $V^{\varepsilon}(X)$ and may allow for computing $V^{\varepsilon}(X)$ in the case when an analytical expression of $C^{\varepsilon}(X,m)$ is available.

A. Entanglement

Let us first consider the resource theory of entangled states, where separable states construct the set $\mathcal{S}$ of free states [1]. In entanglement theory, resource nongenerating operations are conventionally called separability-preserving (or nonentangling) operations, and any physically motivated set of operations such as the set of LOCC is a subset of separability-preserving operations.

The set of separable states does not allow an efficient characterization via semidefinite constraint and thus relevant resource measures are generally hard to compute. To obtain bounds for these quantities, it is often useful to consider a set $\mathcal{S}_{\mathrm{PPT}}$ of positive partial transpose states. Since every separable state is a PPT, $\mathcal{S}$ is a subset of $\mathcal{S}_{\mathrm{PPT}}$ and inclusion is known to be strict. The resource nongenerating operations for the set $\mathcal{S}_{\mathrm{PPT}}$ are called PPT-preserving maps. A relevant useful class which forms a subset of PPT-preserving maps is known as PPT operations, which are the channels whose Choi operators are PPTs.

As a target state ψ, we take the bipartite qubit Bell state $\Phi = |\Phi\rangle\langle\Phi|$ with $|\Phi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$.

1. General bounds

Let us first study the property for general states. We recall the hypothesis-testing relative entropy of entanglement [66] as

$$E_H^{\varepsilon}(\rho) = \max_{\substack{0\leqslant A\leqslant I \\ \mathrm{Tr}(A\rho)\geqslant 1-\varepsilon}} \min_{\sigma\in\mathcal{S}} [-\log_2 \mathrm{Tr}(A\sigma)]. \tag{71}$$

Then we get the following upper bound for the overhead.

Proposition 3. For an arbitrary state ρ, the overhead for virtual resource distillation with separability-preserving operations is upper bounded as

$$C^{\varepsilon}(\rho,m) \leqslant 2^{m-E_H^{\varepsilon}(\rho)+1} - 1 \tag{72}$$

for $m \geqslant E_H^{\varepsilon}(\rho)$.

Proof. Following a similar argument in Theorem 1, we explicitly construct the distillation protocol saturating the upper bound. We consider two separability-preserving channels as

$$\begin{aligned}\Lambda_1(\rho) &= \frac{1}{2^{m-E_H^{\varepsilon}(\rho)}}\mathrm{Tr}(A\rho)\Phi^{\otimes m} \\ &\quad + \left(1 - \frac{1}{2^{m-E_H^{\varepsilon}(\rho)}}\mathrm{Tr}(A\rho)\right)\frac{I-\Phi^{\otimes m}}{2^{2k}-1}, \\ \Lambda_2(\rho) &= \frac{I-\Phi^{\otimes m}}{2^{2k}-1},\end{aligned} \tag{73}$$

where A is the operator that achieves the optimal solution of Eq. (71). For any separable state $\sigma \in \mathcal{S}$, we have

$$\frac{1}{2^{m-E_H^{\varepsilon}(\rho)}}\mathrm{Tr}(A\sigma) \leqslant \frac{1}{2^m}, \tag{74}$$

indicating that the output state $\Lambda_1(\sigma)$ is also separable and hence Λ_1 is separability preserving. Since the output state of Λ_2 is separable, Λ_2 is also separability preserving. Letting $\lambda_1 = 2^{m-E_H^{\varepsilon}(\rho)}$ and $\lambda_2 = \lambda_1 - 1$, we have

$$\begin{aligned}\Phi' &= \lambda_1\Lambda_1(\rho) - \lambda_2\Lambda_2(\rho) \\ &= \mathrm{Tr}(A\rho)\Phi^{\otimes m} + [1-\mathrm{Tr}(A\rho)]\frac{I-\Phi^{\otimes m}}{2^{2m}-1}.\end{aligned} \tag{75}$$

Since $\mathrm{Tr}(A\rho) \geqslant 1-\varepsilon$, we have

$$\tfrac{1}{2}\|\Phi' - \Phi^{\otimes m}\| \leqslant \varepsilon. \tag{76}$$

Therefore, this constructs a virtual resource distillation protocol with an overhead of

$$\lambda_1 + \lambda_2 = 2^{m-E_H^{\varepsilon}(\rho)+1} - 1. \tag{77}$$

■

We thus obtain a lower bound to the distillation rate as

$$V^{\varepsilon}(\rho) \geqslant \max_m \frac{m}{\left(2^{m-E_H^{\varepsilon}(\rho)+1}-1\right)^2}. \tag{78}$$

We can also evaluate the right-hand side as follows. Letting $c := 2^{-E_H^{\varepsilon}(\rho)+1}$ and $g(m) := m/(2^m c - 1)^2$, we then get

$$\frac{d}{dm}g(m) = \frac{c2^m(1-2m\ln 2)-1}{(2^m c-1)^3}. \tag{79}$$

The fact that $0 \leqslant E_H^{\varepsilon}(\rho) \leqslant 1$ for every ρ gives $1 \leqslant c \leqslant 2$. Since $c2^m(1-2m\ln 2) - 1 < 0$ for $m \geqslant 1$ and $c \geqslant 0$, we get $\frac{d}{dm}g(m) < 0$ for all $m \geqslant 1$. Therefore, the maximum happens

when $m = 1$, which gives the tightest form of (78) as

$$V^{\varepsilon}(\rho) \geqslant \frac{1}{\left(2^{2-E_H^{\varepsilon}(\rho)} - 1\right)^2}. \tag{80}$$

2. *Exact expressions via semidefinite programming*

Here we focus on the zero-error distillation, i.e., $\varepsilon = 0$, and provide the exact expressions for the virtual resource distillation overhead.

Proposition 4. Let O be the class of separability-preserving operations. Then, for any state ρ, the distillation overhead is given by the convex optimization program

$$C^0(\rho, m) = \min\left\{\mu_+ + \mu_- \,\middle|\, 0 \leqslant Q_+ \leqslant \mu_+ I,\ 0 \leqslant Q_- \leqslant \mu_- I,\ \mathrm{Tr} Q_+ \sigma \leqslant \frac{\mu_+}{2^m},\ \mathrm{Tr} Q_- \sigma \leqslant \frac{\mu_-}{2^m}\ \forall\, \sigma \in \mathcal{S},\ \mathrm{Tr}\rho(Q_+ - Q_-) = 1 = \mu_+ - \mu_-\right\}. \tag{81}$$

Proof. The proof is direct application of Theorem 1. ∎

By focusing on PPT operations, the above optimization can take a simpler form and become an SDP.

Proposition 5. Let O be the class of positive partial transpose operations. Then, for any state ρ, the virtual distillation overhead is given by the semidefinite program

$$C^0(\rho, m) = \min\left\{\mu_+ + \mu_- \,\middle|\, 0 \leqslant Q_+ \leqslant \mu_+ I,\ 0 \leqslant Q_- \leqslant \mu_- I,\ \|Q_+^{\Gamma}\|_\infty \leqslant \frac{\mu_+}{2^m},\ \|Q_-^{\Gamma}\|_\infty \leqslant \frac{\mu_-}{2^m},\ \mathrm{Tr}\rho(Q_+ - Q_-) = 1 = \mu_+ - \mu_-\right\}. \tag{82}$$

Proof. For any PPT channels $\Lambda_\pm$ such that $\lambda_+\Lambda_+(\rho) - \lambda_-\Lambda_-(\rho) = \Phi^{\otimes m}$, we define $Q_\pm = \lambda_\pm \Lambda_\pm^\dagger(\Phi^{\otimes m})$ and $\mu_\pm = \lambda_\pm$, where $\Lambda^\dagger$ is the dual (adjoint) map. Then

$$\begin{aligned}
\|Q_+^{\Gamma}\|_\infty &= \max_{\omega\in\mathcal{D}} |\mathrm{Tr} Q_+^{\Gamma}\omega| \\
&= \lambda_\pm \max_{\omega\in\mathcal{D}} |\mathrm{Tr}\Phi^{\otimes m}\Lambda_\pm(\omega^{\Gamma})| \\
&= \lambda_\pm \max_{\omega\in\mathcal{D}} |\mathrm{Tr}(\Phi^{\otimes m})^{\Gamma}\Lambda_\pm(\omega^{\Gamma})^{\Gamma}| \\
&\leqslant \lambda_\pm \max_{\omega'\in\mathcal{D}} |\mathrm{Tr}(\Phi^{\otimes m})^{\Gamma}\omega'| \\
&= \frac{\lambda_\pm}{2^m},
\end{aligned} \tag{83}$$

where in the first line we used $\mathcal{D}$ to denote all density matrices, in the fourth line we used that $\Lambda_\pm(\omega^{\Gamma})^{\Gamma}$ must be a valid state because $\Lambda_\pm$ is a PPT operation, and in the fourth line we used the fact that the partial transpose of the maximally entangled state Φ is the SWAP operator with eigenvalues $\pm\frac{1}{2}$. Since $\Lambda_\pm$ are CPTP maps, the other conditions on $Q_\pm$ are also satisfied, meaning that $Q_\pm$ constitute feasible solutions to (82). Optimizing over all feasible PPT protocols yields $C^0(\rho, m) \geqslant \mu_+ + \mu_-$ for optimal $\mu_\pm$.

For the other direction, take any feasible $Q_\pm$ and $\mu_\pm$. Define

$$\Lambda_\pm(\omega) = \mathrm{Tr}\left(\frac{Q_\pm}{\mu_+}\omega\right)\Phi^{\otimes m} + \mathrm{Tr}\left[\left(I - \frac{Q_\pm}{\mu_+}\right)\omega\right]\frac{I - \Phi^{\otimes m}}{2^{2m} - 1}. \tag{84}$$

Writing $(\Phi^{\otimes m})^{\Gamma} = \frac{1}{2^m}(\Pi_S - \Pi_A)$, where Π_S and Π_A denote the projectors onto the symmetric and antisymmetric spaces, respectively, we have that

$$\begin{aligned}
J_{\Lambda_\pm}^{\Gamma} &= \frac{Q_\pm^{\Gamma}}{\mu_\pm} \otimes \frac{\Pi_S - \Pi_A}{2^m} + \frac{I - \frac{Q_-}{\mu_+}}{2^{2m} - 1} \otimes \frac{(2^m - 1)\Pi_S + (2^m + 1)\Pi_A}{2^m} \\
&= \left[\frac{Q_\pm^{\Gamma}}{\mu_+} + \frac{1}{2^m + 1}\left(I - \frac{Q_\pm^{\Gamma}}{\mu_+}\right)\right] \otimes \frac{\Pi_S}{2^m} + \left[-\frac{Q_\pm^{\Gamma}}{\mu_+} + \frac{1}{2^m - 1}\left(I - \frac{Q_\pm^{\Gamma}}{\mu_+}\right)\right] \otimes \frac{\Pi_A}{2^m} \\
&\geqslant \left[-\frac{I}{2^m} + \frac{1}{2^m + 1}\left(I - \frac{I}{2^m}\right)\right] \otimes \frac{\Pi_S}{2^m} + \left[-\frac{I}{2^m} + \frac{1}{2^m - 1}\left(I - \frac{I}{2^m}\right)\right] \otimes \frac{\Pi_A}{2^m} \\
&= 0,
\end{aligned} \tag{85}$$

where in the third line we used that $-\frac{\mu_\pm}{2^m} \leqslant Q_\pm^\Gamma \leqslant \frac{\mu_\pm}{2^m}$ and Π_S and Π_A are orthogonal to each other. The operations $\Lambda_\pm$ are therefore PPT, and since

$$\mu_+\Lambda_+(\rho) - \mu_-\Lambda_-(\rho) = \Phi^{\otimes m}, \tag{86}$$

we obtain $C^0(\rho, m) \leqslant \mu_+ + \mu_-$ as desired. ■

3. Exact expressions via the singlet fraction

We can obtain an alternative exact expression with the singlet fraction. The singlet fraction of a quantum state ρ is the maximum overlap with the maximally entangled state realized by applying an arbitrary LOCC operation to ρ. Here we slightly generalize this concept and call

$$f_O(\rho, m) \coloneqq \max_{\Lambda\in O} \mathrm{Tr}[\Lambda(\rho)\Phi^{\otimes m}] \tag{87}$$

the singlet fraction with respect to a set of free operations O. Then we get the following characterization.

