Virtual Quantum Resource Distillation

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

[1]*Center on Frontiers of Computing Studies, Peking University, Beijing 100871, China*
[2]*School of Computer Science, Peking University, Beijing 100871, China*
[3]*Mathematical Quantum Information RIKEN Hakubi Research Team, RIKEN Cluster for Pioneering Research (CPR) and RIKEN Center for Quantum Computing (RQC), Wako, Saitama 351-0198, Japan*
[4]*Department of Physics, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*
[5]*Department of Basic Science, The University of Tokyo, Tokyo 153-8902, Japan*
[6]*Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, 637371, Singapore*
[7]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543, Singapore*
[8]*CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, Singapore 117543, Singapore*

Distillation, or purification, is central to the practical use of quantum resources in noisy settings often encountered in quantum communication and computation. Conventionally, distillation requires using some restricted "free" operations to convert a noisy state into one that approximates a desired pure state. Here, we propose to relax this setting by only requiring the approximation of the measurement statistics of a target pure state, which allows for additional classical postprocessing of the measurement outcomes. We show that this extended scenario, which we call "virtual resource distillation," provides considerable advantages over standard notions of distillation, allowing for the purification of noisy states from which no resources can be distilled conventionally. We show that general states can be virtually distilled with a cost (measurement overhead) that is inversely proportional to the amount of existing resource, and we develop methods to efficiently estimate such cost via convex and semidefinite programming, giving several computable bounds. We consider applications to coherence, entanglement, and magic distillation, and an explicit example in quantum teleportation (distributed quantum computing). This work opens a new avenue for investigating generalized ways to manipulate quantum resources.

One important aspect in the manipulation of any valuable physical resource is distillation—a means to extract this resource in optimal, purified form from some crude, noisy source. This finds use in quantum information processing, where various quantum resources [1–3]—entanglement [4–6], coherence [7–9], or magic [10–12], among many others—have gained interest for their role in empowering various quantum information protocols [1–3]. However, the inevitable imperfections that permeate near-term quantum technologies mean that noiseless resources may not be readily available. Thus, distillation of such resources is critical for developing practical quantum computation and communication schemes, allowing for systematic means to obtain ideal resources from ones distorted by noise.

Powerful theoretical results have been obtained for both one-shot and asymptotic resource distillation [13–28]. Yet, especially in the one-shot scenario, strong limitations exist that prohibit resource distillation even from highly resourceful states, either demanding many copies of the resource state to enable a successful conversion, or incurring large errors in the process [22,24,25,28]. While those results are fundamental no-go theorems based on the laws of quantum mechanics, could there nevertheless be means to sidestep them by somewhat bending the rules?

Here, we give an affirmative answer to this question by proposing a new paradigm of *virtual resource distillation*. Since the output of many quantum information protocols is ultimately purely classical, it often does not matter whether a particular ideal resource state (e.g., a Bell state) was actually synthesized, provided we can reproduce the same expected outcome statistics. Therefore, we focus on distilling a "virtual" target resource state, in the sense that any operation followed by a measurement of the target resource can be approximated to a desired accuracy. This combines the power of both classical data processing and resource manipulation, performing better than either of these two approaches on their own. We show that virtual distillation enables us to effectively increase the distillation efficiency at an increased cost in the measurement samples. We study properties of the distillation overhead and show how it can be tightly bounded using resource monotones and semidefinite programs. We show that various limitations on distillation schemes can be circumvented by employing our

framework, enabling distillation even from states from which no resource can be extracted conventionally. We give examples in the virtual distillation of coherence, entanglement, and magic, and discuss applications.

Here, we focus on the case where the manipulated objects are quantum states, and we leave the broader framework and complete technical proofs to the companion paper [29].

Background.—A resource theory of states consists of the set $\mathcal{F}$ of free states and the set $\mathcal{O}$ of free operations. A weak and undemanding assumption about the free operations $\Lambda \in \mathcal{O}$ is the so-called "golden rule," which states that $\Lambda(\sigma) \in \mathcal{F}$ for all $\sigma \in \mathcal{F}$. The set of all operations that satisfy the golden rule, called resource nongenerating operations, is then the largest consistent set of free operations. A resource monotone μ is a function that measures the amount of resource a state possesses, satisfying the monotonicity requirement $\mu(\rho) \geq \mu(\Lambda(\rho))$, $\forall \Lambda \in \mathcal{O}$.

