No-Go Theorems for Quantum Resource Distillation

Kun Fang[1,*] and Zi-Wen Liu[2,†]

[1]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, CB3 0WA, UK*

[2]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*

(Dated: September 25, 2019)

Distillation, which generally means converting noisy states into better approximations of some valuable pure states by low-cost operations, is a fundamental type of resource trading protocol that plays major roles in quantum computation and information processing. In this work, we establish a series of fundamental limitations to the efficiency of distillation with finite resources which apply to any reasonable resource theory, using an information-theoretic tool called the quantum hypothesis testing relative entropy. We show that, for highly generic noisy inputs and any target pure resource state, there is an insuperable limitation to the accuracy (lower bound on the error) of any distillation protocol, even probabilistic ones. More explicitly, we prove an "uncertainty relation" between the accuracy and success probability, which indicates a forbidden regime that no protocol with the corresponding parameters can exist. We use these no-go theorems to derive explicit bounds which imply that the total overhead i.e. the number of noisy inputs required to distill a pure resource state up to small error ϵ scale at least as $\Omega(\log(1/\epsilon))$. In particular, we establish specific lower bounds on the cost of magic state distillation, a task crucial to fault-tolerant quantum computation (which address an open problem raised by Hastings and Haah [PRL 120, 050504 (2018)]). Finally, we showcase a basic no-go result for simulating unitary resource channels to demonstrate that the analogous understandings may extend to channel resource theories.

I. INTRODUCTION

The field of quantum information takes a pragmatic approach to the mysteries of quantum mechanics, seeking to obtain a better understanding of which information processing tasks can or cannot be accomplished in this new arena. A landmark result is the no-cloning theorem [1, 2], which states that there is no universal quantum operation duplicating an unknown quantum state. The discovery of this no-go result represents a key advance in our understanding of quantum mechanics, and has profound impact on the development of quantum information and computation. On the one hand, the result prevents several important extensions of classical results such as classical error correction, which for some time was thought to be a fatal limitation for practical quantum computing. But on the other hand, it also stimulates the invention of quantum error correcting codes [3, 4], and lays the foundation for plenty of other major quantum applications, such as the BB84 protocol in quantum cryptography [5].

This work focuses on another fundamental task in quantum information processing that plays particularly important roles in practical scenarios, and reveals its limitations through a host of no-go theorems which prohibit the existence of any allowed operation achieving certain goals, in analogy to the no-cloning theorem. This task is often called *resource distillation*, which aims at "purifying" the resource encoded in noisy quantum systems. In practice, quantum states prepared from experiments or stored in memories are unavoidably contaminated by various kinds of noises, making them unreliable for direct usage. Therefore, distillation comes into play as a standard procedure which helps us extract high-quality resources better suited for application. Most notably, the distillation of entanglement [6–8], coherence [9–11] and magic states [12] have been extensively studied as key subroutines in quantum communication and computation. Understanding the limitations and costs of distillation tasks, especially in the finite-copy (one-shot) setting, are thus of great theoretical and practical importance.

To address this problem in a simple and general manner, we shall use the language of resource theory (see [13] for an introduction of this framework): A state resource theory is defined by *free states* (in contrast to resource states) and *free operations*, with a *golden rule* that any free operation can only map a free state to another free state. This simple rule selects the largest possible set of free operations allowed in resource manipulation such as distillation, since any other operation can by definition create resource and thus trivializing the theory. Whether some resource state can be transformed into another via certain free operations is a fundamental type of problem in quantum information theory. Practical distillation tasks mostly concern the following case: Is it possible to convert a finite amount of mixed (noisy) states to pure resource states to arbitrary accuracy by free operations?

We first give a negative answer to the above question in general, manifesting that the production of any pure resource state with arbitrarily small error from a full-rank noisy state, however weak this target state is, is inherently prohibited by the golden rule embedded in any resource theory. More formally, we establish quantitative bounds which limit the achievable accuracy for free operations that work with a certain probability. It turns out that there is a non-trivial trade-off between the accuracy and success probability, akin to the uncertainty relations. The proofs follow from analyzing the peculiar properties of the hypothesis testing relative entropy monotone, a quantity known to characterize the efficiency of one-shot distillation in many cases [11, 14–20] but not studied in great depth. Using the above results, we find lower bounds on the

* kf383@cam.ac.uk
† zliu1@perimeterinstitute.ca

overhead of distillation, namely, the number of copies of a certain primitive state needed. As a particularly important application, we provide specific lower bounds for the cost of magic state distillation [12], which is estimated to be a dominant source of resource cost in current fault-tolerant quantum computation schemes. The consequent limitations to the common distillation protocols based on quantum error correction are discussed in relation to key progresses in the search for better codes [21–24]. Lastly, we prove a no-go theorem for the simulation of unitary resource channels, which is analogous to the distillation of pure states, in accordance with the recent interest in extending conventional resource theory approaches for quantum states to quantum channels (see e.g. [25–40]).

II. NO-GO THEOREMS FOR DISTILLATION

In this section, we establish fundamental limits to the achievable accuracy of any such distillation procedure in terms of certain properties of the input and output states. Our arguments make use of the following information-theoretic quantities. For any two quantum state ρ and σ, their *quantum hypothesis testing relative entropy* is defined as $D_H^\epsilon(\rho\|\sigma) \equiv -\log\beta_\epsilon(\rho\|\sigma)$, where $\beta_\epsilon(\rho\|\sigma) = \min\left\{\operatorname{Tr} M\sigma \,|\, \operatorname{Tr}\rho M \geq 1-\epsilon, 0 \leq M \leq \mathbb{1}\right\}$. The *fidelity* between two quantum states ρ and σ is defined as $F(\rho,\sigma) \equiv \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$. Denote the maximum overlap between ρ and the set of free states $\mathcal{F}$ as $f_\rho \equiv \max_{\omega\in\mathcal{F}} F(\rho,\omega)$.

Let us first consider the deterministic distillation scenario, where the transformation process is required to succeed with certainty and thus represented by a free CPTP map.

Theorem 1 (No-go, deterministic) *Given any full rank primitive state $\rho \notin \mathcal{F}$ and any pure target resource state $\psi \notin \mathcal{F}$, there is no free operation transforming ρ to ψ within $\varepsilon(\rho,\psi)$ error, where $\varepsilon(\rho,\psi) = \lambda_{\min}(\rho)\left(1-f_\psi\right) > 0$ and $\lambda_{\min}(\rho)$ is the smallest non-zero eigenvalue of ρ.*

The proof of this theorem will use the following continuity bound of the hypothesis testing relative entropy. We delegate its detailed proof to the Supplemental Material.

