

One-shot distillation with constant overhead using catalysts

Received: 17 October 2025

Accepted: 22 April 2026

Published online: 04 May 2026

Kun Fang¹  & Zi-Wen Liu² 

Quantum resource distillation is a fundamental task in quantum information science and technology. Minimizing the overhead of distillation is crucial for the realization of quantum computation and other technologies. Here we explicitly demonstrate how, for general quantum resources, suitably designed quantum catalysts (i.e., auxiliary systems that remain unchanged before and after the process) can lift the multi-shot or asymptotic average distillation overhead to the one-shot setting. This can enable one-shot distillation with constant overhead, thereby overcoming the established logarithmic lower bound. In particular, for magic state distillation, our catalysis method paves a path for tackling the diverging batch size problem associated with code-based low-overhead protocols by enabling arbitrary reduction of the protocol size for any desired accuracy. Notably, this yields constant-overhead magic state distillation with controllable protocol size. Furthermore, we demonstrate a tunable spacetime trade-off between overhead and success probability enabled by catalysts which offers significant versatility for practical implementation. Finally, we extend catalysis techniques to dynamical quantum resources and show that channel mutual information determines one-shot catalytic channel transformation, thereby advancing our understanding for both dynamical catalysis and information theory.

Quantum technologies hold the potential of offering revolutionary advantages over classical methods in crucial technological tasks, including computation and communication, driven by accurate manipulation of various kinds of quantum resources, including quantum entanglement¹, quantum coherence², and nonstabilizerness/magic³. However, the fragile nature of quantum information resources presents a formidable challenge, as the microscopic systems are highly susceptible to various sources of interference, such as environmental noises and control imperfections⁴. These noise and error effects can significantly undermine the effectiveness and security of quantum computation and communication, thereby limiting the scalability and applicability of quantum technologies. In addressing these challenges, quantum resource distillation^{3,5}—which entails converting noisy quantum resources into a smaller amount of purer ones with certain free operations—has emerged as a ubiquitously important

procedure. For instance, magic state distillation^{3,6–18} is of major interest in quantum computation as a leading scheme for the implementation of logical non-Clifford gates, which constitutes the main bottleneck in building fault-tolerant quantum computers. Other notable examples include entanglement distillation^{5,19–21} and coherence distillation^{22–25}, which play key roles in quantum communication, networking and cryptography.

Given the wide importance of resource distillation procedures, optimizing their efficiency constitutes a vital problem for quantum information technology. More specifically, the main goal is to reduce the distillation overhead, i.e., the amount of noisy inputs needed to produce certain desired outputs, for which the key figure of merit is the exponent γ in the $O(\log^\gamma(1/\epsilon))$ asymptotic scaling with respect to target error rate ϵ as $\epsilon \rightarrow 0$. Mainly for theoretical interest and simplicity, researchers have traditionally focused on the idealized infinite-

¹School of Data Science, The Chinese University of Hong Kong, Shenzhen, Guangdong, China. ²Yau Mathematical Sciences Center, Tsinghua University, Beijing, China. ✉e-mail: kunfang@cuhk.edu.cn; zwliu0@tsinghua.edu.cn

shot setting where one considers an average notion of overhead over an uncontrollably large amount of outputs. In particular, there have been intensive efforts and advances that gradually reduce the average overhead of magic state distillation since the earliest proposal achieving $\gamma \approx 2.46^3$. It was long conjectured and believed that γ cannot be reduced to below one⁷; however, striking breakthroughs in recent years have broken this barrier, achieving sublogarithmic average overhead^{8,9} and culminating in recent results that eventually attain (almost) constant average overhead^{14–16,18}.

Nonetheless, in realistic scenarios, both the number of available inputs and the desired outputs are always finite. This corresponds to the one-shot regime, which is fundamentally distinct from the above infinite-shot setting and directly captures resource-processing tasks in practice. In this practical setting, the recent universal no-go theorems for resource purification (which applies to virtually any type of resource, including entanglement, coherence, asymmetry, thermal nonequilibrium, and magic)^{11,12} rigorously imply the existence of a $\gamma \geq 1$ barrier, i.e., a logarithmic lower bound on distillation overhead. Notably, although there exist protocols achieving sublogarithmic average overhead in the asymptotic setting (as mentioned above), they inevitably suffer from a common “batch size problem” from a practical perspective: they usually involve a prohibitively large number of states or dimensions (e.g., the Hastings–Haah sublogarithmic protocol⁸ works with a code that entails $\geq 2^{58}$ input and $\geq 2^{44}$ output states), and the protocol volume must diverge rapidly as $\varepsilon \rightarrow 0$, which incurs enormous practical costs.

In this work, we introduce a distillation scheme utilizing the idea of quantum catalysis, a phenomenon analogous to chemical catalysis where catalysts can facilitate reactions by providing alternative pathways with lower activation energies, demonstrating how it helps circumvent the above limitations. In the realm of quantum information, the catalysis phenomenon involves an auxiliary quantum system, namely the catalyst, which participates in the transformation process, enabling transformations that would otherwise be unachievable while remaining unaltered before and after the process and thus can be reused. Quantum catalysis was first discussed in the context of entanglement transformation²⁶ more than two decades ago. For example, by the majorization criterion^{27,28}, an entangled state $|\psi_1\rangle = \sqrt{0.4}|00\rangle + \sqrt{0.4}|11\rangle + \sqrt{0.1}|22\rangle + \sqrt{0.1}|33\rangle$ cannot be transformed into $|\psi_2\rangle = \sqrt{0.5}|00\rangle + \sqrt{0.25}|11\rangle + \sqrt{0.25}|22\rangle$ via local quantum operations and classical communication. However, there exists a catalyst state $|\phi\rangle = \sqrt{0.6}|44\rangle + \sqrt{0.4}|55\rangle$ that enables the exact transformation from $|\psi_1\rangle \otimes |\phi\rangle$ to $|\psi_2\rangle \otimes |\phi\rangle$. Beyond entanglement manipulation, quantum catalysis has since been extensively studied across various resource theories, including quantum coherence^{29,30}, thermodynamics^{31–33}, purity³⁴ and magic^{35–37}. This area remains active as evidenced by recent works^{32,38–42}, with comprehensive reviews available in refs. 43,44. While catalytic methods are widely useful, recent studies have also revealed fundamental limitations of their power, e.g., they cannot overcome bound entanglement⁴⁵, and cannot increase distillable entanglement⁴⁶. These “negative” results highlight that whether catalysis can introduce advantages for a certain task is a nontrivial question⁴⁷. Furthermore, existing literature on distillation mostly focuses on the distillation rate, namely, how much pure target resource can be extracted from a fixed amount of noisy input resource. This is conceptually dual to the overhead considered in this work, where the aim is to minimize the amount of input for some given target.

The main contribution of this work is to demonstrate the effectiveness of catalytic methods in enhancing distillation efficiency, particularly highlighting their striking capability to arbitrarily reduce one-shot distillation overhead, which is unknown before. This, in some sense, completes the picture of distillation overhead optimization. Specifically, we establish a general and rigorous ordering of resource overheads for one-shot (unassisted), multi-shot

average, and one-shot catalytic distillation settings, as illustrated in Fig. 1, by showing how any multi-shot distillation protocol can be converted into a one-shot catalytic protocol while maintaining the overhead by designing suitable catalysts with reusability guarantees. In particular, our catalytic approach can be applied to make the batch size of any protocol arbitrarily small (e.g., an n -to- m protocol can be converted into a $\lceil n/m \rceil$ -to-1 protocol) for any accuracy. Through magic state distillation, this provides a strategy for alleviating the cost of fault-tolerant quantum computation. Furthermore, we demonstrate that the failure probability and overhead, corresponding to time and space costs respectively, can be flexibly traded for each other with the help of catalysts. Notably, we prove that the optimal constant for constant-overhead magic state distillation can even be reduced to 1 at the price of sacrificing the success probability by a constant factor, pushing the efficiency of distillation to its ultimate limit. Lastly, we extend the catalytic technique to the channel (dynamical) setting. As an application, we provide a one-shot operational interpretation of channel mutual information, showing that it determines the catalytic convertibility of quantum channels and thereby resolving the channel version of a conjecture of Wilming³³. In Fig. 2, we illustrate the fundamental advantages of our catalytic methods more clearly, including a comparison with prominent existing distillation protocols. Furthermore, from a practical perspective, our catalytic schemes offer substantial flexibility in experiment and architecture design by enabling one to freely tailor the qubit and time costs to suit specific hardware features or capabilities. It is worth noting that our protocols consist of quite simple operations and are thus expected to be particularly friendly for experiments.