Proposition 6. Let O be an arbitrary subset of separability-preserving operations that contains local unitary operations assisted by shared classical randomness, e.g., (one-way) LOCC. Then, for every state ρ,

$$C^\varepsilon(\rho, m) = \max\left\{\frac{2(1-\varepsilon)}{f_O(\rho, m)} - 1, 1\right\}. \tag{88}$$

Proof. This is a direct application of Theorem 2, as the local twirling $\mathcal{T}(\cdot) \coloneqq \int dU\, U \otimes U^* \cdot U^\dagger \otimes U^{*\dagger}$ serves as a free generalized twirling operation. ■

Proposition 6 implies that the performance of virtual resource distillation does not change over different choices of free operations for pure states.

Corollary 3. For any bipartite pure state ϕ, the overhead $C^\varepsilon(\phi, m)$ is the same for any choice of operations ranging from one-way LOCC to separability- or PPT-preserving operations. It admits the analytical expression

$$C^\varepsilon(\phi, m) = \max\left\{\frac{2^{m+1}(1-\varepsilon)}{\| |\phi\rangle \|^2_{[2^m]}} - 1,\ 1\right\}, \tag{89}$$

where

$$\| |\phi\rangle \|_{[m]} \coloneqq \left\|\zeta^\downarrow_{1:m-k^\star}\right\|_{\ell_1} + \sqrt{k^\star}\left\|\zeta^\downarrow_{m-k^\star+1:d}\right\|_{\ell_2} \tag{90}$$

defines the so-called m-distillation norm [67]. Here d is the local dimension, $\zeta^\downarrow_{1:k}$ stands for the vector consisting of the k largest (by magnitude) Schmidt coefficients of $|\phi\rangle$, analogously $|\zeta^\downarrow_{k+1:d}\rangle$ denotes the vector of the $d-k$ smallest Schmidt coefficients of $|\phi\rangle$ with $\zeta^\downarrow_{1:0}$ the zero vector, and

$$k^\star \coloneqq \arg\min_{1\leqslant k\leqslant m} \frac{1}{k}\left\|\zeta^\downarrow_{m-k+1:d}\right\|^2_{\ell_2}. \tag{91}$$

Proof. The proof follows since the singlet fraction of any pure state is the same under these operations and is given by $f(\phi, m) = 2^{-m}\| |\phi\rangle \|^2_{[2^m]}$ (see Theorem 15 in [67]). ■

Proposition 6 can also be employed to provide alternative exact expressions for the virtual resource distillation overhead in terms of entanglement measures previously studied.

Proposition 7. Consider separability-preserving operations as the set of free operations. Then, for every state ρ,

$$C^\varepsilon(\rho, m) = \max\left\{\frac{2(1-\varepsilon)}{G_S(\rho; 2^m)} - 1, 1\right\}, \tag{92}$$

where

$$G_S(\rho; k) \coloneqq \sup\left\{\mathrm{Tr}(\rho W)\,\middle|\, 0 \leqslant W \leqslant I,\ \mathrm{Tr}(W\sigma) \leqslant \frac{1}{k}\,\forall\,\sigma \in \mathcal{S}\right\}. \tag{93}$$

When ρ and $\Phi^{\otimes m}$ are defined on the same space (i.e., they have the same dimension),

$$C^\varepsilon(\rho, m) = \max\left\{\frac{2^{m+1}(1-\varepsilon)}{1 + R^g_S(\rho)} - 1, 1\right\}, \tag{94}$$

where $R^g_S(\rho)$ is the generalized robustness of entanglement.

The same result holds for PPT-preserving operations by replacing $\mathcal{S}$ with $\mathcal{S}_{\mathrm{PPT}}$.

Proof. It is known that, when O is the set of separability-preserving operations, the maximum overlap $f_O(\rho, m) = \max_{\Lambda\in O}\mathrm{Tr}[\Lambda(\rho)\Phi^{\otimes m}]$ can be characterized by $f_O(\rho, m) = G_S(\rho; 2^m)$ for a general state ρ and $f_O(\rho, m) = [1 + R^g_S(\rho)]\,2^{-m}$ if ρ and $\Phi^{\otimes m}$ act on the same space [50]. The result then follows by applying Proposition 6. ■

4. Isotropic states

The exact characterization of virtual distillation overhead with respect to the singlet fraction allows us to derive analytical expressions for the class of isotropic states. Let ρ^I_α be an isotropic state defined as

$$\rho^I_\alpha(k) \coloneqq (1-\alpha)\,\Phi^{\otimes k} + \alpha\,\frac{I - \Phi^{\otimes k}}{2^{2k} - 1}. \tag{95}$$

Proposition 8. Let O be an arbitrary subset of separability-preserving or PPT-preserving operations that includes all separable operations. Then, for any $1 \leqslant m \leqslant k$,

$$C^\varepsilon\left(\rho^I_\alpha(k), m\right) = \begin{cases} \max\{2^{m+1}(1-\varepsilon) - 1, 1\}, & \alpha \geqslant 1 - 2^{-k} \\ \max\left\{\frac{2(1-\varepsilon)}{1-\alpha c} - 1, 1\right\}, & \alpha \leqslant 1 - 2^{-k}, \end{cases} \tag{96}$$

where $c \coloneqq (2^k - 2^{k-m})/(2^k - 1)$. In particular,

$$C^\varepsilon\left(\rho^I_\alpha(k), k\right) = \begin{cases} \max\{2^{k+1}(1-\varepsilon) - 1, 1\}, & \alpha \geqslant 1 - 2^{-k} \\ \max\left\{\frac{1+\alpha-2\varepsilon}{1-\alpha}, 1\right\}, & \alpha \leqslant 1 - 2^{-k}, \end{cases} \tag{97}$$

and this latter result holds also for any set of free operations that contains local unitary operations assisted by shared classical randomness, in particular for LOCC.

Proof. By Proposition 6 we have that

$$C^\varepsilon\left(\rho^I_\alpha(k), m\right) = \max\left\{\frac{2(1-\varepsilon)}{f_O\left(\rho^I_\alpha(k), m\right)} - 1, 1\right\}. \tag{98}$$

For $\alpha \geqslant 1 - 2^{-k}$, the state $\rho^I_\alpha(k)$ is separable [68]. Then $\Lambda(\rho^I_\alpha(k))$ is separable (PPT) for any separability-preserving (PPT-preserving) map Λ; using the fact that the overlap of

$\psi^{\otimes m}$ with any PPT state is at most 2^{-m} and it is achieved by a separable state, it follows that $f_O(\rho_\alpha^I(k), m) = 2^{-m}$. For $\alpha < 1 - 2^{-k}$, we use the result of Theorem 18 in [67], which states that

$$f_O\big(\rho_\alpha^I(k), m\big) = 1 - \alpha\frac{2^k - 2^{k-m}}{2^k - 1}. \tag{99}$$

Plugging these values into Eq. (98) concludes the first part of the proof.

For the second part, we notice that $\mathrm{Tr}[\rho_\alpha^I(k)\psi^{\otimes k}] = 1 - \alpha$. Since the value of this overlap cannot be increased by any free operation (see Corollary 15 in [40]), we get $f_O(\rho_\alpha^I(k), k) = 1 - \alpha$ as desired. ■

5. *Bound entanglement does not help virtual distillation*

The virtual distillation overhead is governed by the size of the coefficients in a linear combination of accessible states that form a decomposition of a target state. Intuitively, a smaller overhead could be realized if one were given a larger set of accessible states. In the context of entanglement distillation, if we are given some entangled state, the set of accessible states obtained by applying LOCC operations to the given entangled state is strictly greater than the set of separable states. This leads to a natural question: Is every entangled state useful for virtual distillation? The following result answers this question in the negative. A similar restriction holds also beyond LOCC operations, applying to all PPT-preserving maps.

Proposition 9. Consider LOCC as the set of free operations. Then, for every bound-entangled state ρ and every $\varepsilon \in [0, 1)$,

$$C^\varepsilon(\rho, m) = C_{\mathcal{S}}^\varepsilon(m) = \max\{2^{m+1}(1-\varepsilon) - 1, 1\}, \tag{100}$$

where $C_{\mathcal{S}}^\varepsilon(m)$ is the virtual distillation overhead for separable states, which takes the same value for all separable states. If ρ is a PPT, then (100) holds for an arbitrary subset of PPT-preserving operations that can prepare all separable states.

Proof. Recall that the LOCC singlet fraction for states with zero distillable entanglement satisfies $f_O(\rho, m) \leqslant 2^{-m}$ [68]. Noting that LOCC can prepare every separable state shows $f_O(\rho, m) = 2^{-m}$. Taking $O = \mathrm{LOCC}$ in Proposition 6 then proves the first statement. To show the latter statement, note that for every PPT-preserving map Λ and every PPT state ρ,

$$\begin{aligned}\mathrm{Tr}[\Lambda(\rho)\Phi^{\otimes m}] &\leqslant \max_{\sigma\in\mathrm{PPT}} \mathrm{Tr}(\sigma\Phi^{\otimes m}) \\ &= \max_{\sigma\in\mathrm{PPT}} \mathrm{Tr}[\sigma^\Gamma(\Phi^{\otimes m})^\Gamma] \\ &\leqslant \max_{\omega\in\mathcal{D}} \mathrm{Tr}[\omega\,(\Phi^\Gamma)^{\otimes m}] \\ &= 2^{-m},\end{aligned} \tag{101}$$

where we used that the eigenvalues of Φ^Γ are $\pm\frac{1}{2}$. Noting that the same overlap can be achieved by optimizing over separable states [69] and invoking Proposition 6, the result follows. ■

B. Coherence

We next consider the resource theory of coherence (superposition) [2], where the set $\mathcal{I}$ of free states consists of diagonal states with respect to a given preferred basis $\{|i\rangle\}_i$, i.e., $\mathcal{I} := \{\sum_i p_i|i\rangle\langle i| \mid \sum_i p_i = 1,\ p_i \geqslant 0\,\forall\, i\}$.