In most resource theories, we can define an optimal unit pure resource state, denoted as ψ. Distillation then captures the task of synthesizing ψ from an imperfect state ρ. The one-shot resource distillation rate

$$D^\varepsilon(\rho) = \sup_{\Lambda \in \mathcal{O}} \left\{ m : \frac{1}{2} \|\Lambda(\rho) - \psi^{\otimes m}\|_1 \leq \varepsilon \right\} \quad (1)$$

then defines number of optimal states we can synthesize with ρ at allowable error $\varepsilon \in [0, 1)$. Here, $\|A\|_1 = \text{Tr}[\sqrt{A^\dagger A}]$ is the trace norm.

Virtual resource distillation.—In many settings, distillation is followed by a processing of the state ψ to produce an expectation value x of some observable. Thus, ψ does not strictly need to actually exist: a simulation of ψ that enables accurate retrieval of x is sufficient. We will thus assume that the state is ultimately measured. This contrasts with conventional distillation, where ψ is always synthesized physically.

Formally, consider a task that involves applying some operation $\mathcal{N}$ on the resource state (possibly together with other states), and measuring the resulting state in some

observable M to retrieve x. In order for a distillation protocol Λ to be successful, we thus require that measuring $\mathcal{N} \circ \Lambda(\rho)$ approximates the measurement outcomes of $\mathcal{N}(\psi^{\otimes m})$ for any choice of a channel $\mathcal{N}$ and measurement M. Since applying a channel $\mathcal{N}$ cannot make the error any larger, our requirement for distillation is in fact equivalent to the statement that $|\text{Tr} M \Lambda(\rho) - \text{Tr} M \psi^{\otimes m}| \leq \varepsilon$ for any Hermitian operator M satisfying $0 \leq M \leq I$. This condition is the same as the one for conventional distillation in Eq. (1), so we have gained no advantage. However, since $\text{Tr} M \Lambda(\rho)$ is a classical result, we can further apply classical postprocessing with different distillation operations. Specifically, we can consider a linear combination of the classical results $\sum_j \lambda_j \text{Tr}(M \Lambda_j(\rho)) = \text{Tr}(M \sum_j \lambda_j \Lambda_j(\rho))$ using different choices of $\{\Lambda_j\} \subseteq \mathcal{O}$ and real coefficients λ_j satisfying $\sum_j \lambda_j = 1$. Grouping free operations with the same sign together, we have $|\text{Tr}[M(\lambda_+ \Lambda_+(\rho) - \lambda_- \Lambda_-(\rho))] - \text{Tr} M \psi^{\otimes m}| \leq \varepsilon$, where $\lambda_\pm = \sum_{j:\text{sign}(j)=\pm 1} \lambda_j \geq 0$, $\lambda_+ - \lambda_- = 1$, and $\Lambda_\pm = (1/\lambda_\pm) \sum_{j:\text{sign}(j)=\pm 1} \lambda_j \Lambda_j$. This is equivalent to the virtual distillation condition

$$\frac{1}{2} \|\tilde{\Lambda}(\rho) - \psi^{\otimes m}\|_1 \leq \varepsilon, \quad (2)$$

where we define $\tilde{\Lambda} = \lambda_+ \Lambda_+ - \lambda_- \Lambda_-$ to be a virtual operation; see also Fig. 1. We note that the condition is independent of the operation $\mathcal{N}$ or measurement M.