Results

Preliminaries: general resource theory and distillation overhead

To present our results in a rigorous and unified manner, we shall adopt the language of general resource theory, a framework that finds great success in studying different quantum resource features in recent years (background and setup can be found in e.g., refs. 10,11,48,49). A standard resource theory is defined by a set of free operations $\mathcal{O}$ and a set of free states $\mathcal{F} \subseteq \mathcal{D}$ (the complement of which are resource states), where $\mathcal{D}$ denotes the set of all density matrices. Taking the magic theory as an example, $\mathcal{F}$ consists of stabilizer states, and Clifford operations are a standard choice of $\mathcal{O}$. Generally speaking, $\mathcal{F}$ and $\mathcal{O}$ can be adaptively defined, leading to a wide variety of meaningful resource theories, as long as they follow a golden rule: any free operation can only map a free state to another free state, i.e., $\mathcal{E}(\rho) \in \mathcal{F}, \forall \rho \in \mathcal{F}, \forall \mathcal{E} \in \mathcal{O}$.

The general goal of distillation tasks is to transform noisy resources into pure ones by some protocol represented by a free operation. In practice, we are also interested in protocols that only produce desired outputs with a certain probability (e.g., protocols based on error correction, which work upon passing the syndrome measurements). To encompass such probabilistic cases, consider the generalization of free operations $\mathcal{O}$ to the class of free subnormalized quantum operations, i.e., completely positive and trace-nonincreasing maps $\mathcal{O}_{\text{sub}} := \{\mathcal{L} | \forall \rho \in \mathcal{F}, \exists p \geq 0, \psi \in \mathcal{F}, \text{ s.t. } \mathcal{L}(\rho) = p \cdot \psi\}$. A free probabilistic protocol that transforms ρ to γ with probability p is modeled by a quantum operation $\mathcal{E}_{A \rightarrow XB}(\rho_A) = |0\rangle\langle 0|_X \otimes \mathcal{L}_{A \rightarrow B}(\rho_A) + |1\rangle\langle 1|_X \otimes \mathcal{G}_{A \rightarrow B}(\rho_A)$, where $\mathcal{L}, \mathcal{G} \in \mathcal{O}_{\text{sub}}$. Here, X is an external flag register that keeps track of whether the protocol succeeds (0) or not (1); then $\mathcal{L} \in \mathcal{O}_{\text{sub}}$ represents the successful transformation such that $\mathcal{L}_{A \rightarrow B}(\rho_A) = p\gamma_B$ where $p = \text{Tr}[\mathcal{L}(\rho)]$. The case where $\mathcal{L}$ is a completely positive trace-preserving (CPTP) map and thus $p = 1$ corresponds to a deterministic protocol.

As already emphasized in the introduction, previous research on distillation overhead has focused on the asymptotic average overhead in the infinite-shot limit for simplicity, whereas the one-shot setting

with controllable output size is practically more relevant and constitutes the main target in this work.

Now we lay down formal definitions of the different types of overhead. Denote the space for the source and target states as S , and the ancillary system that supports the catalyst as A . For any two quantum states $\rho, \sigma \in \mathcal{D}(S)$, define their trace distance by $\Delta(\rho, \sigma) := \|\rho - \sigma\|_1/2$ with $\|\cdot\|_1$ being the trace norm. Denote $\rho^{\otimes n} \xrightarrow{(\varepsilon, p)} \sigma^{\otimes m}$ if there exists a resource transformation protocol $\mathcal{L} \in \mathcal{O}_{\text{sub}}$ such that $\mathcal{L}(\rho^{\otimes n}) = p \cdot \eta^m$ with $\eta^m \in \mathcal{D}(S^m)$ and $\Delta(\eta_i^m, \sigma) \leq \varepsilon$ for all $i \in \{1, \dots, m\}$, where η^m is some (not necessarily product) state in the m -copy space and η_i^m is the i -th marginal of η^m . Here we follow the convention in distillation studies and focus on the error of each marginal states^{3,7}. Since pure targets are typically of interest, this is equivalent to defining the error over the global states.

Definition 1. (Unassisted overhead). Let $\rho, \sigma \in \mathcal{D}(S)$ be two arbitrary quantum states. Considering ρ and σ as the primitive (source) state and target state of distillation respectively, the (conventional unassisted notions of) one-shot distillation overhead $C_{\varepsilon, p}$ and multi-shot average

distillation overhead $\bar{C}_{\varepsilon, p}$ are defined as

$$C_{\varepsilon, p}(\rho, \sigma) := \min\{n : \rho^{\otimes n} \xrightarrow{(\varepsilon, p)} \sigma\}, \quad (1)$$

$$\bar{C}_{\varepsilon, p}(\rho, \sigma) := \min\{\lceil n/m \rceil : \rho^{\otimes n} \xrightarrow{(\varepsilon, p)} \sigma^{\otimes m}\}, \quad (2)$$

respectively.

Furthermore, denote $\rho^{\otimes n} \otimes \omega \xrightarrow{(\varepsilon, p)} \sigma \otimes \omega$ if there exists a transformation protocol $\mathcal{L} \in \mathcal{O}_{\text{sub}}$ and a catalyst $\omega \in \mathcal{D}(A)$ such that $\mathcal{L}(\rho^{\otimes n} \otimes \omega) = p \cdot \nu$ with $\nu \in \mathcal{D}(SA)$, $\Delta(\nu_S, \sigma_S) \leq \varepsilon$ and $\nu_A = \omega_A$, where ν_A, ν_S are two marginals of ν .

Definition 2. (One-shot catalytic distillation overhead). Let $\rho, \sigma \in \mathcal{D}(S)$ be two arbitrary quantum states. Considering ρ and σ as the primitive (source) state and target state of distillation, respectively, the one-shot catalytic distillation overhead $\tilde{C}_{\varepsilon, p}$ is defined as

$$\tilde{C}_{\varepsilon, p}(\rho, \sigma) := \min\{n : \rho^{\otimes n} \otimes \omega \xrightarrow{(\varepsilon, p)} \sigma \otimes \omega, \exists \omega \in \mathcal{D}\}. \quad (3)$$

The standard distillation setting of usual interest is to distill copies of less noisy versions of some target pure state by consuming more copies of more noisy versions of this target state, namely, σ is pure and $n \geq m$. For example, the typical task in magic state distillation is to distill less but better T states from noisy ones.

Previous results on distillation overhead are mainly bounds on the multi-shot average notion obtained by taking the asymptotic limit $n, m \rightarrow \infty$ which corresponds to the idealistic infinite-resource scenario. In contrast, the one-shot setting captures the practical finite-resource scenario. With the above notations, the no-go theorem of ref. 11 sets the logarithmic fundamental limit $C_{\varepsilon, p}(\rho, \sigma) = \Omega(\log(1/\varepsilon))$ for ordinary unassisted one-shot distillation overhead. Our aim here is to demonstrate how the catalytic version $\tilde{C}_{\varepsilon, p}$ can break this barrier while maintaining the one-shot nature.

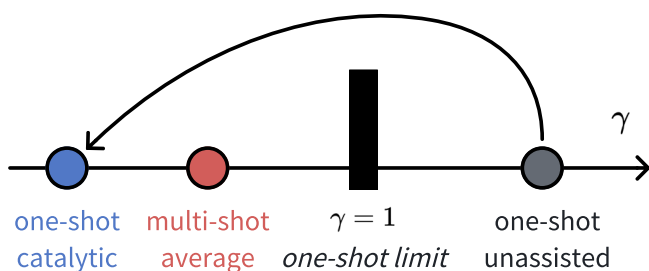


Fig. 1 | Comparison of distillation overhead $\log^\gamma(1/\varepsilon)$ across different settings. A fundamental no-go limit of $\gamma \geq 1$ has been shown to exist for one-shot (unassisted) distillation^{11,12}. We show in this work that this limit can be surpassed with the aid of quantum catalysts.