We first show that the virtual resource distillation overhead admits an analytical expression for an arbitrary single-qubit state.

Proposition 10. For an arbitrary qubit state ρ and an arbitrary set O of free operations that contains probabilistic applications of Pauli X and Z,

$$C^\varepsilon(\rho, 1) = \max\left\{\frac{1-2\varepsilon}{M_{l_1}(\rho)}, 1\right\}, \tag{102}$$

where $M_{l_1}(\rho) = \sum_{i\neq j} |\langle i|\rho|j\rangle|$ is the l_1-norm of coherence.

Proof. Let $\rho = \left(\begin{smallmatrix}\alpha & \beta \\ \beta & 1-\alpha\end{smallmatrix}\right)$. We take $\beta \geqslant 0$ because any state can be brought to this form by the Pauli Z operation, and $C^\varepsilon(\rho, 1)$ is invariant under such an operation. To see $C^\varepsilon(\rho, 1) \leqslant \max\{\frac{1-2\varepsilon}{M_{l_1}(\rho)}, 1\}$, let $\mathcal{T}(\cdot) := \frac{1}{2}\cdot + \frac{1}{2}X\cdot X$ and $\mathcal{Z}\circ\mathcal{T}(\cdot) := \frac{1}{2}Z\cdot Z + \frac{1}{2}ZX\cdot XZ$. Also, let $s(\varepsilon) = \frac{1-2\varepsilon}{4\beta} + \frac{1}{2}$. Then consider the unit trace operator η defined as

$$\eta := s(\varepsilon)\mathcal{T}(\rho) - [s(\varepsilon) - 1]\mathcal{Z}\circ\mathcal{T}(\rho). \tag{103}$$

A direct computation reveals that $\eta = (1-\varepsilon)|+\rangle\langle+| + \varepsilon|-\rangle\langle-|$ and $\frac{1}{2}\|\eta - |+\rangle\langle+|\|_1 \leqslant \varepsilon$. When $\frac{1-2\varepsilon}{M_{l_1}(\rho)} \leqslant 1$, $s(\varepsilon) - 1 \leqslant 0$ and η is a convex combination of $\mathcal{T}(\rho)$ and $\mathcal{Z}\circ\mathcal{T}(\rho)$, giving $C^\varepsilon(\rho, 1) \leqslant 1$. When $\frac{1-2\varepsilon}{M_{l_1}(\rho)} \geqslant 1$ we obtain

$$C^\varepsilon(\rho, 1) \leqslant 2s(\varepsilon) - 1 = \frac{1-2\varepsilon}{2\beta} = \frac{1-2\varepsilon}{M_{l_1}(\rho)}. \tag{104}$$

On the other hand, $C^\varepsilon(\rho, 1) \geqslant \max\{\frac{1-2\varepsilon}{M_{l_1}(\rho)}, 1\}$ can be obtained from Theorem 4 below. ■

We now characterize the virtual resource distillation overhead for general states via semidefinite programming. Recall that a channel $\mathcal{E}$ is called a maximally incoherent operation (MIO) if it maps every incoherent state to an incoherent state, i.e., $\mathcal{E}(\sigma) \in \mathcal{I}\,\forall\,\sigma \in \mathcal{I}$, and $\mathcal{E}$ is called a dephasing-covariant operation (DIO) if it commutes with the completely dephasing map $\Delta(\cdot) = \sum_i |i\rangle\langle i|\cdot|i\rangle\langle i|$, i.e., $\mathcal{E}\circ\Delta = \Delta\circ\mathcal{E}$.

Theorem 3. Let O be the class of maximally incoherent operations or dephasing-covariant incoherent operations and let $\Delta(\cdot) = \sum_i |i\rangle\langle i|\cdot|i\rangle\langle i|$ be the diagonal map. Then, for any state ρ, the virtual distillation overhead is given by the semidefinite program

$$\begin{aligned}C^\varepsilon(\rho, m) = \min\Big\{\mu_+ + \mu_- \Big| & 0 \leqslant Q_+ \leqslant \mu_+ I,\ 0 \leqslant Q_- \leqslant \mu_- I, \\ & \Delta(Q_+) = \frac{\mu_+}{2^m} I,\ \Delta(Q_-) = \frac{\mu_-}{2^m} I, \mu_+ - \mu_- = 1, \mathrm{Tr}\rho(Q_+ - Q_-) \geqslant 1 - \varepsilon\Big\}.\end{aligned} \tag{105}$$

Proof. The lower bound on $C^\varepsilon(\rho, m)$ is essentially an application of Theorem 1, but let us consider it explicitly for completeness. Consider then any feasible MIO protocol such that $\Lambda_\pm \in \mathcal{O}$ and $\frac{1}{2}\|\lambda_+\Lambda_+(\rho) - \lambda_-\Lambda_-(\rho) - |+\rangle\langle+|^{\otimes m}\|_1 \leqslant \varepsilon$. Define $Q_\pm = \lambda_\pm \Lambda_\pm^\dagger(|+\rangle\langle+|^{\otimes m})$ and $\mu_\pm = \lambda_\pm$. For each i, we then have that

$$\begin{aligned}\langle i|Q_\pm|i\rangle &= \lambda_\pm \mathrm{Tr}[\Lambda_\pm(|i\rangle\langle i|)\,|+\rangle\langle+|^{\otimes m}] \\ &= \lambda_\pm \frac{1}{2^m},\end{aligned} \tag{106}$$

where the last line follows since $|+\rangle\langle+|^{\otimes m}$ has a constant overlap 2^{-m} with any incoherent state. This means in particular that $\Delta(Q_\pm) = \frac{\mu_\pm}{2^m} I$. Verifying that other conditions are also satisfied due to the fact that each $\Lambda_\pm$ is a CPTP map, we have that $Q_\pm$ and $\mu_\pm$ are feasible solutions to (105), leading to $C^\varepsilon(\rho, m) \geqslant \mu_+ - \mu_-$ for all MIO maps.

Conversely, let $Q_\pm$ be feasible solutions to (105). Define the quantum channels

$$\Lambda_\pm(\omega) = \mathrm{Tr}\left(\frac{Q_\pm}{\mu_\pm}\omega\right)|+\rangle\langle+|^{\otimes m} + \mathrm{Tr}\left[\left(I - \frac{Q_\pm}{\mu_\pm}\right)\omega\right]\frac{I - |+\rangle\langle+|^{\otimes m}}{2^m - 1}. \tag{107}$$

For any state ω, we have that

$$\begin{aligned}\Lambda_\pm \circ \Delta(\omega) &= \mathrm{Tr}\left[\Delta\left(\frac{Q_\pm}{\mu_\pm}\right)\omega\right]|+\rangle\langle+|^{\otimes m} + \mathrm{Tr}\left\{\left[I - \Delta\left(\frac{Q_\pm}{\mu_\pm}\right)\right]\omega\right\}\frac{I - |+\rangle\langle+|^{\otimes m}}{2^m - 1} \\ &= \frac{1}{2^m}|+\rangle\langle+|^{\otimes m} + \frac{1}{2^m}(I - |+\rangle\langle+|^{\otimes m}) \\ &= \frac{1}{2^m} I \\ &= \Delta \circ \Lambda_\pm(\omega),\end{aligned} \tag{108}$$

and so the constructed maps are both DIOs. As the maps realize the virtual distillation of $|+\rangle\langle+|^{\otimes m}$ from ρ up to error ε, we have that $C^\varepsilon(\rho, m) \leqslant \mu_+ + \mu_-$ under DIO. Since DIO $\subseteq$ MIO, the cost under MIO lower bounds the cost under DIO, and the result follows. ■

For the case of a single qubit, the above program can be analytically solved for every number m of target states.

Theorem 4. Let $\mathcal{O}$ be an MIO or DIO. Then, for every single-qubit state ρ, every integer $m \geqslant 1$, and all $\varepsilon \in [0, 1]$, the virtual distillation overhead is given by

$$C^\varepsilon(\rho, m) = \max\left\{\frac{2^m(1-\varepsilon) - 1}{M_{l_1}(\rho)}, 1\right\}, \tag{109}$$

where $M_{l_1}(\rho)$ is the l_1-norm of coherence.

Proof. Without loss of generality, we assume that ρ is on the XZ plane in the Bloch coordinate, i.e., $\mathrm{Tr}(\rho Y) = 0$, and that $\langle+|\rho|+\rangle \geqslant \langle-|\rho|-\rangle$, as one can always bring any state onto the XZ plane with $\langle+|\rho|+\rangle \geqslant \langle-|\rho|-\rangle$ by applying an incoherent unitary, and $C^\varepsilon(\rho, m)$ is invariant under any incoherent unitary.

To constrain the form of $Q_\pm$, we consider a map Λ_{XZ} defined by

$$\Lambda_{XZ}(\cdot) := \tfrac{1}{2}\mathrm{Tr}(\cdot)I + \tfrac{1}{2}\mathrm{Tr}(X\cdot)X + \tfrac{1}{2}\mathrm{Tr}(Z\cdot)Z. \tag{110}$$

When applied to a quantum state, Λ_{XZ} projects it to the XZ plane. It is straightforward to check that it is unital, i.e., $\Lambda_{XZ}(I) = I$, and self-dual, i.e., $\Lambda_{XZ}^\dagger = \Lambda_{XZ}$. The map Λ_{XZ} is also positive. This is because for an arbitrary single-qubit state σ, $\Lambda_{XZ}(\sigma)$ is also a valid state and hence positive. Since every positive operator acting on the single-qubit system is proportional to a quantum state, their positivity remains under Λ_{XZ}.