In practice, we can effectively implement $\tilde{\Lambda}$ by following a Monte Carlo-based approach that often finds use in the simulation of quantum circuits [12,30–34], quantum error mitigation [35–37], and the implementation of unphysical processes [38–40]. The basic idea is to notice that for any M, we have $\text{Tr} M \tilde{\Lambda}(\rho) = C[\text{sign}(\Lambda_+) p_+ \text{Tr} M \Lambda_+(\rho) + \text{sign}(\Lambda_-) p_- \text{Tr} M \Lambda_-(\rho)]$. Therefore, we can obtain $\text{Tr} M \tilde{\Lambda}(\rho)$ by randomly applying $\Lambda_\pm$ with probability $p_\pm = \lambda_\pm / (\lambda_+ + \lambda_-)$ and multiply each classical outcome by $C\text{sign}(\Lambda_\pm) = \pm 1$. Here, $C := \lambda_+ + \lambda_- \geq 1$ contributes to a larger variance of the outcome distribution. This essentially increases the number of samples by a factor of C^2
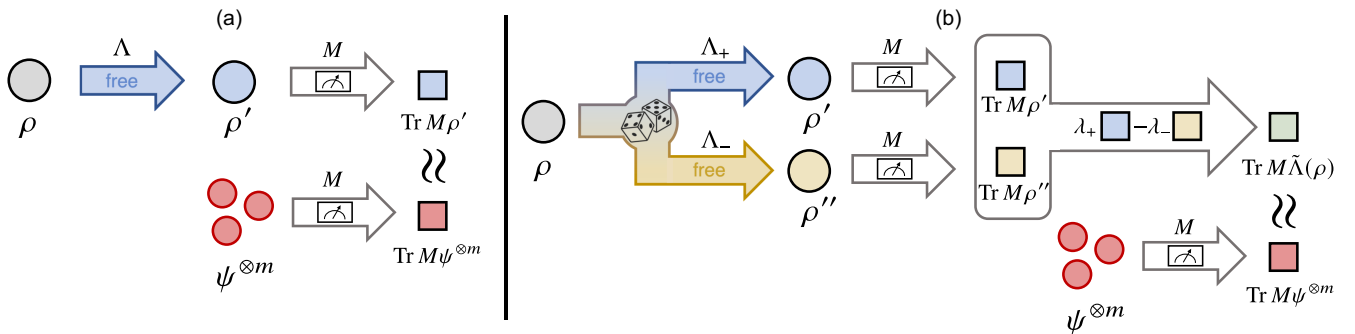

FIG. 1. Two different approaches to resource distillation. (a) Conventional resource distillation employs a free operation Λ to map ρ into a state that approximates the target state $\psi^{\otimes m}$. (b) Virtual distillation approximates the measurement outcomes of $\psi^{\otimes m}$ by using the virtual operation $\tilde{\Lambda} = \lambda_+ \Lambda_+ - \lambda_- \Lambda_-$, a linear combination of free operations.

compared to the case of conventional distillation where resource states $\psi^{\otimes m}$ are available [41]. This means that the effective number of ψ virtually obtained as $\tilde{\Lambda}(\rho)$ is reduced by a factor of $1/C^2$ for the purpose of estimating the expectation value of an observable with the desired accuracy.

This observation motivates us to define the virtual resource distillation rate as

$$V^{\varepsilon}(\rho) = \sup_m \frac{m}{C^{\varepsilon}(\rho, m)^2}, \tag{3}$$

with the overhead $C^{\varepsilon}(\rho, m)$ of virtual operations

$$C^{\varepsilon}(\rho, m) = \inf_{\substack{\tilde{\Lambda} = \lambda_+ \Lambda_+ - \lambda_- \Lambda_- \\ \lambda_+ - \lambda_- = 1 \\ \Lambda_{\pm} \in \mathcal{O}, \lambda_{\pm} \geq 0}} \left\{ \lambda_+ + \lambda_- : \frac{1}{2} \|\tilde{\Lambda}(\rho) - \psi^{\otimes m}\|_1 \leq \varepsilon \right\}. \tag{4}$$

The virtual distillation rate V^{ε} generalizes the conventional one-shot distillation rate D^{ε}, recovering the latter when optimization in (4) is restricted to the case of $\lambda_- = 0$. This immediately implies that $D^{\varepsilon}(\rho) \leq V^{\varepsilon}(\rho)$. Since $V^{\varepsilon}(\rho)$ is fully determined by $C^{\varepsilon}(\rho, m)$, we focus on the estimation of $C^{\varepsilon}(\rho, m)$ in the following.

We note here two major differences between the settings of virtual and conventional distillation. First, our approach is to perform distillation by repeatedly applying free operations to a noisy resource state to obtain a sufficient number of samples. This is very different from conventional distillation protocols used in practice, which work by employing *joint* operations on multiple copies of a state [13,43]. Coherent many-copy manipulation through joint operations is experimentally much more difficult to realize, often making resource distillation challenging or inefficient in practice. Virtual distillation overcomes this by only requiring single-copy operations. Our approach can also be compared with *probabilistic* one-shot distillation protocols in the conventional case, which also repeatedly realize single-copy operations; we address this in more detail in [29]. We will shortly see that virtual distillation does not suffer from limitations that constrain conventional (including probabilistic) distillation in equivalent settings.