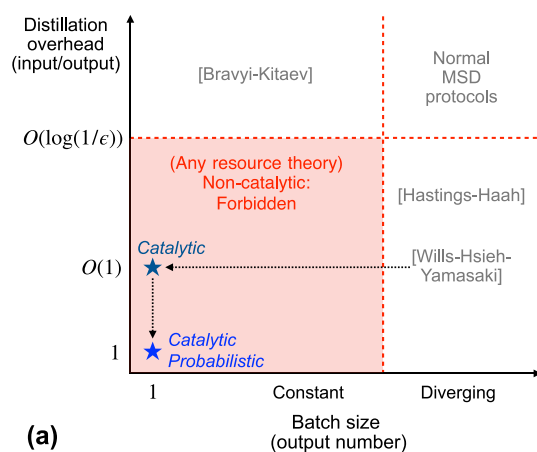
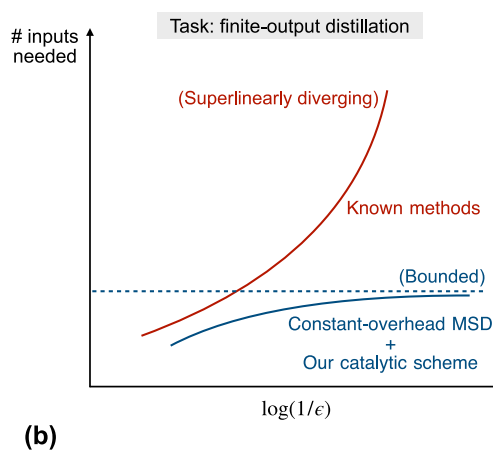


Fig. 2 | Illustration of the fundamental advantages of our catalytic scheme over existing distillation methods. **a** A distillation overhead versus batch size diagram. Due to the no-go theorems in Refs. 11,12, for virtually any well-behaved resource theory, at least one of the following must hold for non-catalytic protocols (where ε denotes the target error rate): (i) the distillation overhead is at least logarithmic, i.e., $\Omega(\log(1/\varepsilon))$; (ii) the distillation cannot be one-shot, that is, the batch size (number of outputs) must diverge as ε decreases (which we refer to as the “batch size problem”). Consequently, the lower-left region is fundamentally forbidden (highlighted in red). For the particularly important case of magic state distillation, normal protocols achieve neither (i) nor (ii) and therefore fall in the upper-right region; the logarithmic-overhead barrier⁷ was only recently overcome by e.g.,



Hastings-Haah⁸ and Wills-Hsieh-Yamasaki¹⁴, which are regarded as major breakthroughs, yet they necessarily suffer from the batch size problem. Our catalytic strategy enables one to bypass these barriers and simultaneously achieve (i) and (ii): for example, our catalytic strategy can make the batch size of the constant-overhead protocol¹⁴ arbitrarily small while preserving the constant overhead (Applications in magic state distillation), and by allowing probabilistic success, we can further achieve arbitrarily small overhead (Catalytic spacetime conversion: trading success probability for reduced overhead). **b** For finite-output distillation tasks, existing protocol methods require a superlinearly diverging amount of input resources with respect to $\log(1/\varepsilon)$ decreases, whereas our catalytic magic state distillation schemes can achieve arbitrarily small ε with finite input resources.

General theory: one-shot catalytic resource distillation

Now we introduce our main results. All proofs and technical analysis are deferred to the Methods section. Here, we deal with the standard distillation setting where more source states are consumed to distill less pure target states, namely, σ is pure and $n \geq m$. Axiomatically, it is assumed that appending free systems, discarding and permuting subsystems, and classically controlled free operations are free, and no other properties of the resource theory are needed, rendering the framework highly flexible and applicable to most relevant theories.

Since a one-shot distillation protocol can be regarded as a multi-shot distillation protocol with further conditions, it follows directly that $\bar{C}_{\varepsilon,p}(\rho, \sigma) \leq C_{\varepsilon,p}(\rho, \sigma)$ for any quantum states ρ and σ . Here we adapt a catalysis technique originally proposed in ref. 50 for LOCC (local operations with classical communication) transformations of quantum entanglement (further explored in recent works such as refs. 30,32,33,38–42) to our general transformation task, which involves an unequal number of source and target states. This is achieved by regrouping the source states into blocks and embedding the target state into a larger space. Particularly, we show that any multi-shot distillation protocol from $\rho^{\otimes n}$ to $\sigma^{\otimes m}$ can effectively be turned into a one-shot catalytic protocol from $\rho^{[n/m]} \otimes \omega$ to $\sigma \otimes \omega$ with the same parameters. Here, the catalyst ω plays a crucial role by enabling new transformation pathways that are otherwise unattainable and absorbing part of the complexity of the transformation process. Although the catalyst temporarily participates in the transformation, it is fully restored afterward. As a result, we have the inequality $\bar{C}_{\varepsilon,p}(\rho, \sigma) \leq \bar{C}_{\varepsilon,p}(\rho, \sigma)$. In conclusion, the general relation of distillation overhead across different settings is as follows:

Theorem 3. For any quantum states ρ and σ , target error $\varepsilon \in [0, 1]$ and success probability $p \in [0, 1]$, the following relation holds: $\bar{C}_{\varepsilon,p}(\rho, \sigma) \leq \bar{C}_{\varepsilon,p}(\rho, \sigma) \leq C_{\varepsilon,p}(\rho, \sigma)$. In particular, any multi-shot distillation protocol can be turned into a one-shot catalytic distillation protocol with the same overhead, implying the first inequality.

It is worth noting that the catalyst technique has been studied in the context of distillation before. However, previous works focused on optimizing the output resource that can be extracted from some fixed noisy input, for which known results are negative. For instance, it has been shown that catalysts cannot overcome bound entanglement⁴⁵ and cannot increase distillable entanglement⁴⁶. In contrast, this work is intended to initiate a different perspective by establishing the advantages of catalysts in enhancing distillation overhead, where the aim is to minimize the amount of noisy input required to achieve a desired target. Theorem 3 establishes a foundational result in this new direction, demonstrating the hitherto unexplored potential of catalysts to reduce the distillation overhead. In the sections that follow, we demonstrate the effectiveness of catalysts in this regard through concrete examples, particularly in the context of magic state distillation, highlighting their capability to surpass previously established limits.

A key advantage of using a catalyst is its recoverability after the transformation, allowing for repeated reuse. We now provide a theoretical guarantee that the catalyst can be reused indefinitely without any degradation in its expected performance.

Theorem 4. For a deterministic one-shot catalytic resource distillation procedure, after $l \geq 1$ repeated uses of the catalyst ω_A , we obtain a joint state $\nu_{S_1 S_2 \dots S_l A}$ such that the catalyst is exactly returned on its marginal $\nu_A = \omega_A$ and the target states $\nu_{S_1} = \nu_{S_2} = \dots = \nu_{S_l}$ with error $\Delta(\nu_{S_i}, \sigma_{S_i}) \leq \varepsilon$ for all $i \in \{1, \dots, l\}$.

An apparent subtlety is that the catalyst may exhibit correlations with the remaining systems. However, when addressing distillation tasks aimed at pure target states, which is typically the case of interest, the correlation between the obtained target state and the remaining systems is under control. This can be made explicit by applying Lemma

8 from the Supplemental Material of ref. 46, which states that an error threshold on the target state $\Delta(\nu_{S_i}, |\phi\rangle\langle\phi|_{S_i}) \leq \varepsilon$ ensures that the correlation between the target state and the catalyst remains small, with $\Delta(\nu_{SA}, |\phi\rangle\langle\phi|_S \otimes \nu_A) \leq \varepsilon + 3\sqrt{\varepsilon}$. Consequently, any intermediate usage of the target state and the catalyst will have a negligible impact on the other systems. In the following, we give a further improved bound for this fact through a simplified proof.

Lemma 5. Suppose $\Delta(\nu_S, |\phi\rangle\langle\phi|_S) \leq \varepsilon$, then $\Delta(\nu_{SA}, |\phi\rangle\langle\phi|_S \otimes \nu_A) \leq 2\sqrt{\varepsilon}$.

We also note that the catalyst is guaranteed to remain effective in deterministic protocols, while for probabilistic ones, there is a risk of losing the validity of the catalyst when the transformation fails, which arises from the probabilistic nature of multi-shot protocols. Nevertheless, it is guaranteed that the probability of losing the catalyst in the converted one-shot catalytic protocol is no greater than p if the multi-shot protocol fails with probability p . Moreover, the failure probability is finite and can be significantly suppressed by e.g., concatenation, and there are ways to make probabilistic distillation protocols deterministic⁵¹. Finally, we stress that the magic state distillation protocols we apply later are already deterministic¹⁴, so the aforementioned risk is not present.

Applications in magic state distillation

As a potential resolution to the non-Clifford operation bottleneck of fault-tolerant quantum computing, magic state distillation has been a subject of extensive study since Bravyi and Kitaev's proposal in 2005³. The primary focus is to reduce the resource overhead of magic state distillation, with numerous improvements achieved over the years^{3,7–9,14–18,52–55}.

It was conjectured⁷ and long believed that the overhead exponent γ cannot be reduced to below one. In the one-shot setting, this can indeed be proven by applying the universal no-go theorem for resource purification in ref. 11. Remarkably, Hastings and Haah first found a code that achieves an average overhead of $n/m = O(\log^{\gamma}(1/\varepsilon))$ for $\gamma \approx 0.678$ ⁸, and very recent progress further shows that the exponent can be made arbitrarily close to 0 or even to equal 0, namely achieving constant average overhead^{14–16}. As explained, despite these advances, the low overhead in these protocols comes with the serious problem that they require uncontrollably large batch sizes as ε tends to zero. By applying the catalytic conversion established in Theorem 3, we can turn the low-overhead distillation protocols into one-shot counterparts that achieve arbitrary accuracy with an arbitrary amount of (even only one) target states. This provides a way to circumvent the fundamental limit of $\Omega(\log(1/\varepsilon))$ for one-shot distillation¹¹.

Corollary 6. There exist one-shot catalytic magic state distillation protocols that achieve any given target error with certainty (i.e., success probability 1) and constant overhead. In particular, by Theorem 4, the catalyst can be reused indefinitely without degradation in its performance.