By assumption, we have $\Lambda_{XZ}(\rho) = \rho$. Let $Q_\pm^\star$ and $\mu_\pm^\star$ be the operators and real numbers that give the optimal solution of (105). Then one can see that $\Lambda_{XZ}(Q_\pm^\star)$ also give the optimal solution $\mu_+^\star + \mu_-^\star$ as follows. First, $0 \leqslant \Lambda_{XZ}(Q_\pm^\star) \leqslant \mu_+^\star I$ follows from Λ_{XZ} being positive and unital. Then $\Delta \circ \Lambda_{XZ}(Q_\pm^\star) = \frac{\mu_\pm^\star}{2^m} I$ follows from the fact that $\Delta \circ \Lambda_{XZ} = \Lambda_{XZ} \circ \Delta$ and $\Lambda_{XZ}(I) = I$. Finally,

$$\begin{aligned}\mathrm{Tr}\rho[\Lambda_{XZ}(Q_+^\star) - \Lambda_{XZ}(Q_-^\star)] &= \mathrm{Tr}\Lambda_{XZ}^\dagger(\rho)(Q_+^\star - Q_-^\star) \\ &= \mathrm{Tr}\Lambda_{XZ}(\rho)(Q_+^\star - Q_-^\star) \\ &= \mathrm{Tr}\rho(Q_+^\star - Q_-^\star) \\ &\geqslant 1 - \varepsilon,\end{aligned} \tag{111}$$

where in the second equality we used that Λ_{XZ} is self-dual and in the third equality we used $\Lambda_{XZ}(\rho) = \rho$ by assumption. Thus, it suffices to restrict our attention to operators $Q_\pm$ such that $Q_\pm = c_\pm^I I + c_\pm^X X + c_\pm^Z Z$, where $c_\pm^P$ are some real numbers. In addition, the condition $\Delta(Q_\pm) \propto I$ further imposes $c_\pm^Z = 0$. This allows us to write $Q_\pm$ in the form

$$Q_\pm = c_\pm^+|+\rangle\langle+| + c_\pm^-|-\rangle\langle-|, \quad c_\pm^+, c_\pm^- \in \mathbb{R}. \tag{112}$$

In terms of this expression, (105) can be rewritten as

$$\begin{aligned}C^\varepsilon(\rho, m) = \min\Bigg\{\mu_+ + \mu_- \,\Bigg|\, &0 \leqslant c_\pm^+, c_\pm^- \leqslant \mu_\pm, c_\pm^+ + c_\pm^- = \frac{\mu_\pm}{2^{m-1}}, \mu_+ - \mu_- = 1, \\ &(c_+^+ - c_-^+)\langle+|\rho|+\rangle + (c_+^- - c_-^-)\langle-|\rho|-\rangle \geqslant 1 - \varepsilon\Bigg\}.\end{aligned} \tag{113}$$

Since $c_\pm^+ + c_\pm^- = \frac{\mu_\pm}{2^{m-1}}$ ensures $c_\pm^+, c_\pm^- \leqslant \mu_\pm$, we can further simplify it to

$$C^\varepsilon(\rho, m) = \min\{2^{m-1}(c_+^+ + c_+^- + c_-^+ + c_-^-) \mid c_\pm^+, c_\pm^- \geqslant 0, c_+^+ - c_-^+ + c_+^- - c_-^- = 1/2^{m-1}, \\ (c_+^+ - c_-^+)\langle +|\rho|+\rangle + (c_+^- - c_-^-)\langle -|\rho|-\rangle \geqslant 1-\varepsilon\}. \tag{114}$$

Since the second and third constraints only involve $c_+^+ - c_-^+$ and $c_+^- - c_-^-$, the minimum occurs when $c_+^+ c_-^+ = c_+^- c_-^- = 0$, because if $c_+^+, c_-^+ \neq 0$ or $c_+^-, c_-^- \neq 0$, one can always make the objective function smaller while keeping the values of $c_+^+ - c_-^+$ and $c_+^- - c_-^-$. Therefore, letting $\alpha \coloneqq c_+^+ - c_-^+$ and $\beta \coloneqq c_+^- - c_-^-$, we get

$$\begin{aligned} C^\varepsilon(\rho, m) &= \min\{2^{m-1}(|\alpha| + |\beta|) \mid \alpha, \beta \in \mathbb{R}, \alpha + \beta = 1/2^{m-1}, \ \alpha\langle +|\rho|+\rangle + \beta\langle -|\rho|-\rangle \geqslant 1-\varepsilon\} \\ &= \min\{2^{m-1}(|\alpha| + |\beta|) \mid \alpha, \beta \in \mathbb{R}, \\ &\qquad \alpha + \beta = 1/2^{m-1}, \ [1 + M_{l_1}(\rho)]/2^m - \beta M_{l_1}(\rho) \geqslant 1-\varepsilon\}, \end{aligned} \tag{115}$$

where in the second equality we rewrote the left-hand side of the third constraint as $(\alpha+\beta)\langle +|\rho|+\rangle + \beta(-\langle +|\rho|+\rangle + \langle -|\rho|-\rangle)$ and used the second constraint as well as the definition of the l_1-norm of coherence $M_{l_1}(\rho) = \langle +|\rho|+\rangle - \langle -|\rho|-\rangle$ and the normalization $\langle +|\rho|+\rangle + \langle -|\rho|-\rangle = 1$.

Suppose that $\alpha < 0$. Then the second constraint enforces $\beta = 1/2^{m-1} - \alpha > 0$. However, this is ensured to be suboptimal because by decreasing $\beta > 0$ so that $\alpha \to 0$, the objective function can be monotonically reduced while the third constraint is ensured to be satisfied, as the left-hand side of the third constraint is monotonically decreasing with $\beta > 0$. Thus, we can set $\alpha \geqslant 0$ and write

$$C^\varepsilon(\rho, m) = \min\{1 + 2^{m-1}(-\beta + |\beta|) \mid \beta \leqslant 1/2^{m-1}, \ [1 + M_{l_1}(\rho)]/2^m - \beta M_{l_1}(\rho) \geqslant 1-\varepsilon\}. \tag{116}$$

The form of the objective function implies that the minimum occurs at the largest β that satisfies both constraints and particularly takes the value 1 if β can become non-negative.

When $1 - \varepsilon < [1 - M_{l_1}(\rho)]/2^m$ we have

$$[1 + M_{l_1}(\rho)]/2^m - \beta M_{l_1}(\rho) \geqslant [1 - M_{l_1}(\rho)]/2^m \tag{117}$$

due to the first constraint. Therefore, the second constraint is always satisfied in this case, and the minimization occurs for any non-negative β, which gives $C^\varepsilon(\rho, m) = 1$.

Suppose now that $1 - \varepsilon \geqslant [1 - M_{l_1}(\rho)]/2^m$. In this case, the minimum occurs when the second constraint becomes equality, which gives

$$\beta = \frac{1}{M_{l_1}(\rho)}\left(\frac{1 + M_{l_1}(\rho)}{2^m} - (1-\varepsilon)\right). \tag{118}$$

When $1 - \varepsilon \leqslant [1 + M_{l_1}(\rho)]/2^m$ we have $\beta \geqslant 0$, which makes $C^\varepsilon(\rho, m) = 1$. On the other hand, when $1 - \varepsilon \geqslant [1 + M_{l_1}(\rho)]/2^m$ we have $\beta \leqslant 0$ and

$$\begin{aligned} C^\varepsilon(\rho, m) &= 1 - \frac{2^m}{M_{l_1}(\rho)}\left(\frac{1 + M_{l_1}(\rho)}{2^m} - (1-\varepsilon)\right) \\ &= \frac{2^m(1-\varepsilon) - 1}{M_{l_1}(\rho)}, \end{aligned} \tag{119}$$

which is greater than or equal to 1.

These cases are summarized as

$$C^\varepsilon(\rho, m) = \max\left\{\frac{2^m(1-\varepsilon) - 1}{M_{l_1}(\rho)}, 1\right\}, \tag{120}$$

concluding the proof. ∎

The analytical expression for the distillation overhead in Theorem 4 allows for an exact characterization of the virtual resource distillation rate.

Corollary 4. Let O be an MIO or DIO. Then for every single-qubit state ρ and $\varepsilon \in [0, 1]$, the virtual distillation rate is given by

$$V^\varepsilon(\rho) = \max\left\{\frac{(\tilde{m}+1)M_{l_1}(\rho)^2}{[2^{\tilde{m}+1}(1-\varepsilon) - 1]^2}, \tilde{m}\right\}, \tag{121}$$

where $M_{l_1}(\rho)$ is the l_1-norm of coherence and

$$\tilde{m} \coloneqq \left\lfloor \log_2\left(\frac{M_{l_1}(\rho) + 1}{1-\varepsilon}\right)\right\rfloor. \tag{122}$$

Proof. We first remark that the function $m/C^\varepsilon(\rho, m)^2$ is monotonically decreasing with m, as we can write $m/C^\varepsilon(\rho, m)^2 = M_{l_1}(\rho)^2 g(m)$, where $g(m) = m/[2^m(1-\varepsilon) - 1]^2$ is the same function that appears in (79) with $c = 1 - \varepsilon$, which is shown to be monotonically decreasing for all $m \geqslant 1$ and $c \geqslant 0$. This implies that the optimal m for the maximization $\sup_m m/C^\varepsilon(\rho, m)^2$ is not greater than the smallest m such that

$$\frac{2^m(1-\varepsilon) - 1}{M_{l_1}(\rho)} \geqslant 1. \tag{123}$$

On the other hand, for m such that

$$\frac{2^m(1-\varepsilon) - 1}{M_{l_1}(\rho)} \leqslant 1, \tag{124}$$

we have $m/C^\varepsilon(\rho, m) = m$, which is an increasing function with m. Therefore, letting $\tilde{m}$ be the maximum integer m satisfying (124), the maximum for $\sup_m m/C^\varepsilon(\rho, m)^2$ happens at either $m = \tilde{m}$, which gives $C^\varepsilon(\rho, m) = 1$, or $m = \tilde{m} + 1$, which gives $C^\varepsilon(\rho, m) = [2^{\tilde{m}+1}(1-\varepsilon) - 1]/M_{l_1}(\rho)$. The result then follows by observing that $\tilde{m}$ can be explicitly obtained in the form of (122). ∎

C. Magic

Here we discuss the resource theory of magic states, which is motivated by the scenario of fault-tolerant quantum computation [70,71]. The stabilizer states are the states that can

be created by Clifford gates and classical randomness, and the resource theory of magic quantifies how much a given state deviates from the set of stabilizer states. Among them, the T state defined by

$$T := |T\rangle\langle T| = \frac{I + (X+Y)/\sqrt{2}}{2},$$
$$|T\rangle := \frac{1}{\sqrt{2}}(|0\rangle + e^{i\pi/4}|1\rangle) \tag{125}$$

plays a major role, as access to the T state together with stabilizer operations is sufficient to realize universal quantum computation. Therefore, magic state distillation protocols usually take the T state as the target state to synthesize. The important class of noisy input states for magic state distillation is the dephased T state,

$$\rho_p^T := (1-p)T + pI/2. \tag{126}$$

An arbitrary state can be brought into this form by the Clifford operation, which applies SX with $S = \mathrm{diag}(1, i)$ with probability $1/2$. Therefore, the design of a magic state distillation protocol can focus on these specific noisy states as its input.

We can apply our virtual resource distillation framework to this class of states, potentially providing better computational accuracy for algorithms run on fault-tolerant quantum computers that aim to obtain expectation values. We note that a closely related setting was discussed in terms of a combination of quantum error mitigation and error correction methods [72–74]. Our framework encompasses this strategy for magic state distillation as an application of the general approach of virtual resource distillation. In particular, the following analytical expressions for the virtual resource distillation extend the result in Ref. [72] to the regime with nonzero error. Here we let O_{STAB} and $\mathcal{F}_{\mathrm{STAB}}$ denote the sets of stabilizer operations and states, respectively.