A second, perhaps counterintuitive difference is that virtual distillation allows for the distillation from free (resourceless) states at a nonzero rate. This becomes possible as we relaxed the distillation requirement to only approximate the outcome statistics: for a free input state, our method essentially reduces to classical data processing, and simply reflects that the cost of classically simulating the target state ψ using free states $\mathcal{F}$ is not infinite.

Estimation of $C^{\varepsilon}(\rho, m)$.—We introduce bounds on $C^{\varepsilon}(\rho, m)$ in general resource theories, which rely on two ingredients. First, we introduce the optimization problem

$$\zeta_{\varepsilon}^s(\rho, k) := \text{minimize} \quad \mu_+ + \mu_-$$
$$\text{s.t.} \quad 0 \leq Q_+ \leq \mu_+ I, \qquad 0 \leq Q_- \leq \mu_- I,$$
$$\text{Tr}Q_+ \sigma \leq \frac{\mu_+}{k}, \qquad \text{Tr}Q_- \sigma \leq \frac{\mu_-}{k} \quad \forall \sigma \in \mathcal{F},$$
$$\mu_+ - \mu_- = 1, \qquad \text{Tr}\rho(Q_+ - Q_-) \geq 1 - \varepsilon, \tag{5}$$

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 in the second line become equality constraints: $\text{Tr}Q_+ \sigma = \mu_+/k$, $\text{Tr}Q_- \sigma = \mu_-/k \; \forall \sigma \in \mathcal{F}$. The second ingredient are three different resource measures for the target pure resource state ψ: the generalized robustness $R_{\mathcal{F}}^g(\rho) := \inf \{\lambda | [(\rho + \lambda \omega)/(1 + \lambda)] \in \mathcal{F}, \omega \in \mathcal{D}\}$, the standard robustness $R_{\mathcal{F}}^s(\rho) := \inf \{\lambda | [(\rho + \lambda \sigma)/(1 + \lambda)] \in \mathcal{F}, \sigma \in \mathcal{F}\}$, and the resource fidelity $F_{\mathcal{F}}(\rho) := \max_{\sigma \in \mathcal{F}} (\text{Tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}})^2$. Here, $\mathcal{D}$ is the set of all density matrices. Our first main result is as follows.

Theorem 1 (Theorem 1 in companion paper [29]).— Consider a convex resource theory and a target pure resource state ψ. Let $\mathcal{O}$ be the class of resource nongenerating operations. If $R_{\mathcal{F}}^s(\psi) < \infty$, then $\zeta_{\varepsilon}^s(\rho, F_{\mathcal{F}}(\psi^{\otimes m})^{-1}) \leq C^{\varepsilon}(\rho, m) \leq \zeta_{\varepsilon}^s(\rho, R_{\mathcal{F}}^s(\psi^{\otimes m}) + 1)$. Furthermore, if it holds that $\langle \psi | \sigma | \psi \rangle$ is constant for all $\sigma \in \mathcal{F}$, then $\zeta_{\varepsilon}^g(\rho, F_{\mathcal{F}}(\psi^{\otimes m})^{-1}) \leq C^{\varepsilon}(\rho, m) \leq \zeta_{\varepsilon}^g(\rho, R_{\mathcal{F}}^g(\psi^{\otimes m}) + 1)$.

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 or multilevel quantum coherence—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 $C^{\varepsilon}(\rho, m)$, i.e.,

$$C^{\varepsilon}(\rho, m) = \zeta_{\varepsilon}^{s/g}(\rho, F_{\mathcal{F}}(\psi^{\otimes m})^{-1}). \tag{6}$$

Recall that the problems $\zeta_{\varepsilon}^{s/g}$ are convex optimization problems and are often (e.g., for coherence, magic states, or nonpositive partial transpose) semidefinite programs, which are efficiently computable in the dimension of the state space [44], allowing for an exact evaluation of C^{ε} in relevant cases. Importantly, they remove the need to optimize over all operations $\Lambda \in \mathcal{O}$, which may be a significantly more difficult problem.

We note that the question of whether there exist suitable states for which the two bounds coincide is, in general, highly dependent on the resource theory in question [29].