Lemma 2 *For any full rank state ρ and any quantum state σ, their quantum hypothesis testing relative entropy is continuous around $\epsilon = 0$. That is, for any $0 \leq \epsilon < \lambda_{\min}(\rho)$ where $\lambda_{\min}(\rho)$ is the smallest eigenvalue of ρ, it holds that*

$$0 \leq D_H^\epsilon(\rho\|\sigma) \leq \log\frac{\lambda_{\min}(\rho)}{\lambda_{\min}(\rho)-\epsilon}.$$

Proof of Theorem 1: The proof is given by contradiction. Suppose there is a free operation $\mathcal{E}$ such that $\mathcal{E}(\rho) = \sigma$ and $\operatorname{Tr}\psi\sigma \geq 1-\epsilon$ with $\epsilon < \varepsilon(\rho,\psi)$. Consider the quantum test $\{\psi, \mathbb{1}-\psi\}$, we have $D_H^\epsilon(\sigma\|\omega) \geq -\log\operatorname{Tr}\psi\omega$ for all $\omega \in \mathcal{F}$. Then the following chain of inequalities holds

$$\begin{aligned}
-\log f_\psi &\leq \min_{\omega\in\mathcal{F}} D_H^\epsilon(\sigma\|\omega) && (1)\\
&= \min_{\omega\in\mathcal{F}} D_H^\epsilon(\mathcal{E}(\rho)\|\omega) && (2)\\
&\leq \min_{\omega\in\mathcal{F}} D_H^\epsilon(\mathcal{E}(\rho)\|\mathcal{E}(\omega)) && (3)\\
&\leq \min_{\omega\in\mathcal{F}} D_H^\epsilon(\rho\|\omega) && (4)\\
&\leq \log\left(\lambda_{\min}(\rho)/(\lambda_{\min}(\rho)-\epsilon)\right) && (5)
\end{aligned}$$

where the second line follows from the assumption that $\mathcal{E}(\rho) = \sigma$, the third line follows since it is minimized over a smaller set due to $\mathcal{E}(\omega) \in \mathcal{F}$, $\forall\omega \in \mathcal{F}$, the fourth line follows from the data-processing inequality of D_H^ϵ [41], the last line follows from the continuity bound in Lemma 2 (applicability guaranteed by the assumption $\epsilon < \varepsilon(\rho,\psi)$). A simple calculation gives us $\varepsilon \geq \lambda_{\min}(\rho)(1-f_\psi)$, which forms a contradiction with our assumption. □

Note that $\lambda_{\min}$ encodes key features of the noise. Again consider depolarizing noise as a basic example, then $\lambda_{\min}$ of a contaminated pure state is proportional to the strength of the noise.

We now move on to a more complicated but practically important setting, in which the distillation process only succeeds with a certain probability. Consider a generalization of the set of free operations $\boldsymbol{O}$ to the class $\boldsymbol{O}_{\text{sub}} \equiv \{\mathcal{L} \,|\, \forall\rho \in \mathcal{F}, \exists\, t \geq 0, \sigma \in \mathcal{F}, \text{ s.t. } \mathcal{L}(\rho) = t\cdot\sigma\}$ of subnormalized quantum operations, that is, completely positive and trace-nonincreasing maps. A free probabilistic protocol transforms ρ to γ with probability p and fidelity $1-\epsilon$ is given by a quantum operation $\mathcal{E}_{A\to FB}$ such that $\mathcal{E}_{A\to FB}(\rho_A) = |0\rangle\langle 0|_F \otimes \mathcal{L}_{A\to B}(\rho_A) + |1\rangle\langle 1|_F \otimes \mathcal{G}_{A\to B}(\rho_A)$, $\mathcal{L}_{A\to B}(\rho_A) = p\sigma_B$ and $F(\sigma,\gamma) \geq 1-\epsilon$, where $\mathcal{L}, \mathcal{G} \in \boldsymbol{O}_{\text{sub}}$ are free sub-operations, and F is a flag register indicating whether the transformation succeeds or not.

Theorem 3 (No-go, probabilistic) *Given any full rank primitive state $\rho \notin \mathcal{F}$ and any pure target resource state $\psi \notin \mathcal{F}$, there is a nontrivial upper bound on the accuracy of distillation that no free probabilistic protocol can exceed, which decreases as the success probability grows. More explicitly, the following trade-off relation between the transformation fidelity $1-\epsilon$ and the success probability p must hold:*

$$\frac{\epsilon}{p} \geq \frac{\varepsilon(\rho,\psi)}{1+R(\rho)} = \frac{\lambda_{\min}(\rho)(1-f_\psi)}{1+R(\rho)}. \tag{6}$$

The proof of this theorem will use another two Lemmas (proofs in the Supplemental Material).

Lemma 4 *For any two flagged quantum states $\rho_i = p_i|0\rangle\langle 0|\otimes\sigma_i + (1-p_i)|1\rangle\langle 1|\otimes\omega_i$ with $i \in \{1,2\}$ and $p_i \in [0,1]$, it holds that $\beta_\epsilon(\rho_1\|\rho_2) \leq p_2\beta_\epsilon(\sigma_1\|\sigma_2) + (1-p_2)\beta_\epsilon(\omega_1\|\omega_2)$.*

Lemma 5 *For any linear suboperation $\mathcal{L}$, there exists a free state $\omega \in \mathcal{F}$ such that $\operatorname{Tr}\mathcal{L}(\omega) \geq (1+R(\rho))^{-1}\operatorname{Tr}\mathcal{L}(\rho)$, where $R(\rho) \equiv \min\{s|\exists\,\sigma, s \geq 0, \text{ s.t. } (\rho+s\sigma)/(1+s) \in \mathcal{F}\}$ is the (generalized) robustness of state ρ.*

Proof of Theorem 3: Suppose there is a probabilistic protocol $\mathcal{E}_{A\to FB}(\rho_A) = |0\rangle\langle 0|_F \otimes \mathcal{L}_{A\to B}(\rho_A) + |1\rangle\langle 1|_F \otimes \mathcal{G}_{A\to B}(\rho_A)$ such that $\mathcal{L}_{A\to B}(\rho_A) = p\sigma_B$ and $\mathrm{Tr}\,\sigma\psi \geq 1-\epsilon$ with $\epsilon < p(1+R(\rho))^{-1}\varepsilon(\rho,\psi)$. Based on Lemma 5, let us consider a free state ω_1 such that $\mathrm{Tr}\,\mathcal{L}(\omega_1) \geq (1+R(\rho))^{-1}\,\mathrm{Tr}\,\mathcal{L}(\rho)$ and take $\omega_2 = \mathcal{L}(\omega_1)/\mathrm{Tr}\,\mathcal{L}(\omega_1)$. Then we have

$$f_\psi = \max_{\omega\in\mathcal{F}} \mathrm{Tr}\,\omega\psi \geq \mathrm{Tr}\,\omega_2\psi \geq \beta_\epsilon(\sigma\|\omega_2), \tag{7}$$