Proposition 11. Let $p_{\mathrm{th}} := 1/\sqrt{2}$ be the maximum value such that $\rho_{p_{\mathrm{th}}}^T \in \mathcal{F}_{\mathrm{STAB}}$. Then the virtual resource distillation overhead with respect to the target state $|T\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$ under stabilizer operations is characterized by

$$C^\varepsilon\left(\rho_p^T, 1\right) = \begin{cases} \frac{1-2\varepsilon}{p_{\mathrm{th}}}, & |p| \leqslant p_{\mathrm{th}} \\ \frac{1-2\varepsilon}{|p|}, & |p| > p_{\mathrm{th}}. \end{cases} \tag{127}$$

Proof. Recall that

$$C^\varepsilon\left(\rho_p^T, 1\right) = \min\{\lambda_+ + \lambda_- \mid \eta = \lambda_+\Lambda_+\left(\rho_p^T\right) - \lambda_-\Lambda_-\left(\rho_p^T\right),$$
$$\Lambda_\pm \in O_{\mathrm{STAB}},\ \tfrac{1}{2}\|\eta - T\|_1 \leqslant \varepsilon\}. \tag{128}$$

Let $\mathcal{T}(\cdot) = I\cdot I + SX\cdot(SX)^\dagger$, where $S = \mathrm{diag}(1, i)$ is the phase gate and X is the Pauli X operator. Then defining $\overline{T} = |\overline{T}\rangle\langle\overline{T}|$ with $|\overline{T}\rangle := Z|T\rangle = \frac{1}{\sqrt{2}}(|0\rangle - e^{i\pi/4}|1\rangle)$, we get for any state ρ that

$$\mathcal{T}(\rho) = \mathrm{Tr}(T\rho)\,T + \mathrm{Tr}(\overline{T}\rho)\,\overline{T}. \tag{129}$$

Using $\mathcal{T}(T) = T$ and the monotonicity of the trace norm under quantum channels, we have

$$\tfrac{1}{2}\|\mathcal{T}(\eta) - T\|_1 = \tfrac{1}{2}\|\mathcal{T}(\eta) - \mathcal{T}(T)\|_1 \leqslant \tfrac{1}{2}\|\eta - T\|_1. \tag{130}$$

This ensures that if η is a feasible solution of (128), so is $\mathcal{T}(\eta)$. Thus, the optimal solution η for (128) can be restricted to the form $\eta = \lambda_+\mathcal{T}\circ\Lambda_+(\rho_p^T) - \lambda_-\mathcal{T}\circ\Lambda_-(\rho_p^T)$.

Note that $\rho_p^T = \frac{1+p}{2}T + \frac{1-p}{2}\overline{T}$. We assume $p \geqslant 0$ without a loss of generality because we can always apply Z to flip the sign. If $p \leqslant p_{\mathrm{th}}$, i.e., $\rho_p^T \in \mathcal{F}_{\mathrm{STAB}}$, $\mathcal{T}\circ\Lambda_\pm(\rho_p^T)$ is also a stabilizer state as $\mathcal{T}\circ\Lambda_\pm \in O_{\mathrm{STAB}}$. Therefore, the optimal solutions should take the form

$$\mathcal{T}\circ\Lambda_+\left(\rho_p^T\right) = \rho_{p_{\mathrm{th}}}^T = \frac{1+p_{\mathrm{th}}}{2}T + \frac{1-p_{\mathrm{th}}}{2}\overline{T},$$
$$\mathcal{T}\circ\Lambda_-\left(\rho_p^T\right) = \rho_{-p_{\mathrm{th}}}^T = \frac{1-p_{\mathrm{th}}}{2}T + \frac{1+p_{\mathrm{th}}}{2}\overline{T}. \tag{131}$$

This form specifies the optimal η as

$$\eta = \frac{1+(\lambda_+ + \lambda_-)p_{\mathrm{th}}}{2}T + \frac{1-(\lambda_+ + \lambda_-)p_{\mathrm{th}}}{2}\overline{T}. \tag{132}$$

Therefore, the condition $\frac{1}{2}\|\eta - T\|_1 \leqslant \varepsilon$ is equivalent to $\frac{1-(\lambda_+ + \lambda_-)p_{\mathrm{th}}}{2} \leqslant \varepsilon$. The optimal $\lambda_\pm$ under this condition gives $C^\varepsilon(\rho_p^T, 1) = \frac{1-2\varepsilon}{p_{\mathrm{th}}}$.

When $p > p_{\mathrm{th}}$, there always exists $\Lambda \in O_{\mathrm{STAB}}$ such that $\rho_{p'}^T = \mathcal{T}\circ\Lambda(\rho_p^T)$ for every $p' \in [-p, p]$; such a Λ is realized by either mixing the maximally mixed state with ρ_p^T or applying Z to ρ_p^T to make ρ_{-p}^T and mixing the maximally mixed state to it. On the other hand, no $p' \notin [-p, p]$ can be realized because otherwise the free operation $\mathcal{T}\circ\Lambda \in O_{\mathrm{STAB}}$ would increase a resource monotone [e.g., trace-distance measure $R_{\mathrm{tr}}(\rho) := \min_{\sigma\in\mathcal{F}_{\mathrm{STAB}}}\frac{1}{2}\|\rho - \sigma\|_1$].

Therefore, following a similar argument for the case of $p \leqslant p_{\mathrm{th}}$, the optimal solutions should take the form

$$\mathcal{T}\circ\Lambda_+\left(\rho_p^T\right) = \rho_p^T = \frac{1+p}{2}T + \frac{1-p}{2}\overline{T},$$
$$\mathcal{T}\circ\Lambda_-\left(\rho_p^T\right) = \rho_{-p}^T = \frac{1-p}{2}T + \frac{1+p}{2}\overline{T}. \tag{133}$$

This form specifies the optimal η as

$$\eta = \frac{1+(\lambda_+ + \lambda_-)p}{2}T + \frac{1-(\lambda_+ + \lambda_-)p}{2}\overline{T}. \tag{134}$$

Therefore, the condition $\frac{1}{2}\|\eta - T\|_1 \leqslant \varepsilon$ is equivalent to $\frac{1-(\lambda_+ + \lambda_-)p}{2} \leqslant \varepsilon$. The optimal $\lambda_\pm$ under this condition give $C^\varepsilon(\rho_p^T, 1) = \frac{1-2\varepsilon}{p}$. ∎

For qutrit states, one of the magic states that maximize the negativity of the discrete Wigner function [75] is known as the Strange state

$$S = |S\rangle\langle S|, \quad |S\rangle := \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle). \tag{135}$$

We can characterize overhead for the Strange state as follows.

Proposition 12. Let O_{STAB} be the set of stabilizer operations and consider S as the target state for virtual resource distillation. Then we have

$$C^\varepsilon(\rho, m) = \max\left\{\frac{2(1-\varepsilon)}{f_{O_{\mathrm{STAB}}}(\rho, m)} - 1, 1\right\}, \tag{136}$$

where $f_{O_{\mathrm{STAB}}}(\rho, m) := \max_{\Lambda\in O_{\mathrm{STAB}}}\mathrm{Tr}[\Lambda(\rho)S^{\otimes m}]$ is the maximum overlap with the copies of the S state.

Proof. The Strange state admits a free twirling operation of the form (64) [52,70] and hence the result follows as a consequence of Theorem 2. In particular, the twirling operation

for the S state is the random application of Clifford unitaries that correspond to elements in the special linear group $\mathrm{SL}(2, \mathbb{Z}_3)$. ■

D. Quantum memory and error mitigation

Next we consider an example of virtual distillation in channel theories. We consider a case that distills a d-dimensional noisy memory $\mathcal{M}$ into an ideal d-dimensional memory $\mathcal{I}_d$, that is, the identity channel, which perfectly preserves any quantum system, by adding extra gates after the application of the memory channel.

The distillation overhead with respect to the target channel $\mathcal{I}_d$ is

$$C^{\varepsilon}(\mathcal{M}, m) = \min_{\{\mathcal{N}_i\}_i, \{\lambda_i\}_i} \left\{ \sum_i |\lambda_i| \,\middle|\, \frac{1}{2} \left\| \sum_i \lambda_i \mathcal{N}_i \circ \mathcal{M} - \mathcal{I}_d^{\otimes m} \right\|_{\diamond} \leqslant \varepsilon \right\}. \tag{137}$$

We can imagine this as the case where we have a quantum gate $\mathcal{U}$ followed by a noise $\mathcal{M}$. Then we aim to apply extra gates so that the noise is canceled. The strategy we consider here thus contains several error mitigation methods [11,76], and similar performance analysis was also studied [9,10,33].

The diamond norm $\| \sum_i \lambda_i \mathcal{N}_i \circ \mathcal{M} - \mathcal{I}_d^{\otimes m} \|_{\diamond}$ can be written as a semidefinite program as [77]

$$\begin{aligned} &\left\| \sum_i \lambda_i \mathcal{N}_i \circ \mathcal{M} - \mathcal{I}_d^{\otimes m} \right\|_{\diamond} \\ &= \min \left\{ 2\lambda - \sum_i \lambda_i + 1 \,\middle|\, \lambda J_{\mathcal{E}} \geqslant \sum_i \lambda_i J_{\mathcal{N}_i \circ \mathcal{M}} - J_{\mathcal{I}_d^{\otimes m}}, \right. \\ &\qquad \left. \mathcal{E} \in \mathrm{CPTP} \right\}, \end{aligned} \tag{138}$$

where J_{Λ} denotes the Choi state for a channel Λ. Therefore, we can write the virtual distillation overhead as a semidefinite program as

$$\begin{aligned} C^{\varepsilon}(\mathcal{M}, m) = \min_{\mathcal{N}_{\pm}, \lambda_{\pm}} \left\{ \lambda_+ + \lambda_- \,\middle|\, 2\lambda - \sum_i \lambda_i + 1 \leqslant \varepsilon, \right. \\ \left. \lambda J_{\mathcal{E}} \geqslant \sum_i \lambda_i J_{\mathcal{N}_i \circ \mathcal{M}} - J_{\mathcal{I}_d^{\otimes m}},\ \mathcal{E} \in \mathrm{CPTP} \right\}. \end{aligned} \tag{139}$$

In Fig. 3 we compute the overhead for depolarizing channels, dephasing channels, and stochastic replacement channels for $m = 1$.

E. Quantum communication

A similar argument can be applied to the setting of quantum communication, in which Alice aims to send a quantum state to Bob via a noisy channel $\mathcal{E}$. In the usual setting of quantum communication, Alice and Bob apply additional quantum operations available to them so that Bob can recover the quantum state that was initially in Alice's hands. The rate at which noiseless qubits can be successfully sent is known as the quantum capacity of a noisy channel.