Next, we show that the overhead is not only computable numerically, but in fact an exact expression for it can be obtained in terms of another resource monotone, $f_{\mathcal{O}}(\rho, m)$, which measures the maximum overlap between $\Lambda(\rho)$ and $\psi^{\otimes m}$ as

$$f_{\mathcal{O}}(\rho, m) := \max_{\Lambda \in \mathcal{O}} \text{Tr}[\Lambda(\rho)\psi^{\otimes m}]. \tag{7}$$
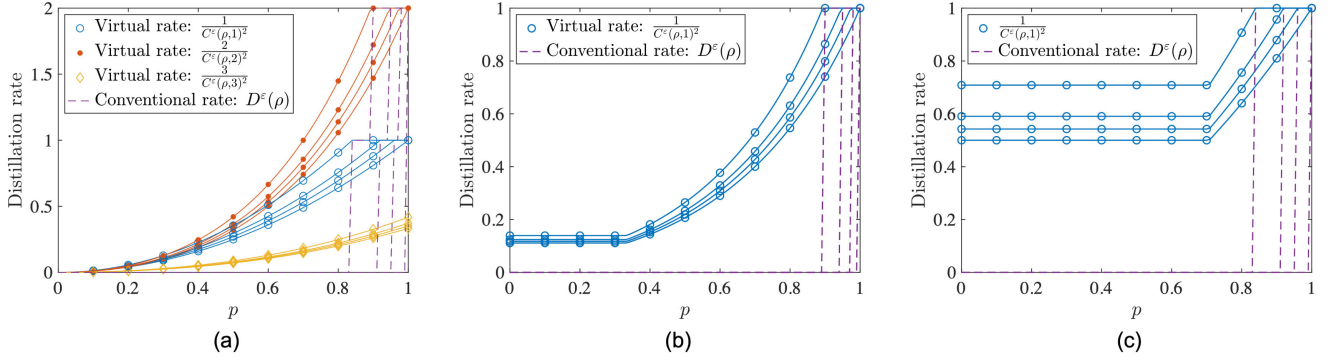
FIG. 2. Comparison between conventional and virtual one-shot resource distillation. We consider four different smoothing parameters $\varepsilon = \{0, 0.02, 0.04, 0.08\}$ corresponding to each of the four lines (with the same color or dot type) from bottom to top. (a) Distillation from the four-dimensional coherent state $\rho_p = p\psi_4 + (1-p)I_4/4$ under MIO and DIO. Here, $|\psi_4\rangle = (|0\rangle + |1\rangle + |2\rangle + |3\rangle)/2$ and I_4 is the four-dimensional identity matrix. We also consider virtual distillation with different m as in Eq. (3) and find that $m = 2$ outperforms the others for most cases (except for very small p where $m = 1$ performs the best). (b) Distillation of the two-qubit entangled state $\rho_p = p\psi + (1-p)I_4/4$ under SEPP operations. Here, $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, and the virtual distillation rate is achieved with LOCC already. (c) Distillation of the noisy magic state $\rho_p = pT + (1-p)I_2/2$ under stabilizer operations. The magic state T is defined as $T = \frac{1}{2}(I + [X + Y]/\sqrt{2})$, where X and Y are Pauli matrices. The virtual distillation rates for (b) and (c) are obtained with $m = 1$ as in Eq. (3). We note that virtual distillation also outperforms probabilistic distillation for these examples. See [29] for details.

This is useful because $f_{\mathcal{O}}$ has been computed exactly in many relevant cases [15,19,21,45–47]. Such an equivalence will be possible whenever there exists a free "generalized twirling" operation [27] $\mathcal{T} \in \mathcal{O}$ of the form $\mathcal{T}(\rho) = \mathrm{Tr}[\psi^{\otimes m}\rho]\psi^{\otimes m} + \mathrm{Tr}[(I - \psi^{\otimes m})\rho]\sigma^{\star}$ for some $\sigma^{\star} \in \mathcal{F}$, which is true for many resource theories of practical interest, such as entanglement and magic theory for specific target states.