where the last inequality follows by considering the quantum test $\{\psi, \mathbb{1}-\psi\}$. On the other hand, we have

$$\beta_\epsilon(\sigma\|\omega_2) = \beta_\epsilon\left(\frac{\mathcal{L}(\rho)}{\mathrm{Tr}\,\mathcal{L}(\rho)}\middle\|\frac{\mathcal{L}(\omega_1)}{\mathrm{Tr}\,\mathcal{L}(\omega_1)}\right) \tag{8}$$

$$\geq \frac{\beta_\epsilon(\mathcal{E}(\rho)\|\mathcal{E}(\omega_1)) - (1-\mathrm{Tr}\,\mathcal{L}(\omega_1))}{\mathrm{Tr}\,\mathcal{L}(\omega_1)} \tag{9}$$

$$\geq \frac{\beta_\epsilon(\rho\|\omega_1) - (1-\mathrm{Tr}\,\mathcal{L}(\omega_1))}{\mathrm{Tr}\,\mathcal{L}(\omega_1)} \tag{10}$$

$$\geq \frac{(\lambda_{\min}(\rho)-\epsilon)/\lambda_{\min}(\rho) - (1-\mathrm{Tr}\,\mathcal{L}(\omega_1))}{\mathrm{Tr}\,\mathcal{L}(\omega_1)}, \tag{11}$$

where the second line follows from Lemma 4, the third line follows from the data-processing inequality, and the last line follows from the continuity bound in Lemma 2 (applicability guaranteed by the assumption $\epsilon < p(1+R(\rho))^{-1}\varepsilon(\rho,\psi) \leq \varepsilon(\rho,\psi)$) and the assumption that ρ is full rank. Combining (11) with (7), we have $\epsilon \geq \mathrm{Tr}\,\mathcal{L}(\omega_1)\varepsilon(\rho,\psi)$. Recall that ω_1 is defined as a free state such that $\mathrm{Tr}\,\mathcal{L}(\omega_1) \geq (1+R(\rho))^{-1}\,\mathrm{Tr}\,\mathcal{L}(\rho) = (1+R(\rho))^{-1}p$, we obtain $\epsilon \geq p(1+R(\rho))^{-1}\varepsilon(\rho,\psi)$, which forms a contradiction with our assumption. □

Note that a slightly weaker version of Theorem 1 for deterministic distillation is recovered by letting $p = 1$. It is also possible to get rid of the $(1+R(\rho))^{-1}$ factor and obtain a stronger bound that covers Theorem 1 under certain restrictions. For example, suppose the theory admits a resource destroying channel [42] Λ, and the allowed free suboperations are those commuting with Λ (such as dephasing-covariant incoherent suboperations for coherence theory [10]). Then for any free suboperation $\mathcal{L}$, it holds that $\mathrm{Tr}\,\mathcal{L}(\rho) = \mathrm{Tr}\,\Lambda\circ\mathcal{L}(\rho) = \mathrm{Tr}\,\mathcal{L}\circ\Lambda(\rho)$, which indicates that there always exists a free state $\omega = \Lambda(\rho)$ such that $\mathrm{Tr}\,\mathcal{L}(\omega) = p$, and therefore the bound reduces to $\epsilon/p \geq \varepsilon(\rho,\psi)$.

In particular, by letting $\epsilon = 0$ we directly obtain the following no-go result for ideal distillation:

Corollary 6 *It is impossible to exactly distill a pure target resource state from a full rank primitive state, even probabilistically.*

Theorems 1 and 3 establish an "uncertainty relation" between the accuracy and success probability of distillation characterized by a regime of $\{\epsilon, p\}$ that is not achievable by any free protocol, as illustrated in Fig. 1. We stress that this phenomenon occurs in any well-defined quantum resource theory (as long as not all pure states are free), and therefore implies universal limitations to distillation of all kinds of resources which are useful in different scenarios.

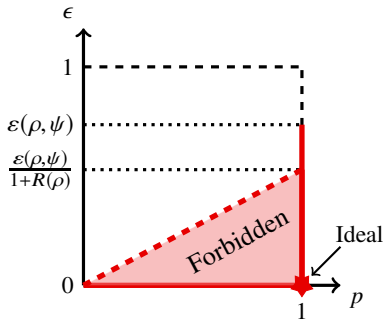

FIG. 1. Interplay between the transformation error ϵ and success probability p. The lower right corner represents the most ideal scenario where ϵ is small and p is large. The red region and solid lines represent the forbidden regime such that no distillation protocol with the corresponding parameters can exist. $\varepsilon(\rho,\psi) = \lambda_{\min}(\rho)(1-f_\psi)$.

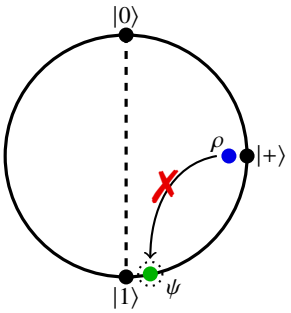

FIG. 2. A qubit coherence theory example illustrated using the Bloch sphere. Here ρ is a mixed state close to the maximally coherent state $|+\rangle$, and ψ is a pure state close to basis state $|1\rangle$. Our no-go theorems indicate that an arbitrarily accurate probabilistic transformation from ρ to ψ is impossible.

Remarkably, the noisy primitive state could "look" much more resourceful (or more precisely, takes much greater value in terms of other resource monotones) than the target state, yet the possibility of an arbitrarily accurate probabilistic transformation is ruled out. An illustrative toy example in terms of qubit coherence theory is given in Fig. 2, where ρ is a slightly noisy version of the maximally coherent state $|+\rangle$ and ψ is a pure target state very close to basis state $|1\rangle$. It is clear that commonly used coherence measures such as the minimum relative entropy or trace distance to incoherent states assign much greater value to ρ than to ψ, which may lead to the false intuition that ρ can be used to trade for ψ. This highlights the special role of the hypothesis testing relative entropy monotone among all resource measures, especially in characterizing distillation tasks.

III. RESOURCE OVERHEAD

The setting of distillation tasks in practice is usually the following: One has access to a certain primitive resource state; The goal is to distill certain "standard" pure resource states to

some desired accuracy while consuming as few copies of the primitive state as possible. Therefore, the amount of primitive states needed to accomplish the desired distillation, namely the *resource cost* or *overhead*, is a key figure of merit for distillation protocols. To state the most general result, we consider error on the entire output state (which could be a collection of unit states) for now. As we now show, our no-go theorems induce fundamental lower bounds on the total overhead of distillation, when the target error is small.