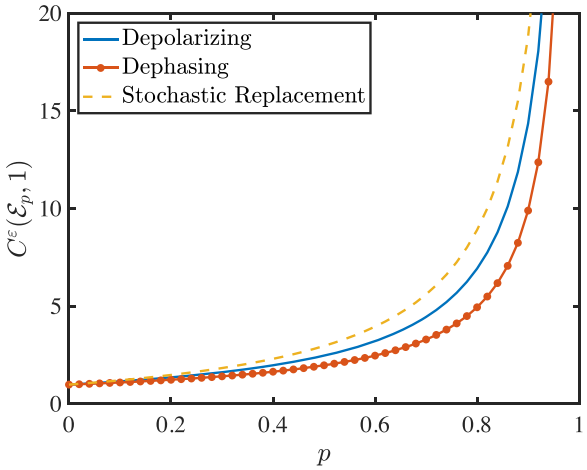

FIG. 3. Distillation overhead for depolarizing channels $\mathcal{E}_p(\rho) = p\rho + (1-p)I/2$, dephasing channels $\mathcal{E}_p(\rho) = p\rho + (1-p)Z\rho Z$, and stochastic replacement channels $\mathcal{E}_p(\rho) = p\rho + (1-p)|0\rangle\langle 0|$. Here we consider $\varepsilon = 0.01$.

We remark that this process can be considered as a channel distillation; the operations Alice and Bob can apply are considered as free operations applied to a noisy channel, with the overall process constructing a superchannel that transforms a channel $\mathcal{E}$ to (approximately) the identity channel. From this perspective, quantum capacity coincides with the distillation rate with respect to available encoding and decoding operations constructing free operations, for which resource-theoretic tools can be employed to study the properties. This resource-theoretic view of quantum communication has recently been actively studied and provided insights into the theory of quantum communication [27,31,32,51].

The framework of virtual resource distillation allows us to extend the conventional setting of quantum communication. The specific restrictions on communication scenarios, reflected in the choice of free operations applied to noisy quantum channels, largely depend on the assisting resource available, such as nonsignaling or entanglement assistance [78,79] or classical communication assistance [80]. Since our framework is applicable to general resource theories, i.e., any choice of free operations, the technique of virtual resource distillation can be applied to communication settings with very general types of physical restrictions. Here we consider the setting of the most physical relevance, which is an unassisted setting in which Alice and Bob can only make local operations on their side.

A subtlety that arises in virtual resource distillation for unassisted communication is that postprocessing after the measurement made by Bob requires one bit of preshared randomness (or classical communication from Alice to Bob) to agree on the operations they apply on each side. To avoid this, here we consider a significantly weaker setting, in which only Bob applies a probabilistic operation followed by measurement and classical postprocessing. In this setting, Bob can generate a random bit on his own and choose his operation and corresponding classical postprocessing.

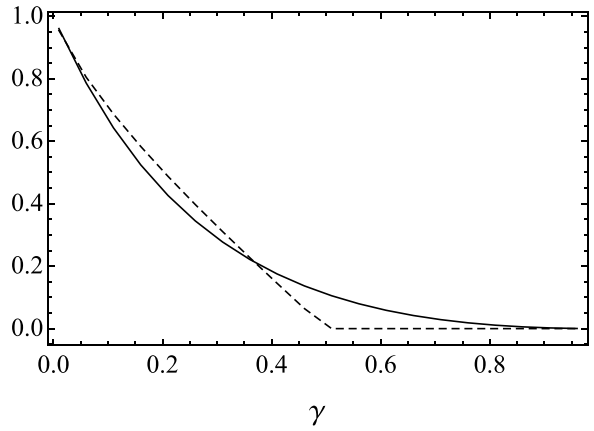

FIG. 4. Lower bound $1/C^0(\mathcal{A}_\gamma, 1)^2$ for the virtual distillation rate (solid line) and quantum capacity $Q(\mathcal{A}_\gamma)$ (dashed line) for the qubit amplitude damping channel $\mathcal{A}_\gamma$

This reduces the estimation of virtual resource distillation rate and overhead to a framework almost identical to the one discussed in the preceding section. Here our target channel is the identity channel $\mathcal{I}$. For a given noisy channel $\mathcal{E}$, the virtual resource distillation overhead with respect to the target channel $\mathcal{I}$ is then characterized by $C^\varepsilon(\mathcal{E}, m)$ defined in (137), with d the dimension of the space that $\mathcal{E}$ acts on.

Let us now focus on the evaluation of distillation overhead for the case of $\varepsilon = 0$ and $m = 1$. In this case, the overhead can be written as

$$C^0(\mathcal{E}, 1) = \min_{\{N_i\}_i, \{\lambda_i\}_i} \left\{ \sum_i |\lambda_i| \,\middle|\, \mathcal{E}^{-1} = \sum_i \lambda_i N_i \right\}, \tag{140}$$

assuming that the inverse map for $\mathcal{E}$ exists. This overhead can then give a lower bound for the virtual resource distillation rate as $V^\varepsilon(\mathcal{E}) \geqslant 1/C^0(\mathcal{E}, 1)^2$ for every $\varepsilon \in [0, 1]$.

This quantity was studied in Refs. [9,10] and shown to coincide with the diamond norm of the inverse map $\|\mathcal{E}^{-1}\|_\diamond$ [10]. The analytical expressions of $C^0(\mathcal{E}, 1)$ for some noisy channels of interest were then obtained. For instance, for the d-dimensional depolarizing channel $\mathcal{D}_p(\rho) \coloneqq (1 - p)\rho + pI/d$ we have

$$C^0(\mathcal{D}_p, 1) = \frac{1 + (1 - 2/d^2)p}{1 - p}. \tag{141}$$

This in particular provides the lower bound for the virtual distillation rate for a qubit depolarizing channel as

$$V^\varepsilon(\mathcal{D}_p) \geqslant \left(\frac{1 - p}{1 + p/2} \right)^2. \tag{142}$$

The exact variant of virtual distillation overhead in (140) is conceptually similar to zero-error quantum communication, where no error is allowed in the protocol. We note, however, that a direct comparison of the lower bound we obtained for the virtual distillation rate and the quantum capacity may not be fair, because the computation of quantum capacity assumes that (i) Alice also applies her operation and (ii) asymptotically many channel uses are allowed. More generally, the error in communication can be nonzero as long as it vanishes in the limit of infinitely many channel uses. We can see that our lower bound $1/C^0(\mathcal{D}_p, 1)^2$ can already be significantly greater than the quantum capacity. To see this, recall that the quantum capacity for qubit depolarizing channel $Q(\mathcal{D}_p)$ has a simple upper bound $Q(\mathcal{D}_p) \leqslant 1 - 4p$ for $p \leqslant 1/4$ and $Q(\mathcal{D}_p) = 0$ for $p \leqslant 1/4$ [81]. It is straightforward to check that $1 - 4p < 1/C^0(\mathcal{D}_p, 1)^2$ for $p \in (0, 1]$, which results in $Q(\mathcal{D}_p) < V^0(\mathcal{D}_p)$ for the whole range of p. In particular, $V^0(\mathcal{D}_p) > 0$ for all $p \in [0, 1]$, which is in stark contrast to the quantum capacity, which becomes $Q(\mathcal{D}_p) = 0$ for $p > 1/4$.

As another example, take the qubit amplitude damping channel $\mathcal{A}_\gamma(\rho) \coloneqq A_0 \cdot A_0^\dagger + A_1 \cdot A_1^\dagger$, with $A_0 \coloneqq |0\rangle\langle 0| + \sqrt{1 - \gamma}|1\rangle\langle 1|$ and $A_1 \coloneqq \sqrt{\gamma}|0\rangle\langle 1|$. Using the results in Refs. [9,10], the overhead can be computed as

$$C^0(\mathcal{A}_\gamma, 1) = \frac{1 + \gamma}{1 - \gamma}, \tag{143}$$

which gives a lower bound for the virtual distillation rate as $V^\varepsilon(\mathcal{A}_\gamma) \geqslant (\frac{1-\gamma}{1+\gamma})^2$. On the other hand, the quantum capacity of amplitude damping is known to be [82]

$$Q(\mathcal{A}_\gamma) = \max_t [h_2((1 - \gamma)t) - h_2(\gamma t)], \tag{144}$$

where $h_2(p) \coloneqq -p \log_2 p - (1 - p) \log_2(1 - p)$ is the binary entropy. Figure 4 plots $Q(\mathcal{A}_\gamma)$ and $1/C^0(\mathcal{A}_\gamma)^2$ for $\gamma \in [0, 1]$. This shows that the virtual distillation rate is ensured to be greater than quantum capacity for $\gamma \geqslant 0.4$ and it remains nonzero while $Q(\mathcal{A}_\gamma) = 0$ for $\gamma \geqslant 1/2$.

F. Dephased non-Markovian processes

Let us now discuss an application of virtual resource distillation in comb theories. Quantum combs become most relevant when the system and environment interact with each other, where one does not have control over the environment. Non-Markovian dynamics particularly appears when the environment sustains quantum memory over multiple time steps. In such a scenario, physically accessible operations, which we take as free operations, should be quantum combs that only act on the accessible system. Here we discuss an example where virtual resource distillation enables us to distill the environment comb that has the perfect quantum memory from the one with inferior memory.

Consider an L-step non-Markovian process that involves a qubit system and environment that interact with each other via two controlled-NOT (CNOT) gates at every time step. We let S_j and E_j refer to the system and environment at the jth time step. Every set of CNOT gates is followed by a partial dephasing channel $\mathcal{Z}_p(\rho) \coloneqq (1 - p)\rho + Z\rho Z$ in the environment, where Z is the Pauli Z operator (Fig. 5).

A quantum comb can be described by its Choi operator, which corresponds to the state obtained by inputting one end of a maximally entangled state into every input port of the comb (up to normalization) [41]. The Choi operator of our given object is written as

$$J_\Upsilon = \star_{j=1}^{L} \left(J_{\mathcal{Z}_p}^{E_j} \star J_{\text{CNOT}}^{S_j E_j} \right), \tag{145}$$

where $J_{\text{CNOT}}^{S_j E_j}$ is the Choi operator for two CNOT gates at the jth time step and $J_{\mathcal{Z}_p}^{E_j}$ is the Choi operator for $\mathcal{Z}_p$. For channels $\mathcal{M} : A \to B$ and $\mathcal{E} : B \to C$, $J_\mathcal{E} \star J_\mathcal{M} \coloneqq \text{Tr}_B(J_\mathcal{E}^{T_B} J_\mathcal{M})$ refers to

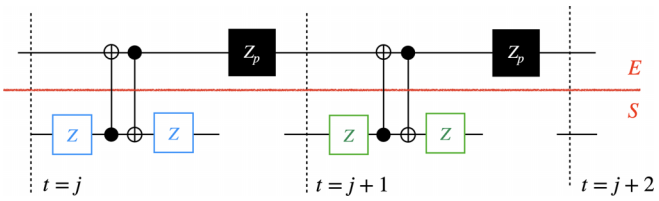

FIG. 5. Diagram showing the jth and $(j+1)$th time steps. The Z operations in blue and green are stochastic operations applied at probability p in the virtual distillation process. Two Z's in the same color are either simultaneously applied or are not applied at all, while the applications of blue and green Z's are independent.

the link product, which gives the Choi operator for the concatenated channel $\mathcal{E} \circ \mathcal{M}$ [41].