As discussed, the correlation between the catalyst and the obtained target state can be made sufficiently small. A related result by Rubboli and Tomamichel⁵⁶ indicates that making such residual correlations arbitrarily small requires a divergent amount of resources in the catalyst if the resource theory has multiplicative maximum fidelity of resource (i.e., $\hat{F}(\rho_1 \otimes \rho_2) = \hat{F}(\rho_1) \cdot \hat{F}(\rho_2)$, where $\hat{F}(\rho) := \max_{\sigma \in \mathcal{F}} F(\rho, \sigma)$ with F being the normal state fidelity). However, the maximum fidelity in magic theory (also named stabilizer fidelity) is known to be non-multiplicative^[57, Section 6.2]. It remains an interesting open question whether, in magic theory, it is possible to find exotic catalysts with a finite amount of magic that achieve vanishing residual correlations or error.

Catalytic spacetime conversion: trading success probability for reduced overhead

In distillation tasks, the resource overhead essentially represents the space requirement, characterizing the size of quantum systems that need to be coherently manipulated in a single experiment. On the other hand, the success probability is associated with the number of experimental repetitions needed to accomplish the tasks and thus determines the time cost. As the size of the qubits and coherent control is limited, especially in the near term, it is usually more feasible to scale up an experiment by repeating the experiment multiple times rather than conducting it with larger systems in fewer trials. By leveraging quantum catalysis, we enable a trade-off between space and time, allowing the resource burden to be shifted from qubit overhead to time.

More specifically, we show that if a transformation from $\rho^{\otimes n}$ to $\sigma^{\otimes m}$ is achievable with success probability p , then a catalytic transformation from $\rho^{\otimes k}$ to σ can also be achieved with success probability $pm/\lceil n/k \rceil$, for any $1 \leq k \leq n/m$, effectively trading time for space. Notably, by setting $k = 1$, we obtain a one-shot catalytic distillation protocol with unit overhead and success probability pm/n , in which case the overhead is minimized to its extreme.

Theorem 7. Suppose there exists a distillation protocol transforming $\rho^{\otimes n}$ to $\sigma^{\otimes m}$ with success probability p and target error ε . Then it holds for any $1 \leq k \leq n/m$ that $\tilde{C}_{\varepsilon, pm/\lceil n/k \rceil}(\rho, \sigma) \leq k$. In particular, $\tilde{C}_{\varepsilon, pm/n}(\rho, \sigma) = 1$.

It should be noted that trading success probability for overhead is not always possible without the use of a catalyst, which underscores the significance of Theorem 7. Consider an example task of entanglement distillation where the goal is to transform a noisy entangled state ρ (e.g., an isotropic state) into a Bell state ψ with success probability p' and target error ε' . There exist nontrivial forbidden tuples (p', ε') according to the no-go theorems established in Refs. 11, 58, which are nevertheless achievable in the average case using the hashing protocol⁵⁹: $\rho^{\otimes mr}$ can be transformed into $\psi^{\otimes m}$ with arbitrarily small target error and success probability 1 for certain r and m , where $1/r$ corresponds to the hashing bound. This shows that achieving a multi-copy transformation does not necessarily guarantee the feasibility of a one-shot transformation by simply compromising the success probability, even when p' is chosen to be close to zero. Yet, this trade-off can always be achieved with the aid of a catalyst by our result.

As mentioned, very recent works^{14–16} demonstrated that magic state distillation can be achieved with constant overhead in the asymptotic limit but left the question of optimality of this constant open for future investigation. In particular, the protocol in ref. 14 achieves $p = 1$ and $n/m = O(1)$. By applying Theorem 7, we find that the optimal constant can be ultimately reduced to 1 with the aid of a catalyst, at the expense of compromising the success probability by a constant factor.

Corollary 8. There exist one-shot catalytic magic state distillation protocols that achieve any given target error with constant success probability using only one copy of the source magic state (i.e., constant overhead 1). In particular, by Theorem 4, the catalyst can be reused indefinitely without degradation in its performance.

Channel catalytic transformation theory

The concept of catalytic transformation can also be extended to the manipulation of quantum channels, which correspond to dynamical resources that play a crucial role in, e.g., quantum communication and quantum error correction^{12, 60–67}. In this section, we extend the catalysis technique to the channel setting.

Let $L(A \rightarrow B)$ be the set of all linear maps from Hilbert space A to Hilbert space B . Let $\text{CTP}(A \rightarrow B)$ be the set of all quantum channels (i.e., completely positive and trace-preserving maps) from A to B . The most

general transformation of a quantum channel is described by a superchannel, which is a linear map that maps quantum channels to quantum channels and has a physical realization with pre- and post-processing on the quantum channel upon which it acts. More specifically, Π is a superchannel⁶¹ if there exists a Hilbert space E with $|E| \leq |A_0| |B_0|$ and two CPTP maps $\Gamma_{\text{pre}} \in \text{CTP}(B_0 \rightarrow A_0 E)$ and $\Gamma_{\text{post}} \in \text{CTP}(A_1 E \rightarrow B_1)$ such that for all linear map $\Psi \in L(A_0 \rightarrow A_1)$, it holds that $\Pi[\Psi] = \Gamma_{\text{post}} \circ (\Psi \otimes \mathcal{I}_E) \circ \Gamma_{\text{pre}}$.

Before addressing catalytic channel manipulation, we first need to clarify the definition of a reduced channel. It is known that if a bipartite channel satisfies the non-signaling (NS) condition $\text{Tr}_{B_2} \circ \mathcal{N}_{A_1 A_2 \rightarrow B_1 B_2} = \text{Tr}_{B_2} \circ \mathcal{N}_{A_1 A_2 \rightarrow B_1 B_2} \circ \mathcal{R}_{A_2}^{\pi}$ where $\mathcal{R}^{\pi}$ is a replacer channel that sends any input to a fixed state π , then we can define a unique channel such that $\text{Tr}_{B_2} \circ \mathcal{N}_{A_1 A_2 \rightarrow B_1 B_2} = \mathcal{N}_{A_1 \rightarrow B_1} \circ \text{Tr}_{A_2}$. This channel is given by

$$\mathcal{N}_{A_1 \rightarrow B_1}(\cdot) := \text{Tr}_{B_2} \mathcal{N}_{A_1 A_2 \rightarrow B_1 B_2}((\cdot) \otimes \pi_{A_2}). \quad (4)$$

By the NS assumption, it can be easily verified that the definition in (4) is independent of the choice of π . Therefore, without loss of generality, we can always take π to be the maximally mixed state of appropriate dimension. See also discussions in ref. [68, General theory: one-shot catalytic resource distillation]. More generally, if the channel $\mathcal{N}_{A_1 A_2 \rightarrow B_1 B_2}$ does not satisfy the NS condition, we can still define a linear map via (4), which is clearly a quantum channel from $A_1 \rightarrow B_1$. The only difference is that the channel will depend on the choice of π .

Define the error between quantum channels $\mathcal{N}$ and $\mathcal{M}$ to be $\Delta(\mathcal{N}, \mathcal{M}) := \frac{1}{2} \|\mathcal{N} - \mathcal{M}\|_{\diamond}$ where $\|\mathcal{E}\|_{\diamond} := \sup_{k \in \mathbb{N}} \sup_{\|X\|_1 \leq 1} \|\mathcal{E} \otimes \mathcal{I}_k(X)\|_1$ denotes the diamond norm⁶⁹. We formally characterize catalytic channel transformations as follows.

Definition 9. Let $\mathcal{N} \in \text{CTP}(A \rightarrow B)$ and $\mathcal{M} \in \text{CTP}(A' \rightarrow B')$ be two quantum channels. We say $\mathcal{N}$ can be catalytically transformed to $\mathcal{M}$ with target error ε if there exists a free superchannel Π and a channel $\mathcal{C} \in \text{CTP}(X \rightarrow Y)$ such that the output channel $\mathcal{P} := \Pi_{\mathcal{N}}[\mathcal{N} \otimes \mathcal{C}_n] \in \text{CTP}(A'X \rightarrow B'Y)$ satisfies

$$\mathcal{P}_{X \rightarrow Y} = \mathcal{C}_{X \rightarrow Y}, \text{ and } \Delta(\mathcal{P}_{A'X \rightarrow B'Y}, \mathcal{M}_{A' \rightarrow B'} \otimes \mathcal{C}_{X \rightarrow Y}) \leq \varepsilon, \quad (5)$$

where $\mathcal{P}_{X \rightarrow Y}(\cdot) := \text{Tr}_{B'} \mathcal{P}_{A'X \rightarrow B'Y}((\cdot) \otimes \pi_{A'})$. Moreover, we say $\mathcal{N}$ can be (asymptotically) catalytically transformed to $\mathcal{M}$ if there exists a sequence of superchannels Π_n and a sequence of catalyst channel $\mathcal{C}_n$ such that $\mathcal{N}^{\otimes n}$ can be catalytically transformed to $\mathcal{M}^{\otimes n}$ with error ε_n and $\lim_{n \rightarrow \infty} \varepsilon_n = 0$.