Theorem 7 (Total overhead) *Consider the task of distilling some pure target resource state ψ with error at most ϵ, from n copies of primitive state $\hat{\rho}$. For sufficiently small ϵ and any full rank $\hat{\rho}$, there does not exist any free operation that accomplishes the task if the following is not satisfied:*

$$n \geq \log_{\frac{1}{\lambda_{\min}(\hat{\rho})}} \frac{1-f_\psi}{\epsilon}. \tag{12}$$

Furthermore, for sufficiently small ϵ and any full rank state $\hat{\rho}'$, there does not even exist any probabilistic protocol with success probability p that accomplishes the task if the following is not satisfied:

$$n \geq \log_{\frac{1+R(\hat{\rho}')}{\lambda_{\min}(\hat{\rho}')}} \frac{(1-f_\psi)p}{\epsilon}. \tag{13}$$

Proof of Theorem 7: Follows by transforming the no-go theorems with input state $\hat{\rho}^{\otimes n}$. Notice that $\lambda_{\min}(\hat{\rho}^{\otimes n}) = \lambda_{\min}(\hat{\rho})^n$. For the deterministic case, Theorem 1 implies that for any full rank $\hat{\rho}$ and $\epsilon < \varepsilon(\hat{\rho}^{\otimes n}, \psi)$, we have

$$\epsilon \geq \lambda_{\min}(\hat{\rho}^{\otimes n})(1-f_\psi) = \lambda_{\min}(\hat{\rho})^n(1-f_\psi). \tag{14}$$

This directly translates to Eq. (12).

To obtain the claimed bound for the probabilistic case, notice the following: By the definition of $R(\hat{\rho})$, there exists some state τ such that $\hat{\rho} + R(\hat{\rho})\tau = (1+R(\hat{\rho}))\omega$ where $\omega \in \mathcal{F}$. By expanding this equation, we obtain

$$\omega^{\otimes n} = \frac{1}{(1+R(\hat{\rho}))^n}\hat{\rho}^{\otimes n} + \frac{(1+R(\hat{\rho}))^n - 1}{(1+R(\hat{\rho}))^n}\tau', \tag{15}$$

where $\omega^{\otimes n} \in \mathcal{F}$ axiomatically [43], and τ' is a density operator. Therefore, $1+R(\hat{\rho}^{\otimes n}) \leq (1+R(\hat{\rho}))^n$. Now, by Theorem 3, for any full rank state $\hat{\rho}'$ and $\epsilon < p(1+R(\hat{\rho}'^{\otimes n}))^{-1}\varepsilon(\hat{\rho}'^{\otimes n}, \psi)$, we have

$$\epsilon/p \geq \frac{\lambda_{\min}(\hat{\rho}'^{\otimes n})(1-f_\psi)}{1+R(\hat{\rho}'^{\otimes n})} \geq \frac{\lambda_{\min}(\hat{\rho}')^n(1-f_\psi)}{(1+R(\hat{\rho}'))^n}. \tag{16}$$

This directly translates to Eq. (13). □

The above theorem indicates that for distillation protocols which succeed with at least a constant probability (that does not vanish when reducing the target ϵ), the total overhead must scale at least as $\Omega(\log(1/\epsilon))$ as $\epsilon \to 0$.

IV. MAGIC STATE DISTILLATION

Optimizing the resource costs for implementing a universal set of logical gates fault-tolerantly is a crucial problem for realizing scalable quantum computation [44–46]. A leading scheme for fault-tolerance is based on the so-called *magic states* [12]. Here the so-called Clifford operations are considered "cheap" since they admit fault-tolerant implementations thanks to stabilizer codes [44, 47–49], but their computational power is very limited—due to the celebrated Gottesman-Knill theorem, they can even be efficiently simulated by classical computers [44, 50, 51]. To achieve universal quantum computation, one needs non-Clifford gates such as $T = \mathrm{diag}(1, e^{i\pi/4})$. A standard approach is to distill high-quality pure magic state $|T\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$ from sufficiently many dirty magic states, which can be directly used to emulate a logical T-gate via state injection [52]. Unfortunately, the production of any magic state requires non-Clifford operations, and to achieve sufficiently low error the amount of these resources needed is substantial, constituting a major source of the cost of such fault-tolerance schemes. Therefore, understanding the ultimate limitations to the efficiency of this distillation component is crucial.

We now address this problem by tailoring our general results to the practical magic state distillation settings and establish explicit lower bounds on the overhead. (Note that the resource theory ideas have advanced our understanding of magic states and quantum computation in various other ways [18, 28, 29, 53–55].) Known protocols for magic state distillation are commonly based on concatenating error correction subroutines using stabilizer codes to (probabilistically) produce an output with sufficiently high quality upon passing the syndrome measurements. Note that such procedures alone cannot create magic states, following the golden rule. Depending on the codes being used, the output could take the form of a large state with each marginal sufficiently close to a unit target state, in which case we are also interested in the *average overhead*, i.e. the total overhead divided by the number of marginals. Here we only showcase the T-state result, but the bounds for other useful magic states (see e.g. [24]) can be similarly obtained by plugging in corresponding parameters.

Theorem 8 *Consider the following general form of magic state distillation task: given n copies of full rank primitive magic states $\hat{\rho}$, output an m-qubit state τ such that $\mathrm{Tr}\,\tau_i T = \langle T|\tau_i|T\rangle \geq 1-\epsilon, \forall i = 1, ..., m$ where $\tau_i = \mathrm{Tr}_{\bar{i}}\tau$ is the i-th qubit. Then for sufficiently small ϵ, the average overhead of any probabilistic protocol that succeeds with probability p must obey*

$$C := n/m \geq \frac{1}{m}\log_{\frac{1+R(\hat{\rho})}{\lambda_{\min}(\hat{\rho})}} \frac{((4-2\sqrt{2})^m - 1)p}{(4-2\sqrt{2})^m m\epsilon}. \tag{17}$$

Proof of Theorem 8: By applying the union bound, we have $\langle T^{\otimes m}|\tau|T^{\otimes m}\rangle \geq 1 - m\epsilon$. Also notice that $f_{T^{\otimes m}} = (4-2\sqrt{2})^{-m}$ [19, 56–58]. By plugging everything into Eq. (13) we obtain the claimed bound. □

In the analyses of magic state distillation protocols, one is particularly interested in the exponent γ in the asymptotic average overhead $O(\log^\gamma(1/\epsilon))$. A subtlety of our lower bound is that the output size m could depend on the target ϵ for specific protocols. Thus, to understand the scaling, one needs

to take into account the behavior of m as well. There are two key implications of our bound to code-based distillation protocols. Assuming non-vanishing success probability (the passing probability of deeper rounds of concatenation converges sufficiently fast to one), then: (i) It is impossible to construct a protocol with sublogarithmic average overhead ($\gamma < 1$) with any $[n, k, d]$ code such that $k \leq d$. This can be seen by plugging $m = k^\nu$ and $\log(1/\epsilon) \sim d^\nu$ into Eq. (17). This in particular implies a $\gamma \geq 1$ bound for $k = 1$ codes, in response to a problem raised in [22]. Note that the best known such codes allow $\gamma \to 2$ [23, 24], so there is still a gap. (ii) Any $\gamma < 1$ protocol must have a scale (size of input and output) that diverges under concatenation. It was actually conjectured that no codes allowing $\gamma < 1$ exist [21], but such codes were recently found [22] (see also [59]). There, indeed, the codes employed have $k > d$. So although the average overhead of such a protocol is considered low, its output size must grow rapidly as we reduce ϵ, which could blow up the overall cost. Our results indicate that this is inevitable.