Theorem 2 (Theorem 2 in companion paper [29]).—Suppose a free generalized twirling operation exists. Then

$$C^{\varepsilon}(\rho, m) = \max\left\{\frac{2(1-\varepsilon)}{f_{\mathcal{O}}(\rho, m)} - 1, 1\right\}. \tag{8}$$

Reference [27] showed that the existence of a suitable resource nongenerating twirling operation is guaranteed when $F_{\mathcal{F}}(\psi^{\otimes m})^{-1}$ equals $1 + R_{\mathcal{F}}^s(\psi^{\otimes m})$. We thus have an alternative characterization of $C^{\varepsilon}(\rho, m)$ via the resource monotone $f_{\mathcal{O}}(\rho, m)$: whenever the condition $F_{\mathcal{F}}(\psi^{\otimes m})^{-1} = 1 + R_{\mathcal{F}}^s(\psi^{\otimes m})$ is satisfied, $C^{\varepsilon}(\rho, m)$ under resource nongenerating maps is given exactly by Eq. (8).

We note that while Theorem 1 provides an exact characterization of the virtual distillation overhead without the need for an explicit optimization over the allowed free operations, Theorem 2 is applicable also for general sets of free operations that are weaker than resource nongenerating operations. These results may therefore be applicable to complementary scenarios.

Surpassing conventional limitations.—Distillation in the conventional sense is constrained by many no-go theorems that restrict what transformations can be achieved in certain regimes. Consider zero-error distillation ($\varepsilon = 0$). In this case, conventional distillation protocols have significant limits: they cannot, for example, distil *any* pure states from states that are highly mixed (full or almost full rank, depending on the theory) [22,24,25], not even when many copies of input states are available, and not even probabilistically [22,28]. Virtual distillation suffers from no such no-go limitation: even full-rank states allow for distillation with a finite overhead cost. An even stronger limitation constrains the one-shot distillation from isotropic states ρ_p in theories such as quantum entanglement, coherence, or magic. No free operation can improve the fidelity of ρ_p with a maximally resourceful state, making distillation impossible from a single copy of ρ_p for all small values of ε [14,48]; virtual distillation allows one to surpass such restrictions.

Figure 2 illustrates this capability in contexts of coherence, entanglement, and magic. Here, we compare the virtual distillation rate with the conventional case for the three types of resources, demonstrating the lack of a noise threshold that makes the conventional rate diminish to zero, and a more continuous dependence of the virtual distillation rate on the smoothing parameter ε. In the case of coherence [7–9,19], we employ maximally incoherent operations (MIOs) and dephasing-covariant incoherent operations (DIOs) and apply Theorem 1 to determine the overhead for MIOs and DIOs. For entanglement [4–6,17], we determine the overhead using Theorem 2 for any arbitrary subset of separability-preserving (SEPP) operations that contains local operation and classical communication (LOCC). For magic [10–12], we examine stabilizer operations and obtain an exact formula for the overhead, as explained in the companion paper [29].

Applications.—We first consider an application in quantum teleportation, where entanglement is a critical resource. Given noisy entangled states, conventional resource distillation generally assumes multiqubit operations on many copies of the noisy state and the existence of quantum memory. Our virtual distillation protocol serves as a more experimentally friendly alternative without these challenging requirements. For example, consider the noisy state $\rho_{A'B}(p) = p\psi_{A'B} + (1-p)I_4/4$ with $|\psi\rangle_{A'B} = (|00\rangle_{A'B} + |11\rangle_{A'B})/\sqrt{2}$. For $p \geq 1/3$ (and assuming $\varepsilon = 0$), we could virtually distill $\rho_{A'B}$ as $\tilde{\Lambda}[\rho_{A'B}(p)] = [4/(1+3p)]\Lambda_+[\rho_{A'B}(p)] - [(3-3p)/(1+3p)]\Lambda_-[\rho_{A'B}(p)]$, where Λ_+ is the identity channel and $\Lambda_-(\rho) = \sigma_{AB} = (I - \psi_{AB})/3$. The distillation overhead is $C^0[\rho_{A'B}(p)] = \min\{(7-3p)/(1+3p), 3\}$. The cost remains a reasonable constant even if we consider virtual distillation to multiple noisy entangled states, a relevant case in distributed quantum computing [49].

Another setting to which our approach immediately applies is fault-tolerant quantum computation, where it is very natural to consider classical outcomes of quantum algorithms. It is then important to understand optimal ways to distill the underlying resource of magic [11,12]. Multiple proposals have been made regarding a combination of quantum error mitigation and error correction methods [50–52]. Our framework encompasses this strategy for magic state distillation as an application of the general approach of virtual resource distillation, in particular allowing us to extend the results in Ref. [50] to more general regimes [29].