The first condition in (5) implies that after the coding strategy Π , the transmission scheme $\mathcal{P}_{X \rightarrow Y}(\cdot)$ functions the same as the channel $\mathcal{C}$. That is, the functioning of $\mathcal{C}$ is not affected by the coding, thus it serves as a catalyst channel. The second condition in (5) implies that the correlation between the source channel and the catalyst channel is within a target error. This particularly implies that the reduced channel $\mathcal{P}_{A' \rightarrow B'}(\cdot) := \text{Tr}_Y \mathcal{P}_{A'X \rightarrow B'Y}((\cdot) \otimes \pi_X)$ will function approximately as $\mathcal{M}$, in the sense that $\Delta(\mathcal{P}_{A' \rightarrow B'}, \mathcal{M}_{A' \rightarrow B'}) \leq \varepsilon$ (see Lemma 13 in “Methods”).

With the above definition, we are now ready to discuss quantum channel simulation. By the well-known quantum reverse Shannon theorem⁶⁰, we know that channel manipulation under entanglement-assisted or NS-assisted codes is reversible. This, in particular, implies that $\mathcal{N}$ can be asymptotically transformed to $\mathcal{M}$ with unit rate if and only if $I(\mathcal{N}) \geq I(\mathcal{M})$, where $I(\mathcal{N}) := \sup_{\mathcal{R}} I(B : R)_{\mathcal{N}_{A \rightarrow B}(\mathcal{R}_A)}$ is the quantum channel mutual information and $I(B : R)_{\rho} = S(\rho_B) + S(\rho_R) - S(\rho_{BR})$ with $S(\rho) = -\text{Tr} \rho \log \rho$, giving an operational meaning of channel mutual information in the conventional i.i.d. setting.

Extending the catalysis technique for quantum states, we obtain the following result, which shows that $\mathcal{N}$ can be catalytically transformed to one $\mathcal{M}$ if and only if $I(\mathcal{N}) \geq I(\mathcal{M})$ given free entanglement and catalyst channels, thereby addressing the channel version of the

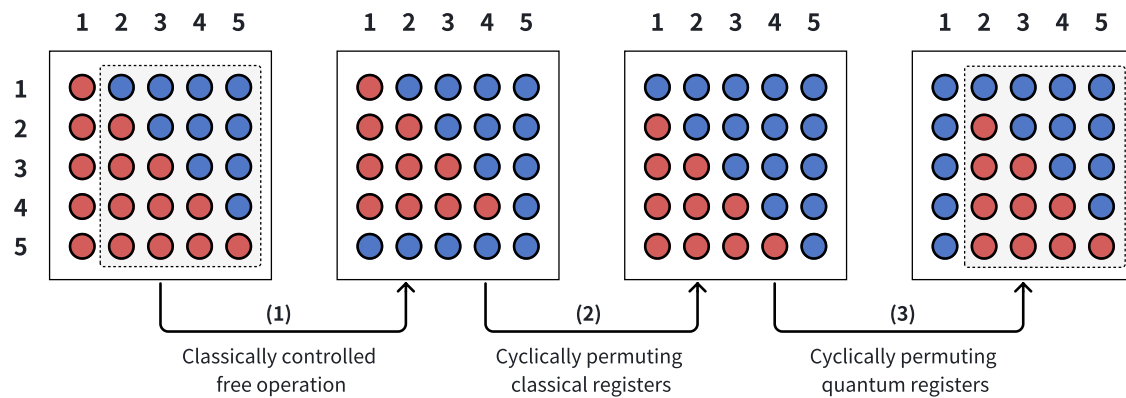


Fig. 3 | An illustration for the steps of our catalytic procedure used to prove Theorem 3. In this example, $n = 15$, $m = 5$, and hence $k = 3$. Each dot represents a quantum state on the system S^3 . Red dots correspond to groups of states $\zeta = \rho^{\otimes 3} \in \mathcal{D}(S^3)$, while blue dots represent embedded states $\hat{\sigma} = \sigma \otimes \pi^{\otimes 2} \in \mathcal{D}(S^3)$ that match the system of ζ . The overall quantum state is a mixture of rows, each labeled with a classical register. Suppose we have a 15-to-5 distillation protocol. In the first step, a classically controlled free operation is applied, using this protocol to

transform the last row from $\zeta^{\otimes 5} = \rho^{\otimes 15}$ to $\hat{\sigma}^{\otimes 5}$. The second step involves cyclically permuting the classical registers (i.e., the rows of dots), and the third step involves cyclically permuting the quantum registers (i.e., the columns of dots). Overall, this gives a 3-to-1 catalytic distillation protocol from $\zeta = \rho^{\otimes 3}$ to $\hat{\sigma}$ (which is equivalent to σ by tracing out the free states $\pi^{\otimes 2}$). The dashed boxes highlight the catalyst state, showing that it remains unchanged before and after the transformation.

open question raised in ref. 33 regarding whether mutual information determines the convertibility of quantum channels.

Theorem 10. Let $\mathcal{N}, \mathcal{M} \in \text{CPTP}(A \rightarrow B)$ be two quantum channels. Then $\mathcal{N}$ can be catalytically transformed to $\mathcal{M}$ with entanglement-assisted or NS-assisted codes if and only if $I(\mathcal{N}) \geq I(\mathcal{M})$.

This endows the channel mutual information with an operational meaning in terms of catalytic transformation at the one-shot level, enriching our understanding of quantum catalysis from a fundamental information-theoretic perspective.

Discussion

In this work, we considered quantum resource distillation with the assistance of catalysis and presented universal protocols with rigorous efficiency and reusability guarantees that surpass known fundamental limitations of one-shot distillation overhead. Our results highlight the utility and power of catalytic methods in enabling flexible configuration of different types of resources in quantum information processing tasks, which has strong practical appeal. Specifically, in the context of magic state distillation, our methods provide a practically valuable route for optimizing the cost of fault-tolerant quantum computing, which addresses the overhead or batch size problems of existing protocols.

There remain various subtle issues that may arise in practical implementations, including noise and correlation effects on the catalysts, which warrant further investigation and optimization. An important step forward is to study the benefits and challenges in implementing our catalytic distillation protocols with regard to specific experimental setups. These protocols are anticipated to be highly experiment-friendly due to the simplicity of required operations, specifically, classically controlled operations and relabeling or possibly collective movements that are e.g., natively efficient with the reconfigurable atom array platform, which emerged as a promising candidate for experimental fault-tolerant quantum computing recently⁷⁰ with unique advantages in overcoming the crucial non-Clifford operation barrier⁷¹.

Regarding the fundamental theory of catalytic distillation, an important avenue for further study is to relax the requirement of exact catalyst preservation and instead allow the catalyst to be returned only approximately, in the spirit of entanglementembezzling^{72–75}. We expect that this more flexible setting would allow further improvements of specific parameters of catalytic protocols, including the overhead constant and the size of the catalyst for a given target error.

In addition, while the catalysts used in this work generally have a size comparable to the system in multi-shot protocols, necessitating coherent manipulations of large quantum systems, the prospect of designing smaller, more efficient catalysts remains a valuable possibility. Conversely, the fundamental bounds on the catalyst size under certain target parameters are also compelling problems to further investigate (see Applications in magic state distillation for relevant discussion).

Moreover, given the connection between magic state distillation and quantum error-correcting (QEC) codes, our results may also provide new perspectives on catalytic QEC^{76,77}. Finally, extending these results to continuous variable systems and conducting a systematic study of channel theory are interesting directions for future work.

Methods

Here, we provide full proofs and technical analysis of the main results.