V. CHANNEL THEORY

As a basic result for the channel resource theory setting [31, 39, 40], we now show that there does not exist any free superchannel that perfectly transforms a noisy channel into a unitary resource channel, which is analogous to pure resource states. Here, again, the only constraint on the allowed superchannels is the golden rule of no resource generation.

Theorem 9 (Channel no-go) *Given any primitive channel $\mathcal{N}_{A\to B} \notin \boldsymbol{O}(A \to B)$ with free component (i.e. can be written as $\mathcal{N} = p\mathcal{E} + (1-p)\mathcal{M}, p \in (0,1)$ where $\mathcal{E} \in \boldsymbol{O}(A \to B)$ and $\mathcal{M}$ is any channel), and any target unitary resource channel $\mathcal{U}_{C\to D} \notin \boldsymbol{O}(C \to D)$, there is no free superchannel Π transforming $\mathcal{N}_{A\to B}$ to $\mathcal{U}_{C\to D}$.*

Proof can be found in the Supplemental Material. A straightforward implication of this result is that the zero-error quantum capacity of a generic quantum channel, such as the quantum depolarizing channel, is zero. More comprehensive studies of the channel setting will be left for future work.

VI. CONCLUDING REMARKS

This work reveals universal limitations to the accuracy and efficiency of quantum resource distillation tasks through investigating the properties of the hypothesis testing relative entropy. Our results do not rely on any particular structure of the resource theory, and only require very mild assumptions on the primitive state that hold in most practically interesting cases. All bounds depend only on very few parameters that concisely encode relevant properties of the noise, the target state, and the resource theory, and are thus easy to analyze. Just as how the no-cloning theorem essentially stem from linearity, a fundamental feature of quantum mechanics, our results are succinct consequences of the golden rule of quantum resource theory at bottom. An important future work is to investigate to what extent the bounds on the distillation fidelity and asymptotic overhead can be approached. It would also be interesting to further understand the approximate and probabilistic regimes of unitary channel simulation, due to its connections to the fields of quantum Shannon theory, gate and circuit synthesis etc.

ACKNOWLEDGMENTS

We thank Daniel Gottesman, Jeongwan Haah, Sirui Lu, Bartosz Regula, Ryuji Takagi for helpful discussions and feedback. We also thank Julio I. de Vicente for pointing out a gap in the proof of Lemma 2 in the previous version of this manuscript. K.F. is supported by the University of Cambridge Isaac Newton Trust Early Career grant RG74916. Z.-W.L. is supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation.

[1] W K Wootters and W H Zurek, "A single quantum cannot be cloned," Nature **299**, 802–803 (1982).
[2] D Dieks, "Communication by EPR devices," Physics Letters A **92**, 271–272 (1982).
[3] Peter W. Shor, "Scheme for reducing decoherence in quantum computer memory," Phys. Rev. A **52**, R2493–R2496 (1995).
[4] A. M. Steane, "Error correcting codes in quantum theory," Phys. Rev. Lett. **77**, 793–797 (1996).
[5] Charles H Bennett and Gilles Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *International Conference on Computers, Systems & Signal Processing, Bangalore, India, Dec 9-12, 1984* (1984) pp. 175–179.
[6] Charles H. Bennett, Herbert J. Bernstein, Sandu Popescu, and Benjamin Schumacher, "Concentrating partial entanglement by local operations," Phys. Rev. A **53**, 2046–2052 (1996).
[7] Charles H. Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A. Smolin, and William K. Wootters, "Purification of noisy entanglement and faithful teleportation via noisy channels," Phys. Rev. Lett. **76**, 722–725 (1996).
[8] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters, "Mixed-state entanglement and quantum error correction," Phys. Rev. A **54**, 3824–3851 (1996).
[9] Andreas Winter and Dong Yang, "Operational resource theory of coherence," Phys. Rev. Lett. **116**, 120404 (2016).
[10] Kun Fang, Xin Wang, Ludovico Lami, Bartosz Regula, and Gerardo Adesso, "Probabilistic Distillation of Quantum Coherence," Physical Review Letters **121**, 070404 (2018), 1804.09500.
[11] Bartosz Regula, Kun Fang, Xin Wang, and Gerardo Adesso, "One-shot coherence distillation," Phys. Rev. Lett. **121**, 010401