The dephasing channel degrades the quantum memory and decoheres quantum states over time. Our goal is to remove the effect of this dephasing noise in the environment by applying operations in the system. Therefore, we set our target comb Θ as

$$J_\Theta = \star_{j=1}^{L} \left(J_{\mathcal{I}}^{E_j} \star J_{\text{CNOT}}^{S_j E_j} \right). \tag{146}$$

We would like to find a set $\{\Lambda_i\}_i$ of free operations, i.e., quantum combs that act only on the system side, so that $\Theta = \sum_i \lambda_i \Lambda_i(\Upsilon)$ for some real numbers $\{\lambda_i\}_i$. To this end, it is useful to note that the inverse map $\mathcal{Z}_p^{-1}$ for the partial dephasing is decomposed as [11]

$$\mathcal{Z}_p^{-1} = \frac{1-p}{1-2p}\mathcal{I} - \frac{p}{1-2p}\mathcal{Z} \tag{147}$$

and the optimal overhead is realized by this decomposition [9,10,33], which gives $C^0(\mathcal{Z}_p, 1) = 1/(1-2p)$ with respect to the target channel $\mathcal{I}$ in light of the discussion in the two preceding sections. The implementation of this inverse map is realized by applying $\mathcal{Z}$ at probability p followed by postprocessing, i.e., multiplying $1/(1-2p)$ to the measurement outcome with a possible sign factor if Pauli Z is applied.

We now observe that the same action can be made on the environment by applying a Z operator on the system side, namely, we apply Z operators both before and after the CNOT gates at probability p, and we do not apply anything at probability $1-p$ (Fig. 5). When a Z operator is applied, the action of Z propagates to the environment through the second CNOT gate, while the effect on the system side cancels out by the second Z operator after the CNOT gates.

We can independently apply the same procedure at every time step, which constructs 2^L free operations $\{\Lambda_{\vec{i}}\}_{\vec{i}\in\{0,1\}^L}$ where the location of $1'$s in $\vec{i}$ specifies the time steps at which Z operators are applied. Letting $|\vec{i}|$ denote the number of $1'$s in $\vec{i}$, the desired linear decomposition of the target comb is written as

$$\Theta = \sum_{\vec{i}\in\{0,1\}^L} (-1)^{|\vec{i}|}(1-p)^{L-|\vec{i}|}p^{|\vec{i}|}\Lambda_{\vec{i}}(\Upsilon), \tag{148}$$

which gives

$$C^0(\Upsilon, 1) \leqslant \frac{1}{(1-2p)^L}. \tag{149}$$

We conjecture that equality holds because this is essentially the most efficient way to counteract the dephasing, although we leave the full investigation to future work.

The implementation of the virtual resource distillation then is realized as follows.

(1) At each step, apply Z operators before and after the CNOT gates at probability p and do nothing at probability $1-p$. Record which operation was applied.

(2) Multiply $(-1)^{\text{sgn}}(1-2p)^L$ to the measurement outcome, where sgn takes the value 0 if the total number of time steps at which Z operators were applied was even and 1 if odd.

(3) Repeat the same procedure many times and take a sample average of the postprocessed measurement outcomes.

This example ensures that our virtual resource distillation framework in comb theories can be applied to memory preservation in the environment by manipulating the accessible system only. This may be seen as an error mitigation protocol applied to non-Markovian dynamics, which was previously studied in several other settings [83,84].

VII. CONCLUSION

We presented a framework of virtual resource distillation applicable to general resource theories with an arbitrary set of convex free objects and free operations, including general types of quantum objects such as quantum states, channels, and higher-order processes represented by quantum combs. We derived various expressions and bounds for virtual resource distillation rate and overhead, in both general settings and concrete theories of practical interest, demonstrating its versatility and broad applicability.

Promising future directions include obtaining explicit evaluations of the performance of probabilistic virtual resource distillation, which may find use in practical settings that are not possible to characterize using only deterministic protocols. It will also be interesting to consider experimental implementations of virtual resource distillation that are possible on today's quantum devices.

A further open question is the relation between virtual distillation rates and asymptotic rates of conventional distillation. Although, as remarked earlier and in [12], the latter are very different from our approach (requiring in particular coherent manipulation of many-copy input states $\rho^{\otimes n}$ with an unbounded number of copies), it would nevertheless be interesting to understand whether virtual distillation can already improve on such rates. Furthermore, the many-copy extension of our virtual framework, i.e., the behavior of $V^\varepsilon(\rho^{\otimes n})$ when more copies of the input state ρ can be manipulated coherently, is an interesting question of its own. Although this sacrifices the experimentally friendly character of virtual distillation protocols, it could be useful to understand to what extent virtual distillation capabilities can be improved through such many-copy protocols, leading to the delineation of the ultimate limits of virtual quantum resource distillation.

ACKNOWLEDGMENTS

We thank Suguru Endo, Patrick Hayden, Jayne Thompson, and Mark M. Wilde for insightful discussions. This work was supported by the National Natural Science Foundation of

China Grant No. 12175003, NSAF (Grant No. U2330201); the Singapore Ministry of Education Tier 1 Grant No. RG77/22; the National Research Foundation, Singapore; Agency for Science, Technology and Research (A*STAR) under its QEP2.0 program (No. NRF2021-QEP2-02-P06); and the Singapore Ministry of Education Tier 2 Grant No. MOE-T2EP50221-0005. B.R. was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant No. 22KF0067 and the JSPS Postdoctoral Fellowship for Research in Japan. R.T. was supported by the Lee Kuan Yew Postdoctoral Fellowship at Nanyang Technological University Singapore.

APPENDIX: DUAL FORMULATION OF C^ε

Here we prove Eq. (36), that is, establish a dual form of the virtual distillation cost C^ε. To begin, we write

$$\begin{aligned} C^\varepsilon(\rho,m) &= \inf_{\tilde\tau\sim_\varepsilon\tau^{\otimes m}} \inf\{\lambda_+ + \lambda_- \mid \tilde\tau = \lambda_+\Lambda_+(\rho) - \lambda_-\Lambda_-(\rho),\ \lambda_\pm \geqslant 0,\ \lambda_+ - \lambda_- = 1,\ \Lambda_\pm \in O\} \\ &= \inf_{\substack{\tilde\tau\sim_\varepsilon\tau^{\otimes m}\\ \mathrm{Tr}\tilde\tau=1}} \inf\{\lambda_+ + \lambda_- \mid \tilde\tau = \lambda_+\Lambda_+(\rho) - \lambda_-\Lambda_-(\rho),\ \lambda_\pm \geqslant 0,\ \Lambda_\pm \in O\} \\ &= \inf_{\substack{\tilde\tau\sim_\varepsilon\tau^{\otimes m}\\ \mathrm{Tr}\tilde\tau=1}} \inf\{\mathrm{Tr}\tilde\Lambda_+(\rho) + \mathrm{Tr}\tilde\Lambda_-(\rho) \mid \tilde\tau = \tilde\Lambda_+(\rho) - \tilde\Lambda_-(\rho),\ \tilde\Lambda_\pm \in \mathrm{cone}(O)\}, \end{aligned} \tag{A1}$$

where $\mathrm{cone}(O) = \{\lambda\Lambda \mid \lambda \geqslant 0,\ \Lambda \in O\}$. We will now take the Lagrange dual of the inner minimisation. The Lagrangian of this problem is

$$\mathcal{L}(\tilde\Lambda_\pm; H, X, Y) = \mathrm{Tr}\tilde\Lambda_+(\rho) + \mathrm{Tr}\tilde\Lambda_-(\rho) - \mathrm{Tr}\{H[\tilde\Lambda_+(\rho) - \tilde\Lambda_-(\rho) - \tilde\tau]\} - \mathrm{Tr}XJ_{\tilde\Lambda_+} - \mathrm{Tr}YJ_{\tilde\Lambda_-}, \tag{A2}$$

where J_Λ denotes the Choi operator of the corresponding map and H, X, Y are Lagrange multipliers satisfying $\mathrm{Tr}XJ_\Lambda \geqslant 0\,\forall\,\Lambda \in O$ and analogously for Y. Using the Choi-Jamiołkowski isomorphism, we can rewrite this as

$$\mathcal{L}(\tilde\Lambda_\pm; H, X, Y) = \mathrm{Tr}(I\otimes\rho^T)J_{\tilde\Lambda_+} + \mathrm{Tr}(I\otimes\rho^T)J_{\tilde\Lambda_-} + \mathrm{Tr}H\tilde\tau - \mathrm{Tr}(H\otimes\rho^T)J_{\tilde\Lambda_+} + \mathrm{Tr}(H\otimes\rho^T)J_{\tilde\Lambda_-(\rho)} - \mathrm{Tr}XJ_{\tilde\Lambda_+} - \mathrm{Tr}YJ_{\tilde\Lambda_-}. \tag{A3}$$

By definition, the dual problem is then [85]

$$\begin{aligned} \sup_{\substack{H\in\mathrm{Herm}\\ X:\mathrm{Tr}XJ_\Lambda\geqslant 0\,\forall\Lambda\in O\\ Y:\mathrm{Tr}YJ_\Lambda\geqslant 0\,\forall\Lambda\in O}} \inf_{\tilde\Lambda_\pm\in\mathrm{Herm}} \mathcal{L}(\tilde\Lambda_\pm; H, X, Y) &= \sup\{\mathrm{Tr}H\tilde\tau \mid \mathrm{Tr}[(I-H)\otimes\rho^T]J_\Lambda \geqslant 0\,\forall\,\Lambda\in O,\ \mathrm{Tr}[(I+H)\otimes\rho^T]J_\Lambda \geqslant 0\,\forall\,\Lambda\in O\} \\ &= \sup\{\mathrm{Tr}H\tilde\tau \mid -1 \leqslant \mathrm{Tr}H\Lambda(\rho) \leqslant 1\,\forall\,\Lambda\in O\}. \end{aligned}$$

Since $H = 0$ is strictly feasible for the above, by Slater's theorem we have that the optimal values of the primal and dual optimization problems are equal. A change of variables $W := 2H - I$ gives the form stated in Eq. (36).

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).
[2] A. Streltsov, G. Adesso, and M. B. Plenio, *Colloquium*: Quantum coherence as a resource, Rev. Mod. Phys. **89**, 041003 (2017).
[3] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. **91**, 025001 (2019).
[4] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).
[5] J. Cotler, S. Choi, A. Lukin, H. Gharibyan, T. Grover, M. E. Tai, M. Rispoli, R. Schittko, P. M. Preiss, A. M. Kaufman, M. Greiner, H. Pichler, and P. Hayden, Quantum virtual cooling, Phys. Rev. X **9**, 031013 (2019).
[6] B. Koczor, Exponential error suppression for near-term quantum devices, Phys. Rev. X **11**, 031057 (2021).
[7] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, Phys. Rev. X **11**, 041036 (2021).
[8] F. Buscemi, M. Dall'Arno, M. Ozawa, and V. Vedral, Direct observation of any two-point quantum correlation function, arXiv:1312.4240
[9] J. Jiang, K. Wang, and X. Wang, Physical implementability of linear maps and its application in error mitigation, Quantum **5**, 600 (2021).
[10] B. Regula, R. Takagi, and M. Gu, Operational applications of the diamond norm and related measures in quantifying the non-physicality of quantum maps, Quantum **5**, 522 (2021).