Proof of Theorem 3

The second inequality follows from their definitions. We now prove the first inequality by constructing an explicit catalytic transformation with the desired performance. As mentioned, the catalyst structure we use here was originally proposed in ref. 50 in the context of LOCC transformations of quantum entanglement. As mentioned, while similar catalyst structures have been used in various contexts recently^{30,32,33,38–42}, existing works mostly consider the transformation of source states to an equal number of target states. Here we adapt the strategy to the distillation of general resource types with arbitrary numbers of input and output states, to show that any multi-shot distillation protocol that transforms $\rho^{\otimes n}$ to $\sigma^{\otimes m}$ can be turned into a one-shot catalytic protocol that transforms $\rho^{[n/m]} \otimes \omega$ to $\sigma \otimes \omega$ with the same parameters (which then implies the inequalities in the theorem statement). In the proof, we present the procedure in a general and rigorous manner, and provide an illustrative example of its steps in Fig. 3. Consider any protocol $\mathcal{L} \in \mathcal{O}_{\text{sub}}$ such that $\mathcal{L}(\rho^{\otimes n}) = p\eta^m$ and $\Delta(\eta_i^m, \sigma) \leq \varepsilon$ for all $i \in \{1, \dots, m\}$, where $\eta^m \in \mathcal{D}(S^m)$ and η_i^m is the i -th marginal of η^m . For our purposes, we prepare m blocks of quantum states, each block containing $k = \lceil n/m \rceil$ copies of ρ . We denote a full block by $\zeta = \rho^{\otimes k}$. Since $mk = m\lceil n/m \rceil \geq n$, we have a free operation $\mathcal{L}_1$ such that

$$\mathcal{L}_1(\zeta^{\otimes m}) = \mathcal{L}_1(\rho^{\otimes mk}) = p\eta^m \text{ and } \Delta(\eta_i^m, \sigma) \leq \varepsilon, \quad (6)$$

for all $i \in \{1, \dots, m\}$. This can be done by simply discarding the residual $mk - n$ copies of the source states and then performing the transformation $\mathcal{L}$ on $\rho^{\otimes n}$. Note that η_m and $\sigma^{\otimes m}$ live in $\mathcal{D}(S^m)$, but we can embed them into a larger space $\mathcal{D}((S^k)^m)$ so that the individual subsystem matches the space of ζ . More precisely, consider embedding $\mathcal{E}(\cdot) = (\cdot) \otimes \pi_S^{\otimes(k-1)}$ where $\pi \in \mathcal{F}$ is a free state, and let $\hat{\eta}^m = \mathcal{E}^{\otimes m}(\eta^m)$ and $\hat{\sigma} = \mathcal{E}(\sigma)$. Consider $\mathcal{L}_2 = \mathcal{E} \circ \mathcal{L}_1$ we get a transformation

$$\mathcal{L}_2(\zeta^{\otimes m}) = p\hat{\eta}^m \text{ and } \Delta(\hat{\eta}_i^m, \hat{\sigma}) \leq \varepsilon, \quad (7)$$

where $\hat{\eta}_i^m$ is the marginal state on the i -th system and the inequality follows since trace distance is invariant under our embedding.

Now, we construct a catalyst ω to transform ζ to $\hat{\sigma}$ within the given error. Let $\hat{\eta}_{1,i}^m$ be the marginal state of $\hat{\eta}_m$ on the first i systems. Consider the catalyst

$$\omega := \frac{1}{m} \sum_{i=1}^m \zeta^{\otimes i-1} \otimes \hat{\eta}_{1,m-i}^m \otimes |i\rangle\langle i|, \quad (8)$$

where $|i\rangle\langle i| \in \mathcal{D}(F)$ are states in a classical flag register F . This gives an overall state

$$\zeta \otimes \omega = \frac{1}{m} \sum_{i=1}^m \zeta^{\otimes i} \otimes \hat{\eta}_{1,m-i}^m \otimes |i\rangle\langle i|. \quad (9)$$

We now aim to use this catalyst to complete the desired transformation. Let $\mathcal{P}_0(\cdot) := P_0(\cdot)P_0$ with $P_0 := \sum_{i=1}^{m-1} |i\rangle\langle i|$ and $\mathcal{P}_1(\cdot) := P_1(\cdot)P_1$ with $P_1 := |m\rangle\langle m|$ be two projecting maps on the flag register F . First, performing a classically controlled transformation $\mathcal{I} \otimes \mathcal{P}_0 + \frac{1}{p}\mathcal{L}_2 \otimes \mathcal{P}_1$ on $\zeta \otimes \omega$, we get

$$\nu_1 := \left(\mathcal{I} \otimes \mathcal{P}_0 + \frac{1}{p}\mathcal{L}_2 \otimes \mathcal{P}_1 \right) (\zeta \otimes \omega) \quad (10)$$

$$= \frac{1}{m} \sum_{i=1}^{m-1} \zeta^{\otimes i} \otimes \hat{\eta}_{1,m-i}^m \otimes |i\rangle\langle i| + \frac{1}{m} \hat{\eta}^m \otimes |m\rangle\langle m|, \quad (11)$$

where $\mathcal{I}$ is an identity map. This succeeds with probability p . Note that this transformation is simply a one-way LOCC operation⁷⁸, where we measure the classical flag register and feed forward the classical outcome to adaptively apply the quantum operation $\mathcal{I}$ or $\mathcal{L}_2$. Therefore, this classically controlled operation is clearly free. Second, by cyclically permuting the classical registers of ν_1 such that $i \rightarrow i+1$ and $m \rightarrow 1$, we get

$$\nu_2 := \frac{1}{m} \sum_{i=1}^m \zeta^{\otimes i-1} \otimes \hat{\eta}_{1,m-i+1}^m \otimes |i\rangle\langle i|. \quad (12)$$

Third, permuting the quantum registers in a similar way, such that $i \rightarrow i+1$ and $m \rightarrow 1$, we get the resulting state ν_3 . Note that ν_3 and ν_2 are the same state but with different system orders. Finally, note that the embedding operation is reversible by removing the ancillary registers. Denote $\mathcal{E}^{-1}(\cdot) = \text{Tr}_{S^{k-1}}(\cdot)$ to be the reverse embedding operation, and $\nu := \mathcal{E}^{-1} \otimes \mathcal{I}^{\otimes m-1}(\nu_3)$ be the final state after reverse embedding.

After these transformations, we can show that (i) the marginal state on the first system of ν is our target state satisfying the target error; (ii) the marginal state on the last $m-1$ systems of ν is exactly the catalyst ω . For the first claim, the marginal state on the first system of ν (or equivalently, the marginal state on the last system of ν_2 after reverse embedding) is given by

$$\nu' := \mathcal{E}^{-1} \left(\frac{1}{m} \sum_{i=1}^m \hat{\eta}_i^m \right) = \frac{1}{m} \sum_{i=1}^m \eta_i^m. \quad (13)$$

Since $\Delta(\eta_i^m, \sigma) \leq \varepsilon$ we have $\Delta(\nu', \sigma) \leq \varepsilon$ by the convexity of trace distance. The second claim is evident by checking that the first $m-1$ systems of ν_2 gives ω in Eq. (8). This gives a one-shot catalytic distillation protocol with the same performance in target error ε and success probability p , thereby completing the proof.

Proof of Theorem 4

Let $T = S^k$ be the systems on the source state and $\zeta_T = \rho_S^{\otimes k}$. Let $\mathcal{L}$ be the catalytic transformation. In the first round, we get $\mathcal{L}_{TA \rightarrow S_1 A}(\zeta_T \otimes \omega_A) = \nu_{S_1 A}$ with $\Delta(\nu_{S_1}, \sigma_{S_1}) \leq \varepsilon$, and $\nu_A = \omega_A$. Then we apply $\mathcal{L}$ again to a fresh copy of the source state ζ_T together with the catalyst and leave the state on the system S_1 untouched. Denote the output systems as $S_2 A$ where S_2 is isomorphic to S_1 . This gives the global state $\mathcal{L}_{TA \rightarrow S_2 A}(\zeta_T \otimes \nu_{S_1 A})$. Then we can check the catalyst on the system A by $\text{Tr}_{S_2 S_1} \mathcal{L}_{TA \rightarrow S_2 A}(\zeta_T \otimes \nu_{S_1 A}) = \text{Tr}_{S_2} \mathcal{L}_{TA \rightarrow S_2 A}(\zeta_T \otimes \omega_A) = \omega_A$. So the marginal state on the catalytic system is unchanged. We can also check the target state on the system S_2 by $\text{Tr}_{S_1 A} \mathcal{L}_{TA \rightarrow S_2 A}(\zeta_T \otimes \nu_{S_1 A}) = \text{Tr}_A \mathcal{L}_{TA \rightarrow S_2 A}(\zeta_T \otimes \omega_A) = \text{Tr}_A \mathcal{L}_{TA \rightarrow S_1 A}(\zeta_T \otimes \omega_A) = \nu_{S_1}$, which means that the marginal state we distill remains exactly the same as in the first round. The stated result is obtained by applying the above process repeatedly l times.