(2018).
[12] Sergey Bravyi and Alexei Kitaev, "Universal quantum computation with ideal clifford gates and noisy ancillas," Phys. Rev. A **71**, 022316 (2005).
[13] Eric Chitambar and Gilad Gour, "Quantum resource theories," Rev. Mod. Phys. **91**, 025001 (2019).
[14] F. G. S. L. Brandao and N. Datta, "One-shot rates for entanglement manipulation under non-entangling maps," IEEE Transactions on Information Theory **57**, 1754–1760 (2011).
[15] K. Fang, X. Wang, M. Tomamichel, and R. Duan, "Non-asymptotic entanglement distillation," IEEE Transactions on Information Theory , 1–1 (2019).
[16] Q. Zhao, Y. Liu, X. Yuan, E. Chitambar, and A. Winter, "One-shot coherence distillation: Towards completing the picture," IEEE Transactions on Information Theory , 1–1 (2019).
[17] Nicole Yunger Halpern and Joseph M. Renes, "Beyond heat baths: Generalized resource theories for small-scale thermodynamics," Phys. Rev. E **93**, 022126 (2016).
[18] Xin Wang, Mark M. Wilde, and Yuan Su, "Efficiently computable bounds for magic state distillation," arXiv e-prints , arXiv:1812.10145 (2018), arXiv:1812.10145 [quant-ph].
[19] Zi-Wen Liu, Kaifeng Bu, and Ryuji Takagi, "One-shot operational quantum resource theory," Phys. Rev. Lett. **123**, 020401 (2019).
[20] Bartosz Regula, Kun Fang, Xin Wang, and Mile Gu, "One-shot entanglement distillation beyond locc," arXiv preprint arXiv:1906.01648 (2019).
[21] Sergey Bravyi and Jeongwan Haah, "Magic-state distillation with low overhead," Phys. Rev. A **86**, 052329 (2012).
[22] Matthew B. Hastings and Jeongwan Haah, "Distillation with sublogarithmic overhead," Phys. Rev. Lett. **120**, 050504 (2018).
[23] Jeongwan Haah, Matthew B. Hastings, D. Poulin, and D. Wecker, "Magic state distillation with low space overhead and optimal asymptotic input count," Quantum **1**, 31 (2017).
[24] Jeongwan Haah and Matthew B. Hastings, "Codes and Protocols for Distilling T, controlled-S, and Toffoli Gates," Quantum **2**, 71 (2018).
[25] Khaled Ben Dana, María García Díaz, Mohamed Mejatty, and Andreas Winter, "Resource theory of coherence: Beyond states," Phys. Rev. A **95**, 062327 (2017).
[26] María García Díaz, Kun Fang, Xin Wang, Matteo Rosati, Michalis Skotiniotis, John Calsamiglia, and Andreas Winter, "Using and reusing coherence to realize quantum processes," Quantum **2**, 100 (2018).
[27] Quntao Zhuang, Peter W. Shor, and Jeffrey H. Shapiro, "Resource theory of non-gaussian operations," Phys. Rev. A **97**, 052317 (2018).
[28] James R. Seddon and Earl T. Campbell, "Quantifying magic for multi-qubit operations," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **475**, 20190251 (2019), https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.2019.0251.
[29] Xin Wang, Mark M. Wilde, and Yuan Su, "Quantifying the magic of quantum channels," arXiv e-prints , arXiv:1903.04483 (2019), arXiv:1903.04483 [quant-ph].
[30] Xin Wang and Mark M. Wilde, "Resource theory of asymmetric distinguishability for quantum channels," arXiv e-prints , arXiv:1907.06306 (2019), arXiv:1907.06306 [quant-ph].
[31] Gilad Gour, "Comparison of Quantum Channels with Superchannels," arXiv e-prints , arXiv:1808.02607 (2018), arXiv:1808.02607 [quant-ph].
[32] Kun Fang, Xin Wang, Marco Tomamichel, and Mario Berta, "Quantum channel simulation and the channel's smooth max-information," arXiv preprint arXiv:1807.05354 (2018).
[33] Mark M. Wilde, "Entanglement cost and quantum channel simulation," Phys. Rev. A **98**, 042338 (2018).
[34] Lu Li, Kaifeng Bu, and Zi-Wen Liu, "Quantifying the resource content of quantum channels: An operational approach," arXiv e-prints , arXiv:1812.02572 (2018), arXiv:1812.02572 [quant-ph].
[35] Thomas Theurer, Dario Egloff, Lijian Zhang, and Martin B. Plenio, "Quantifying operations with an application to coherence," Phys. Rev. Lett. **122**, 190405 (2019).
[36] Ryuji Takagi and Bartosz Regula, "General resource theories in quantum mechanics and beyond: operational characterization via discrimination tasks," arXiv e-prints , arXiv:1901.08127 (2019), arXiv:1901.08127 [quant-ph].
[37] Stefano Pirandola, Riccardo Laurenza, Carlo Ottaviani, and Leonardo Banchi, "Fundamental limits of repeaterless quantum communications," Nature communications **8**, 15043 (2017).
[38] Gilad Gour and Andreas Winter, "How to quantify a dynamical resource?" arXiv e-prints , arXiv:1906.03517 (2019), arXiv:1906.03517 [quant-ph].
[39] Yunchao Liu and Xiao Yuan, "Operational Resource Theory of Quantum Channels," arXiv e-prints , arXiv:1904.02680 (2019), arXiv:1904.02680 [quant-ph].
[40] Zi-Wen Liu and Andreas Winter, "Resource theories of quantum channels and the universal role of resource erasure," arXiv e-prints , arXiv:1904.04201 (2019), arXiv:1904.04201 [quant-ph].
[41] Ligong Wang and Renato Renner, "One-shot classical-quantum capacity and hypothesis testing," Phys. Rev. Lett. **108**, 200501 (2012).
[42] Zi-Wen Liu, Xueyuan Hu, and Seth Lloyd, "Resource destroying maps," Phys. Rev. Lett. **118**, 060502 (2017).
[43] Fernando G. S. L. Brandão and Gilad Gour, "Reversible framework for quantum resource theories," Phys. Rev. Lett. **115**, 070503 (2015).
[44] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).
[45] P. W. Shor, "Fault-tolerant quantum computation," in *Proceedings of 37th Conference on Foundations of Computer Science* (1996) pp. 56–65.
[46] Earl T. Campbell, Barbara M. Terhal, and Christophe Vuillot, "Roads towards fault-tolerant universal quantum computation," Nature **549**, 172 EP – (2017).
[47] Daniel Gottesman, "Class of quantum error-correcting codes saturating the quantum hamming bound," Phys. Rev. A **54**, 1862–1868 (1996).
[48] Daniel Gottesman, *Stabilizer codes and quantum error correction*, Ph.D. thesis, California Institute of Technology (1997).
[49] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, "Quantum error correction and orthogonal geometry," Phys. Rev. Lett. **78**, 405–408 (1997).
[50] Daniel Gottesman, "The Heisenberg Representation of Quantum Computers," arXiv e-prints , quant-ph/9807006 (1998), arXiv:quant-ph/9807006 [quant-ph].
[51] Scott Aaronson and Daniel Gottesman, "Improved simulation of stabilizer circuits," Phys. Rev. A **70**, 052328 (2004).
[52] Daniel Gottesman and Isaac L. Chuang, "Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations," Nature **402**, 390–393 (1999).
[53] Victor Veitch, S A Hamed Mousavian, Daniel Gottesman, and Joseph Emerson, "The resource theory of stabilizer quantum computation," New Journal of Physics **16**, 013009 (2014).
[54] Mark Howard and Earl Campbell, "Application of a resource theory for magic states to fault-tolerant quantum computing,"

Phys. Rev. Lett. **118**, 090501 (2017).
[55] Michael Beverland, Earl Campbell, Mark Howard, and Vadym Kliuchnikov, "Lower bounds on the non-Clifford resources for quantum computations," arXiv e-prints , arXiv:1904.01124 (2019), arXiv:1904.01124 [quant-ph].
[56] Earl T. Campbell, "Catalysis and activation of magic states in fault-tolerant architectures," Phys. Rev. A **83**, 032317 (2011).
[57] Sergey Bravyi and David Gosset, "Improved classical simulation of quantum circuits dominated by clifford gates," Phys. Rev. Lett. **116**, 250501 (2016).
[58] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard, "Simulation of quantum circuits by low-rank stabilizer decompositions," arXiv e-prints , arXiv:1808.00128 (2018), arXiv:1808.00128 [quant-ph].
[59] Anirudh Krishna and Jean-Pierre Tillich, "Towards low overhead magic state distillation," arXiv e-prints , arXiv:1811.08461 (2018), arXiv:1811.08461 [quant-ph].
[60] Roger A Horn and Charles R Johnson, *Matrix analysis* (Cambridge university press, 2012).
[61] Kun Fang, Xin Wang, Marco Tomamichel, and Mario Berta, "Quantum channel simulation and the channel's smooth max-information," in *2018 IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2018) pp. 2326–2330.
[62] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Transforming quantum operations: Quantum supermaps," EPL (Europhysics Letters) **83**, 30004 (2008).