Discussion.—We introduced virtual resource distillation, an extended experimentally friendly framework of resource distillation that integrates classical linear postprocessing into free operations. In doing so, we provided a fundamental tool for distilling the resources needed for quantum advantages in noisy current and near-future quantum technology. The results are general and applicable to many specific resource theories, such as coherence, entanglement, magic of quantum states, quantum communication, uniformity, and athermality, as well as any other resource theory where the target state is pure. We also detail the much more general setting of virtual resource distillation of quantum channels and combs in the companion paper [29], enabling virtual distillation of non-Markovianity and quantum memory. This work considers discrete-variable resource theories, and an extension to continuous variables [53,54] would also be interesting.

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

[1] M. Horodecki and J. Oppenheim, (Quantumness in the context of) resource theories, Int. J. Mod. Phys. B **27**, 1345019 (2013).

[2] B. Coecke, T. Fritz, and R. W. Spekkens, A mathematical theory of resources, Inf. Comput. **250**, 59 (2016).

[3] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. **91**, 025001 (2019).

[4] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, Phys. Rev. A **53**, 2046 (1996).

[5] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Quantifying entanglement, Phys. Rev. Lett. **78**, 2275 (1997).

[6] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[7] J. Aberg, Quantifying superposition, arXiv:quant-ph/0612146.

[8] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying coherence, Phys. Rev. Lett. **113**, 140401 (2014).

[9] A. Streltsov, G. Adesso, and M. B. Plenio, Colloquium: Quantum coherence as a resource, Rev. Mod. Phys. **89**, 041003 (2017).

[10] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, Negative quasi-probability as a resource for quantum computation, New J. Phys. **14**, 113011 (2012).

[11] V. Veitch, S. H. Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer quantum computation, New J. Phys. **16**, 013009 (2014).

[12] M. Howard and E. Campbell, Application of a resource theory for magic states to fault-tolerant quantum computing, Phys. Rev. Lett. **118**, 090501 (2017).

[13] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, Phys. Rev. A **53**, 2046 (1996).

[14] A. Kent, Entangled mixed states and local purification, Phys. Rev. Lett. **81**, 2839 (1998).

[15] E. M. Rains, A semidefinite program for distillable entanglement, IEEE Trans. Inf. Theory **47**, 2921 (2001).

[16] F. G. Brandao and M. B. Plenio, Entanglement theory and the second law of thermodynamics, Nat. Phys. **4**, 873 (2008).

[17] F. G. S. L. Brandao and N. Datta, One-shot rates for entanglement manipulation under non-entangling maps, IEEE Trans. Inf. Theory **57**, 1754 (2011).

[18] Q. Zhao, Y. Liu, X. Yuan, E. Chitambar, and X. Ma, One-shot coherence dilution, Phys. Rev. Lett. **120**, 070403 (2018).

[19] B. Regula, K. Fang, X. Wang, and G. Adesso, One-shot coherence distillation, Phys. Rev. Lett. **121**, 010401 (2018).

[20] Z.-W. Liu, K. Bu, and R. Takagi, One-shot operational quantum resource theory, Phys. Rev. Lett. **123**, 020401 (2019).

[21] 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).

[22] K. Fang and Z.-W. Liu, No-go theorems for quantum resource purification, Phys. Rev. Lett. **125**, 060405 (2020).

[23] M. K. Vijayan, E. Chitambar, and M.-H. Hsieh, Simple bounds for one-shot pure-state distillation in general resource theories, Phys. Rev. A **102**, 052403 (2020).

[24] B. Regula and R. Takagi, Fundamental limitations on distillation of quantum channel resources, Nat. Commun. **12**, 4411 (2021).

[25] K. Fang and Z.-W. Liu, No-go theorems for quantum resource purification: New approach and channel theory, PRX Quantum **3**, 010337 (2022).

[26] B. Regula and R. Takagi, One-shot manipulation of dynamical quantum resources, Phys. Rev. Lett. **127**, 060402 (2021).

[27] R. Takagi, B. Regula, and M. M. Wilde, One-shot yield-cost relations in general quantum resource theories, PRX Quantum **3**, 010348 (2022).

[28] B. Regula, Probabilistic transformations of quantum resources, Phys. Rev. Lett. **128**, 110505 (2022).