Proof of Lemma 5

Let $F(\rho, \sigma) = \text{Tr} \sqrt{\rho^{1/2} \sigma \rho^{1/2}}$ be the fidelity. Since $\Delta(\nu_S, |\phi\rangle\langle\phi|_S) \leq \varepsilon$, we have the relation that $F(\nu_S, |\phi\rangle\langle\phi|_S) \geq \sqrt{1-\varepsilon}$. As ν_{SA} is an extension of ν_S , by Uhlmann's theorem⁷⁹, there exists an extension of $|\phi\rangle\langle\phi|_S$ such that its fidelity with ν_{SA} is the same as $F(\nu_S, |\phi\rangle\langle\phi|_S)$. However, since $|\phi\rangle\langle\phi|_S$ is a pure state, its extension can only be in the form of a tensor product. Therefore, we can assume its extension as $|\phi\rangle\langle\phi|_S \otimes \mu_A$ such that $F(\nu_{SA}, |\phi\rangle\langle\phi|_S \otimes \mu_A) = F(\nu_S, |\phi\rangle\langle\phi|_S)$. By the data processing inequality of fidelity, we get $F(\nu_A, \mu_A) \geq F(\nu_{SA}, |\phi\rangle\langle\phi|_S \otimes \mu_A) \geq \sqrt{1-\varepsilon}$. This implies that $\Delta(\nu_{SA}, |\phi\rangle\langle\phi|_S \otimes \mu_A) \leq \sqrt{\varepsilon}$ and $\Delta(\nu_A, \mu_A) \leq \sqrt{\varepsilon}$. By the triangle inequality of trace distance, we have $\Delta(\nu_{SA}, |\phi\rangle\langle\phi|_S \otimes \mu_A) + \Delta(|\phi\rangle\langle\phi|_S \otimes \mu_A, |\phi\rangle\langle\phi|_S \otimes \nu_A) \leq 2\sqrt{\varepsilon}$.

Proof of Theorem 7

Let $\mathcal{L} \in \mathcal{O}_{\text{sub}}$ be a resource distillation protocol that transforms n copies of the source state ρ_S to m ($m \leq n$) copies of the target state σ_S with success probability p and within target error ε . That is, there exists a quantum state $\eta^m \in \mathcal{D}(S^m)$, such that $\mathcal{L}(\rho^{\otimes n}) = p\eta^m$ and $\Delta(\eta_i^m, \sigma) \leq \varepsilon$ for all $i \in \{1, \dots, m\}$, where η_i^m is the i -th marginal of η^m . Then we can consider $g = \lceil n/k \rceil \geq m$ groups of $\zeta = \rho^{\otimes k}$. Since $gk = k \lceil n/k \rceil \geq n$, there exists a free operation $\mathcal{L}_1$ such that $\mathcal{L}_1(\zeta^{\otimes g}) = \mathcal{L}_1(\rho^{\otimes gk}) = p\eta^m$ and $\Delta(\eta_i^m, \sigma) \leq \varepsilon$, for any $i \in \{1, \dots, m\}$. This can be done by simply abandoning the residual $gk - n$ copies of the source states and then performing the transformation $\mathcal{L}$ on $\rho^{\otimes n}$. Then we need to prove that there exists a catalytic transformation $\mathcal{L}'$ that transforms one copy of ζ to one copy of σ with success probability pm/g and within target error ε . An illustration for the steps of the procedure is given in Fig. 4.

Before the actual transformation, we need to perform two embedding steps to ensure that the systems match. First, as $\zeta \in \mathcal{D}(S^k)$ and $\sigma \in \mathcal{D}(S)$, we need to embed σ into a larger space by using the embedding $\mathcal{E}(\cdot) = (\cdot) \otimes \pi_S^{\otimes(k-1)}$ where $\pi \in \mathcal{F}$ is a free state. Let $\hat{\eta}^m = \mathcal{E}^{\otimes m}(\eta^m)$ and $\hat{\sigma} = \mathcal{E}(\sigma)$. Then, since $m \leq g$, we need to do a second embedding to compensate this system mismatching by using free states. Moreover, to avoid mixing the free state with our target state, which could compromise the final fidelity, we need to append the free states on an orthogonal Hilbert space so that we can distinguish them via projective measurements later. More explicitly, let θ be a free state on the system S^k . We embed it into a joint Hilbert space $T = S^k W$ by $\theta_T = \theta_{S^k} \otimes |1\rangle\langle 1|_W$. We then embed $\zeta, \hat{\sigma} \in \mathcal{D}(S^k)$ into the larger space T by $\tilde{\zeta}_T = \zeta_{S^k} \otimes |0\rangle\langle 0|_W$ and $\tilde{\sigma}_T = \hat{\sigma}_{S^k} \otimes |0\rangle\langle 0|_W$, respectively. Similarly, we embed $\hat{\eta}^m$ into $\mathcal{D}(T^m)$ by $\tilde{\eta}^m = \hat{\eta}^m \otimes |0\rangle\langle 0|_{W^m}$.

Then there exists a distillation protocol $\mathcal{L}_2$ that transforms $\tilde{\zeta}^{\otimes g} \in \mathcal{D}(T^g)$ to $\tilde{\eta}_m \otimes \tilde{\theta}^{\otimes g-m} \in \mathcal{D}(T^g)$. Let $\tilde{\eta}_{1,i}^m$ be the marginal state of $\tilde{\eta}_m$ on

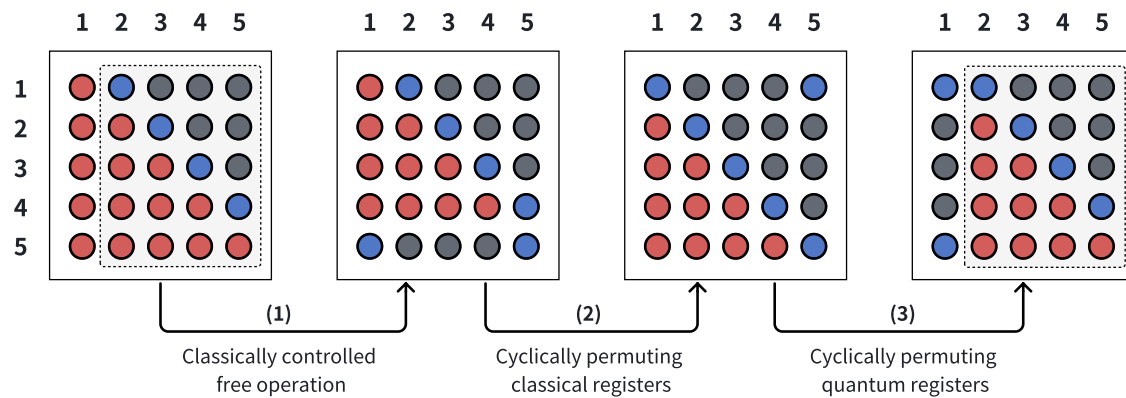


Fig. 4 | An illustration for the steps of our catalytic procedure used to prove Theorem 7, for the case where $n = 5$, $m = 2$, and $k = 1$. Each dot represents a quantum state on the system T . Red dots correspond to the embedded source state $\tilde{\rho} \in \mathcal{D}(T)$, while blue dots represent the embedded target state $\tilde{\sigma} \in \mathcal{D}(T)$. Gray dots represent the embedded free state $\tilde{\theta} \in \mathcal{D}(T)$, which is orthogonal to $\tilde{\sigma}$. The overall quantum state is a mixture of rows, each labeled with a classical register. Suppose we have a 5-to-2 distillation protocol with success probability p . In the first step, a classically controlled free operation is applied, transforming the last row

from $\tilde{\rho}^{\otimes 5}$ to $\tilde{\sigma}^{\otimes 2} \otimes \tilde{\theta}^{\otimes 3}$ with success probability p . The second step involves cyclically permuting the classical registers (i.e., the rows of dots), and the third step involves cyclically permuting the quantum registers (i.e., the columns of dots). Since $\tilde{\sigma}$ is orthogonal to $\tilde{\theta}$, we can perform projective measurement and post-select $\tilde{\sigma}$ with success probability $2/5$. Overall, this gives a 1-to-1 catalytic distillation protocol from ρ to σ with success probability $2p/5$. The dashed boxes highlight the catalyst state, showing that it remains unchanged before and after the transformation.

the first i systems and $\tilde{\eta}_i^m$ be the marginal state on the i -th system. Define the following catalyst state on $\mathcal{D}(T^{g-1}F)$ with classical register F :

$$\omega := \frac{1}{g} \sum_{i=1}^{g-m-1} \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_i^m \otimes \tilde{\theta}^{\otimes g-m-i} \otimes |i\rangle\langle i| + \frac{1}{g} \sum_{i=g-m}^g \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_{1,g-i}^m \otimes |i\rangle\langle i|. \quad (14)$$

Note that for $m = g$ or $m = g - 1$ the first term vanishes. We have the overall state on $\mathcal{D}(T^g F)$,

$$\tilde{\zeta} \otimes \omega = \frac{1}{g} \sum_{i=1}^{g-m-1} \tilde{\zeta}^{\otimes i} \otimes \tilde{\eta}_i^m \otimes \tilde{\theta}^{\otimes g-m-i} \otimes |i\rangle\langle i| + \frac{1}{g} \sum_{i=g-m}^g \tilde{\zeta}^{\otimes i} \otimes \tilde{\eta}_{1,g-i}^m \otimes |i\rangle\langle i|. \quad (15)$$