Supplemental Material: No-Go Theorems for Quantum Resource Distillation

This supplemental material provides a more detailed analysis and proofs of some results omitted in the main text.

I. TECHNICAL LEMMAS

Lemma 2 *For any full rank states ρ and any quantum state σ, their quantum hypothesis testing relative entropy is continuous around $\epsilon = 0$. That is, for any $0 \le \epsilon < \lambda_{\min}(\rho)$ where $\lambda_{\min}(\rho)$ is the smallest eigenvalue of ρ, it holds that*

$$0 \le D_H^\epsilon(\rho\|\sigma) \le \log \frac{\lambda_{\min}(\rho)}{\lambda_{\min}(\rho) - \epsilon}.$$

Proof. Suppose M is an optimal measurement operator that attains $D_H^\epsilon(\rho\|\sigma)$. Then we have $0 \le M \le \mathbb{1}$, $\operatorname{Tr}\rho M \ge 1-\epsilon$ and $D_H^\epsilon(\rho\|\sigma) = -\log \operatorname{Tr} M\sigma$. Denote the non-zero eigenvalues of ρ and M as $\{\lambda_i\}_{i=1}^d$ and $\{m_j\}_{j=1}^k$, which are both sorted in an non-increasing order. Let $\lambda_{\min} = \min_i \lambda_i$ and $m_{\min} = \min_j m_j$. Since ρ is full rank, we have $k \le d$. We first argue that $k = d$, i.e., M is full rank. Suppose $k < d$, then we have

$$1 - \epsilon \le \operatorname{Tr}\rho M \le \sum_{i=1}^{k} \lambda_i m_i \le \sum_{i=1}^{k} \lambda_i \le 1 - \lambda_{\min}, \tag{S1}$$

where the second inequality follows from the von Neumann's trace theorem [60, Theorem 7.4.1.1], and the third inequality follows since $m_i \le 1$. Therefore, Eq. (S1) contradicts to the assumption that $\epsilon < \lambda_{\min}$ and thus $k = d$. Similar to Eq. (S1), we have

$$1 - \epsilon \le \operatorname{Tr}\rho M \le \sum_{i=1}^{d} \lambda_i m_i \le \left(\sum_{i=1}^{d-1} \lambda_i\right) + \lambda_{\min} m_{\min}, \tag{S2}$$

where the last inequality follows from the rearrangement inequality and the fact that $m_i \le 1$. This implies $m_{\min} \ge 1 - \epsilon/\lambda_{\min}$. Then we have $M \ge m_{\min}\mathbb{1} \ge (1 - \epsilon/\lambda_{\min})\mathbb{1}$, and thus

$$D_H^\epsilon(\rho\|\sigma) = -\log \operatorname{Tr} M\sigma \le -\log \operatorname{Tr}(1 - \epsilon/\lambda_{\min})\sigma = \log \frac{\lambda_{\min}(\rho)}{\lambda_{\min}(\rho) - \epsilon}. \tag{S3}$$

This completes the proof. ∎

Lemma 4 *For any two flagged quantum states $\rho_i = p_i|0\rangle\langle 0| \otimes \sigma_i + (1-p_i)|1\rangle\langle 1| \otimes \omega_i$ with $i \in \{1,2\}$ and $p_i \in [0,1]$, it holds that $\beta_\epsilon(\rho_1\|\rho_2) \le p_2\beta_\epsilon(\sigma_1\|\sigma_2) + (1-p_2)\beta_\epsilon(\omega_1\|\omega_2)$.*

Proof. Suppose $\beta_\epsilon(\sigma_1\|\sigma_2)$ and $\beta_\epsilon(\omega_1\|\omega_2)$ are achieved by optimal measurement operators M and N respectively. Then we can take $Q = |0\rangle\langle 0| \otimes M + |1\rangle\langle 1| \otimes N$. It is clear that $0 \le Q \le \mathbb{1}$ and $\operatorname{Tr} Q\rho_1 = p_1 \operatorname{Tr} M\rho + (1-p_1)\operatorname{Tr} N\omega_1 \ge p_1(1-\epsilon) + (1-p_1)(1-\epsilon) = 1 - \epsilon$. So Q is a feasible measurement operator for $\beta_\epsilon(\rho_1\|\rho_2)$. Thus we have

$$\beta_\epsilon(\rho_1\|\rho_2) \le \operatorname{Tr} Q\rho_2 = p_2 \operatorname{Tr} M\sigma_2 + (1-p_2)\operatorname{Tr} N\omega_2 = p_2\beta_\epsilon(\sigma_1\|\sigma_2) + (1-p_2)\beta_\epsilon(\omega_1\|\omega_2), \tag{S4}$$

which completes the proof. ∎

Lemma 5 *For any linear suboperation $\mathcal{L}$, there exists a free state $\omega \in \mathcal{F}$ such that $\operatorname{Tr}\mathcal{L}(\omega) \ge (1 + R(\rho))^{-1} \operatorname{Tr}\mathcal{L}(\rho)$, where $R(\rho) \equiv \min\{s | \exists\, \sigma, s \ge 0,\ s.t.\ (\rho + s\sigma)/(1+s) \in \mathcal{F}\}$ is the (generalized) robustness of state ρ.*

Proof. By definition of $R(\rho)$, there exists $\omega \in \mathcal{F}$ such that $\omega = \frac{1}{1+R(\rho)}\rho + \frac{R(\rho)}{1+R(\rho)}\sigma$. By linearity of $\mathcal{L}$,

$$\operatorname{Tr}\mathcal{L}(\omega) = \frac{1}{1+R(\rho)} \operatorname{Tr}\mathcal{L}(\rho) + \frac{R(\rho)}{1+R(\rho)} \operatorname{Tr}\mathcal{L}(\sigma), \tag{S5}$$

and thus the desired bound directly follows. ∎

II. RESOURCE THEORY OF QUANTUM CHANNELS

Analogous to the resource theories of quantum states, a resource theory of quantum channels can also be built upon two basic ingredients: the set of free channels $\boldsymbol{O}$ and the set of free superchannels (maps from channels to channels) $\boldsymbol{\Theta}$, with the golden rule $\boldsymbol{\Theta}(\boldsymbol{O}) \subseteq \boldsymbol{O}$ which selects all superchannels that can be possibly allowed (including those admitting implementations by free combs considered in [40]). Now the general question becomes whether there exists a free superchannel $\Pi \in \boldsymbol{\Theta}$ that maps one quantum channel $\mathcal{N}$ to another quantum channel $\mathcal{M}$, i.e., $\Pi(\mathcal{N}_{A\to B}) = \mathcal{M}_{C\to D}$. If so, we say that $\mathcal{M}$ can be simulated by $\mathcal{N}$.