[29] R. Takagi, X. Yuan, B. Regula, and M. Gu, companion paper, Virtual quantum resource distillation: General framework and applications, Phys. Rev. A **109**, 022403 (2024).

[30] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating outcome probabilities of quantum circuits using quasiprobabilities, Phys. Rev. Lett. **115**, 070501 (2015).

[31] J. R. Seddon, B. Regula, H. Pashayan, Y. Ouyang, and E. T. Campbell, Quantifying quantum speedups: Improved classical simulation from tighter magic monotones, PRX Quantum **2**, 010345 (2021).

[32] Y. Yang, B.-N. Lu, and Y. Li, Accelerated quantum Monte Carlo with mitigated error on noisy quantum computer, PRX Quantum **2**, 040361 (2021).

[33] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, Simulating large quantum circuits on a small quantum computer, Phys. Rev. Lett. **125**, 150504 (2020).

[34] L. Brenner, C. Piveteau, and D. Sutter, Optimal wire cutting with classical communication, arXiv:2302.03366.

[35] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[36] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, J. Phys. Soc. Jpn. **90**, 032001 (2021).

[37] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[38] F. Buscemi, M. Dall'Arno, M. Ozawa, and V. Vedral, Direct observation of any two-point quantum correlation function, arXiv:1312.4240.

[39] J. Jiang, K. Wang, and X. Wang, Physical implementability of linear maps and its application in error mitigation, Quantum **5**, 600 (2021).

[40] 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).

[41] Specifically, according to the Hoeffding inequality [42], $\mathcal{O}[\log(1/\delta)/\beta^2]$ and $\mathcal{O}[C^2 \log(1/\delta)/\beta^2]$ number of samples can estimate $\text{Tr}[M\psi^{\otimes m}]$ and $\text{Tr}[M\tilde{\Lambda}(\rho)]$ to achieve an accuracy $\beta \geq 0$ with a failure probability less than $\delta \geq 0$.

[42] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. **58**, 13 (1963).

[43] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).

[44] L. Vandenberghe and S. Boyd, Semidefinite programming, SIAM Rev. **38**, 49 (1996).

[45] F. G. S. L. Brandão and M. B. Plenio, A reversible theory of entanglement and its relation to the second law, Commun. Math. Phys. **295**, 829 (2010).

[46] 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).

[47] Q. Zhao, Y. Liu, X. Yuan, E. Chitambar, and A. Winter, One-shot coherence distillation: Towards completing the picture, IEEE Trans. Inf. Theory **65**, 6441 (2019).

[48] B. Regula, Tight constraints on probabilistic convertibility of quantum states, Quantum **6**, 817 (2022).

[49] When $\rho_{A'B}$ is a separable state, this represents the cost of classical simulation. An intuitive yet nonoptimal simulation protocol would be using state tomography for Alice and sending the classical density matrix information to Bob. Our result instead gives an optimal simulation protocol. It is not hard to see that the overhead would increase consider multi-round quantum teleportation between different parties, which reflects the necessity of using entangled states. For example, consider sequential teleportation from party 0 to party n, the overhead to simulate perfect teleportation using separable states would be 3^n. On the other hand, suppose nearly entangled states $\rho(\varepsilon) = (1 - \varepsilon)\psi_{A'B} + \varepsilon I_4/4$ are shared (with small ε), the overhead for simulating n sequential rounds of teleportation is $[1 + 6\varepsilon/(4 - 3\varepsilon)]^n \approx e^{6n\varepsilon/(4-3\varepsilon)}$. Therefore, whenever $n\varepsilon = \mathcal{O}(1)$, the overhead is still a constant even for large n. For example, suppose $n = 50$ and $\varepsilon = 1\%$, the overhead is about $e^3 \approx 20$.

[50] M. Lostaglio and A. Ciani, Error mitigation and quantum-assisted simulation in the error corrected regime, Phys. Rev. Lett. **127**, 200506 (2021).

[51] 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).

[52] 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).

[53] B. Yadin, F. C. Binder, J. Thompson, V. Narasimhachar, M. Gu, and M. S. Kim, Operational resource theory of continuous-variable nonclassicality, Phys. Rev. X **8**, 041038 (2018).

[54] B. Regula, L. Lami, G. Ferrari, and R. Takagi, Operational quantification of continuous-variable quantum resources, Phys. Rev. Lett. **126**, 110403 (2021).