[11] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).
[12] X. Yuan, B. Regula, R. Takagi, and M. Gu, companion paper, Virtual quantum resource distillation, Phys. Rev. Lett. **132**, 050203 (2024).
[13] K. Ben Dana, M. García Díaz, M. Mejatty, and A. Winter, Resource theory of coherence: Beyond states, Phys. Rev. A **95**, 062327 (2017).
[14] M. G. Díaz, K. Fang, X. Wang, M. Rosati, M. Skotiniotis, J. Calsamiglia, and A. Winter, Using and reusing coherence to realize quantum processes, Quantum **2**, 100 (2018).
[15] D. Rosset, F. Buscemi, and Y.-C. Liang, Resource theory of quantum memories and their faithful verification with minimal assumptions, Phys. Rev. X **8**, 021033 (2018).
[16] S. Bauml, S. Das, X. Wang, and M. M. Wilde, Resource theory of entanglement for bipartite quantum channels, arXiv:1907.04181.
[17] G. Gour and C. M. Scandolo, Dynamical entanglement, Phys. Rev. Lett. **125**, 180505 (2020).
[18] G. Saxena, E. Chitambar, and G. Gour, Dynamical resource theory of quantum coherence, Phys. Rev. Res. **2**, 023298 (2020).
[19] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Fundamental limits of repeaterless quantum communications, Nat. Commun. **8**, 15043 (2017).
[20] P. Faist and R. Renner, Fundamental work cost of quantum processes, Phys. Rev. X **8**, 021011 (2018).
[21] T. Theurer, D. Egloff, L. Zhang, and M. B. Plenio, Quantifying operations with an application to coherence, Phys. Rev. Lett. **122**, 190405 (2019).
[22] R. Takagi and B. Regula, General resource theories in quantum mechanics and beyond: Operational characterization via discrimination tasks, Phys. Rev. X **9**, 031053 (2019).
[23] Y. Liu and X. Yuan, Operational resource theory of quantum channels, Phys. Rev. Res. **2**, 012035(R) (2020).
[24] Z.-W. Liu and A. Winter, Resource theories of quantum channels and the universal role of resource erasure, arXiv:1904.04201.
[25] X. Wang, M. M. Wilde, and Y. Su, Quantifying the magic of quantum channels, New J. Phys. **21**, 103002 (2019).
[26] X. Wang and M. M. Wilde, Resource theory of asymmetric distinguishability for quantum channels, Phys. Rev. Res. **1**, 033169 (2019).
[27] R. Takagi, K. Wang, and M. Hayashi, Application of the resource theory of channels to communication scenarios, Phys. Rev. Lett. **124**, 120502 (2020).
[28] H. Kristjánsson, G. Chiribella, S. Salek, D. Ebler, and M. Wilson, Resource theories of communication, New J. Phys. **22**, 073014 (2020).
[29] G. Gour and A. Winter, How to quantify a dynamical quantum resource, Phys. Rev. Lett. **123**, 150401 (2019).
[30] X. Yuan, Y. Liu, Q. Zhao, B. Regula, J. Thompson, and M. Gu, Universal and operational benchmarking of quantum memories, npj Quantum Inf. **7**, 108 (2021).
[31] B. Regula and R. Takagi, Fundamental limitations on distillation of quantum channel resources, Nat. Commun. **12**, 4411 (2021).
[32] K. Fang and Z.-W. Liu, No-go theorems for quantum resource purification: New approach and channel theory, PRX Quantum **3**, 010337 (2022).
[33] R. Takagi, Optimal resource cost for error mitigation, Phys. Rev. Res. **3**, 033178 (2021).
[34] G. D. Berk, A. J. P. Garner, B. Yadin, K. Modi, and F. A. Pollock, Resource theories of multi-time processes: A window into quantum non-Markovianity, Quantum **5**, 435 (2021).
[35] M. Horodecki, P. Horodecki, and R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, Phys. Rev. A **60**, 1888 (1999).
[36] D. Jonathan and M. B. Plenio, Minimal conditions for local pure-state entanglement manipulation, Phys. Rev. Lett. **83**, 1455 (1999).
[37] K. Fang, X. Wang, L. Lami, B. Regula, and G. Adesso, Probabilistic distillation of quantum coherence, Phys. Rev. Lett. **121**, 070404 (2018).
[38] K. Fang and Z.-W. Liu, No-go theorems for quantum resource purification, Phys. Rev. Lett. **125**, 060405 (2020).
[39] B. Regula, Probabilistic transformations of quantum resources, Phys. Rev. Lett. **128**, 110505 (2022).
[40] B. Regula, Tight constraints on probabilistic convertibility of quantum states, Quantum **6**, 817 (2022).
[41] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum circuit architecture, Phys. Rev. Lett. **101**, 060401 (2008).
[42] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Transforming quantum operations: Quantum supermaps, Europhys. Lett. **83**, 30004 (2008).
[43] G. D. Berk, S. Milz, F. A. Pollock, and K. Modi, Extracting quantum dynamical resources: Consumption of non-Markovianity for noise reduction, arXiv:2110.02613.
[44] M. Horodecki and J. Oppenheim, (Quantumness in the context of) resource theories, Int. J. Mod. Phys. B **27**, 1345019 (2013).
[45] R. Takagi, B. Regula, K. Bu, Z.-W. Liu, and G. Adesso, Operational advantage of quantum resources in subchannel discrimination, Phys. Rev. Lett. **122**, 140402 (2019).
[46] R. Uola, T. Kraft, J. Shang, X.-D. Yu, and O. Gühne, Quantifying quantum resources with conic programming, Phys. Rev. Lett. **122**, 130404 (2019).
[47] M. Oszmaniec and T. Biswas, Operational relevance of resource theories of quantum measurements, Quantum **3**, 133 (2019).
[48] A. Anshu, M.-H. Hsieh, and R. Jain, Quantifying resources in general resource theory with catalysts, Phys. Rev. Lett. **121**, 190504 (2018).
[49] Z.-W. Liu, K. Bu, and R. Takagi, One-shot operational quantum resource theory, Phys. Rev. Lett. **123**, 020401 (2019).
[50] B. Regula, K. Bu, R. Takagi, and Z.-W. Liu, Benchmarking one-shot distillation in general quantum resource theories, Phys. Rev. A **101**, 062315 (2020).
[51] B. Regula and R. Takagi, One-shot manipulation of dynamical quantum resources, Phys. Rev. Lett. **127**, 060402 (2021).
[52] R. Takagi, B. Regula, and M. M. Wilde, One-shot yield-cost relations in general quantum resource theories, PRX Quantum **3**, 010348 (2022).
[53] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russ. Math. Surv. **52**, 1191 (1997).
[54] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018), Theorem 3.46.

[55] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Memory effects in quantum channel discrimination, Phys. Rev. Lett. **101**, 180501 (2008).
[56] G. Gutoski, On a measure of distance for quantum strategies, J. Math. Phys. **53**, 032202 (2012).
[57] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. **58**, 13 (1963).
[58] B. Regula and L. Lami, Functional analytic insights into irreversibility of quantum resources, arXiv:2211.15678.
[59] J. Burniston, M. Grabowecky, C. M. Scandolo, G. Chiribella, and G. Gour, Necessary and sufficient conditions on measurements of quantum channels, Proc. R. Soc. A **476**, 20190832 (2020).
[60] G. Vidal and R. Tarrach, Robustness of entanglement, Phys. Rev. A **59**, 141 (1999).
[61] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).
[62] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying coherence, Phys. Rev. Lett. **113**, 140401 (2014).
[63] M. Steiner, Generalized robustness of entanglement, Phys. Rev. A **67**, 054305 (2003).
[64] A. W. Harrow and M. A. Nielsen, Robustness of quantum gates in the presence of noise, Phys. Rev. A **68**, 012308 (2003).
[65] N. Johnston, C.-K. Li, S. Plosker, Y.-T. Poon, and B. Regula, Evaluating the robustness of k-coherence and k-entanglement, Phys. Rev. A **98**, 022328 (2018).
[66] N. Datta, Min- and max-relative entropies and a new entanglement monotone, IEEE Trans. Inf. Theor. **55**, 2816 (2009).
[67] B. Regula, K. Fang, X. Wang, and M. Gu, One-shot entanglement distillation beyond local operations and classical communication, New J. Phys. **21**, 103017 (2019).
[68] M. Horodecki and P. Horodecki, Reduction criterion of separability and limits for a class of distillation protocols, Phys. Rev. A **59**, 4206 (1999).
[69] A. Shimony, Degree of entanglement, Ann. N.Y. Acad. Sci. **755**, 675 (1995).
[70] V. Veitch, S. H. Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer quantum computation, New J. Phys. **16**, 013009 (2014).
[71] M. Howard and E. Campbell, Application of a resource theory for magic states to fault-tolerant quantum computing, Phys. Rev. Lett. **118**, 090501 (2017).
[72] M. Lostaglio and A. Ciani, Error mitigation and quantum-assisted simulation in the error corrected regime, Phys. Rev. Lett. **127**, 200506 (2021).
[73] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, Error mitigation for universal gates on encoded qubits, Phys. Rev. Lett. **127**, 200505 (2021).
[74] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error reduction technique: Applications from the NISQ to the fault-tolerant quantum computing eras, PRX Quantum **3**, 010345 (2022).
[75] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, Negative quasi-probability as a resource for quantum computation, New J. Phys. **14**, 113011 (2012).
[76] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).
[77] J. Watrous, Semidefinite programs for completely bounded norms, Theor. Comput. **5**, 217 (2009).
[78] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem, IEEE Trans. Inf. Theory **48**, 2637 (2002).
[79] D. Leung and W. Matthews, On the power of PPT-preserving and non-signalling codes, IEEE Trans. Inf. Theory **61**, 4486 (2015).
[80] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, Phys. Rev. A **54**, 3824 (1996).
[81] G. Smith and J. A. Smolin, *2008 IEEE Information Theory Workshop* (IEEE, Piscataway, 2008), pp. 368–372.
[82] V. Giovannetti and R. Fazio, Information-capacity description of spin-chain correlations, Phys. Rev. A **71**, 032314 (2005).
[83] H. Hakoshima, Y. Matsuzaki, and S. Endo, Relationship between costs for quantum error mitigation and non-Markovian measures, Phys. Rev. A **103**, 012611 (2021).
[84] M. Ho, R. Takagi, and M. Gu, Enhancing quantum models of stochastic processes with error mitigation, arXiv:2105.06448.
[85] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004).