In analogy to the state distillation tasks where one aims to turn a noisy state into a pure one, here we want to turn a noisy channel into a unitary one, which preserves information. Below we show an elementary channel version of the no-go results, which says that perfect simulation of unitary channels are generically impossible. Similarly, we say a quantum channel $\mathcal{N}$ contains a free component if there exists free channel $\mathcal{E} \in \boldsymbol{O}$ and another quantum channel $\mathcal{M}$ such that $\mathcal{N} = p\mathcal{E} + (1-p)\mathcal{M}$ with $p > 0$.

Theorem 9 (Channel no-go) *Given any primitive channel $\mathcal{N}_{A\to B} \notin \boldsymbol{O}(A \to B)$ with free component (i.e. can be written as $\mathcal{N} = p\mathcal{E} + (1-p)\mathcal{M}, p \in (0,1)$ where $\mathcal{E} \in \boldsymbol{O}(A \to B)$ and $\mathcal{M}$ is any channel) and any target unitary resource channel $\mathcal{U}_{C\to D} \notin \boldsymbol{O}(C \to D)$, there is no free superchannel Π transforming $\mathcal{N}_{A\to B}$ to $\mathcal{U}_{C\to D}$.*

If our target $\mathcal{U}$ is a noiseless quantum channel, the channel transformation task will correspond to the well-studied channel capacity problem (see e.g. [32, 61]). A straightforward implication of the above result shows that the zero-error quantum capacity of a generic quantum channel, such as the quantum depolarizing channel, is zero.

The proof of this no-go result requires the definition of the channel's min-relative entropy and its monotonicity under superchannels. The channel's min-relative entropy is defined as

$$D_{\min}(\mathcal{N}\|\mathcal{M}) \equiv \sup_{\rho_{AA'}} D_{\min}(\mathcal{N}_{A'\to B}(\rho_{AA'})\|\mathcal{M}_{A'\to B}(\rho_{AA'})) = \sup_{\rho_A} D_{\min}(\mathcal{N}_{A'\to B}(\phi_{AA'})\|\mathcal{M}_{A'\to B}(\phi_{AA'})) \tag{S6}$$

where $\phi_{AA'}$ is a purification of ρ_A. The second equality follows from the joint-convexity of $D_{\min} \equiv D_H^0$. The monotonicity of the channel's min-relative entropy holds as follows:

Lemma S1 *For any superchannel Π and quantum channels $\mathcal{N}_{A'\to B}$ and $\mathcal{M}_{A'\to B}$, it holds that $D_{\min}(\Pi(\mathcal{N})\|\Pi(\mathcal{M})) \le D_{\min}(\mathcal{N}\|\mathcal{M})$.*

Proof. Note that any superchannel Π can be implemented by pre- and post- quantum processings as $\Pi(\mathcal{N}_{A\to B}) = \mathcal{D}_{BR\to D} \circ (\mathcal{N}_{A\to B} \otimes \mathbb{1}_{R\to R}) \circ \mathcal{E}_{C\to AR}$, where $\mathcal{E}, \mathcal{D}$ are channels [62]. Suppose Π is realized by $\{\mathcal{E}, \mathcal{D}\}$ and the optimal solution of $D_{\min}(\Pi(\mathcal{N})\|\Pi(\mathcal{M}))$ is achieved by state ρ_{CE}. Denote $\sigma_{ARE} = \mathcal{E}_{C\to AR}(\rho_{CE})$. Then we have the following chain of inequalities,

$$\begin{aligned}
&D_{\min}(\Pi(\mathcal{N})\|\Pi(\mathcal{M})) \\
&= D_{\min}(\mathcal{D}_{BR\to D} \circ \mathcal{N}_{A\to B}(\sigma_{ARE})\|\mathcal{D}_{BR\to D} \circ \mathcal{M}_{A\to B}(\sigma_{ARE})) &\text{(S7)}\\
&\le D_{\min}(\mathcal{N}_{A\to B}(\sigma_{ARE})\|\mathcal{M}_{A\to B}(\sigma_{ARE})) &\text{(S8)}\\
&\le D_{\min}(\mathcal{N}\|\mathcal{M}), &\text{(S9)}
\end{aligned}$$

where the first inequality follows from the data-processing inequality of $D_{\min}$, and the second inequality follows by definition. ■

Proof of Theorem 9: We prove this by contradiction. Suppose there is a free superchannel Π such that $\Pi(\mathcal{N}) = \mathcal{U}$. Denote $J_{\mathcal{U}}$ as the corresponding Choi state of $\mathcal{U}$. Since $\mathcal{U}$ is a unitary channel, we know that $J_{\mathcal{U}}$ is a pure state. By the assumption of $\mathcal{U} \notin \boldsymbol{O}$, we have for any $\mathcal{E} \in \boldsymbol{O}$ it holds $D_{\min}(\mathcal{U}\|\mathcal{E}) \ge D_{\min}(J_{\mathcal{U}}\|J_{\mathcal{E}}) = -\log \mathrm{Tr}\, J_{\mathcal{U}} J_{\mathcal{E}} > 0$. On the other hand, suppose $\mathcal{N} = p\mathcal{E} + (1-p)\mathcal{M}$ with $\mathcal{E} \in \boldsymbol{O}$, we have $D_{\min}(\mathcal{N}\|\mathcal{E}) = D_{\min}(p\mathcal{E} + (1-p)\mathcal{M}\|\mathcal{E}) = 0$. Then we have

$$0 < \min_{\mathcal{E}\in\boldsymbol{O}} D_{\min}(\mathcal{U}\|\mathcal{E}) \le \min_{\mathcal{E}\in\boldsymbol{O}} D_{\min}(\mathcal{U}\|\Pi(\mathcal{E})) = \min_{\mathcal{E}\in\boldsymbol{O}} D_{\min}(\Pi(\mathcal{N})\|\Pi(\mathcal{E})) \le \min_{\mathcal{E}\in\boldsymbol{O}} D_{\min}(\mathcal{N}\|\mathcal{E}) = 0, \tag{S10}$$

where the second inequality follows since the minimization is restrict to $\mathcal{E} \in \Pi(\boldsymbol{O})$ on the r.h.s., and the third inequality follows from the monotonicity in Lemma S1. This forms a contradiction. □

The approximate and probabilistic versions are left for future work.