$= |g\rangle\langle g|$ be two projecting maps on the flag register F . We then proceed with the following steps. First, performing a classically controlled transformation $\mathcal{I} \otimes \mathcal{P}_0 + \frac{1}{p} \mathcal{L}_2 \otimes \mathcal{P}_1$ on $\tilde{\zeta} \otimes \omega$, we obtain

$$\nu_1 := \left(\mathcal{I} \otimes \mathcal{P}_0 + \frac{1}{p} \mathcal{L}_2 \otimes \mathcal{P}_1 \right) (\tilde{\zeta} \otimes \omega) \quad (16)$$

$$= \frac{1}{g} \sum_{i=1}^{g-m-1} \tilde{\zeta}^{\otimes i} \otimes \tilde{\eta}_i^m \otimes \tilde{\theta}^{\otimes g-m-i} \otimes |i\rangle\langle i| \quad (17)$$

$$+ \frac{1}{g} \sum_{i=g-m}^g \tilde{\zeta}^{\otimes i} \otimes \tilde{\eta}_{1,g-i}^m \otimes |i\rangle\langle i| + \frac{1}{g} \tilde{\eta}^m \otimes \tilde{\theta}^{\otimes g-m} \otimes |g\rangle\langle g|. \quad (18)$$

As explained in the proof of Theorem 3, this procedure is a free operation. Second, by cyclically permuting the classical registers of ν_1 such that $i \rightarrow i + 1$ and $g \rightarrow 1$, we get

$$\nu_2 := \frac{1}{g} \sum_{i=1}^{g-m} \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_i^m \otimes \tilde{\theta}^{\otimes g-m-i+1} \otimes |i\rangle\langle i| + \frac{1}{g} \sum_{i=g-m+1}^g \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_{1,g-i+1}^m \otimes |i\rangle\langle i|. \quad (19)$$

Third, permuting the quantum registers in a similar way, such that $i \rightarrow i + 1$ and $g \rightarrow 1$, we get the resulting state ν_3 . Note that ν_3 and ν_2 are the same state but with different system orders. Then we can check that the last $g - 1$ quantum systems of ν_3 returns the catalyst, or equivalently, we can check the first $g - 1$ systems of ν_2 and get

$$\frac{1}{g} \sum_{i=1}^{g-m} \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_i^m \otimes \tilde{\theta}^{\otimes g-m-i} \otimes |i\rangle\langle i| + \frac{1}{g} \sum_{i=g-m+1}^g \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_{1,g-i}^m \otimes |i\rangle\langle i| = \omega. \quad (20)$$

The first quantum system of ν_3 or equivalently the last quantum system of ν_2 gives our target state:

$$\nu' := \frac{1}{g} \sum_{i=1}^{g-m} \tilde{\theta} + \frac{1}{g} \sum_{i=g-m+1}^g \tilde{\eta}_{1,g-i+1}^m = \frac{m}{g} \nu'' + \left(1 - \frac{m}{g} \right) \tilde{\theta}, \text{ with } \nu'' := \frac{1}{m} \sum_{i=1}^m \tilde{\eta}_i^m. \quad (21)$$

Since $\nu'' = \frac{1}{m} \sum_{i=1}^m \tilde{\eta}_i^m = \frac{1}{m} \sum_{i=1}^m \tilde{\eta}_i^m \otimes |0\rangle\langle 0|_W$ and $\tilde{\theta} = \theta_{S^k} \otimes |1\rangle\langle 1|_W$ are orthogonal, we can perform a projective measurement on the flag system W to distinguish these two states. So we will have a probability m/g to obtain the state ν' . Finally, as all appended states are uncorrelated to the target system, we get $\nu''' = \mathcal{E}^{-1}(\text{Tr}_W \nu'') = \frac{1}{m} \sum_{i=1}^m \tilde{\eta}_i^m$ by removing all the appended systems for free. Since $\Delta(\tilde{\eta}_i^m, \sigma) \leq \varepsilon$, we have $\Delta(\nu''', \sigma) \leq \varepsilon$ by the convexity of trace distance. This concludes the proof.

Some remarks concerning our above result are in order. First, the catalyst construction extends beyond the standard ones in refs. 30,32,33,38–42,50. Appending the free state θ on the orthogonal subspace is crucial here, as it allows the target state to be post-selected after the projective measurement, ensuring that the fidelity is not compromised in the final step. Second, the catalyst used in the proof is not unique. Other options could also work in the same manner, such as

$$\omega = \frac{1}{g} \sum_{i=1}^{g-m} \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_{1,m-1}^m \otimes \tilde{\theta}^{\otimes g-m-i+1} \otimes |i\rangle\langle i| + \frac{1}{g} \sum_{i=g-m+1}^g \tilde{\zeta}^{\otimes i-1} \otimes \tilde{\eta}_{1,g-i}^m \otimes |i\rangle\langle i|. \quad (22)$$

Interestingly, such a catalyst is independent on the target state $\tilde{\eta}^m$ if $m = 1$. Third, if $p = 1$, the catalyst is always returned on the marginal before the post-selection on the target system. So its status is preserved by ignoring the measurement outcome and therefore can be reused in the subsequent distillation rounds.

Proof of Theorem 10

The proof of this result requires the following lemmas.

Lemma 11. For any quantum channel $\mathcal{N}$ and superchannel Π , the output channel $\Pi(\mathcal{N})$ satisfies $I(\Pi(\mathcal{N})) \leq I(\mathcal{N})$.

Proof. This can be understood from the existing operational interpretation of a channel's mutual information, which represents its entanglement-assisted quantum capacity. Since the channel $\Pi(\mathcal{N})$ is noisier than $\mathcal{N}$, it results in reduced communication capability. A similar argument is provided in ref. [65, Remark 5]. $\square$

Moreover, the quantum channel mutual information is continuous with respect to the diamond norm.

Lemma 12. Let $\mathcal{N}, \mathcal{M} \in \text{CPTP}(A \rightarrow B)$ and d_{AB} be the dimension of $A \otimes B$. If $\|\mathcal{N} - \mathcal{M}\|_{\diamond} \leq \varepsilon$, it holds that

$$|I(\mathcal{N}) - I(\mathcal{M})| \leq 3\varepsilon \log d_{AB} + 3h_2(\varepsilon) =: f(\varepsilon). \quad (23)$$

Proof. For any $\delta > 0$, choose φ_1 and φ_2 be such that $|I(\mathcal{N}) - I(B : R)_{\mathcal{N}(\varphi_1)}| \leq \delta$ and $|I(\mathcal{M}) - I(B : R)_{\mathcal{M}(\varphi_2)}| \leq \delta$ where $d_R = d_A$. Since $\|\mathcal{N} - \mathcal{M}\|_{\diamond} \leq \varepsilon$, we have $\|\mathcal{N}(\varphi_1) - \mathcal{M}(\varphi_1)\|_1 \leq \varepsilon$ and $\|\mathcal{N}(\varphi_2) - \mathcal{M}(\varphi_2)\|_1 \leq \varepsilon$. Recall the Fannes inequality $|S(\rho) - S(\psi)| \leq \varepsilon \log d + h_2(\varepsilon)$ if $\|\rho - \psi\|_1 \leq \varepsilon$. We have the continuity of the quantum mutual information as

$$|I(B : R)_{\rho} - I(B : R)_{\psi}| \leq |S(\rho_B) - S(\psi_B)| + |S(\rho_R) - S(\psi_R)| + |S(\rho_{BR}) - S(\psi_{BR})| \quad (24)$$

$$\leq 3\varepsilon \log d_{AB} + 3h_2(\varepsilon) =: f(\varepsilon). \quad (25)$$

if $\|\rho_{BR} - \psi_{BR}\|_1 \leq \varepsilon$. Putting everything together, we have

$$I(\mathcal{N}) - \delta \leq I(B : R)_{\mathcal{N}(\varphi_1)} \leq I(B : R)_{\mathcal{M}(\varphi_1)} + f(\varepsilon) \leq I(\mathcal{M}) + f(\varepsilon), \quad (26)$$

$$I(\mathcal{M}) - \delta \leq I(B : R)_{\mathcal{M}(\varphi_2)} \leq I(B : R)_{\mathcal{N}(\varphi_2)} + f(\varepsilon) \leq I(\mathcal{N}) + f(\varepsilon). \quad (27)$$

So we get $|I(\mathcal{N}) - I(\mathcal{M})| \leq \delta + f(\varepsilon)$. Since δ can be made arbitrarily small, we get $|I(\mathcal{N}) - I(\mathcal{M})| \leq f(\varepsilon)$. $\square$

Lemma 13. Let $\mathcal{P} \in \text{CPTP}(A'X \rightarrow B'Y)$, $\mathcal{M} \in \text{CPTP}(A' \rightarrow B')$, and $\mathcal{C} \in \text{CPTP}(X \rightarrow Y)$. Let $\mathcal{P}_{A' \rightarrow B'}(\cdot) := \text{Tr}_Y \mathcal{P}_{A'X \rightarrow B'Y}(\cdot \otimes \pi_X)$ be the reduced channel of $\mathcal{P}$ on the system A' . If $\Delta(\mathcal{P}_{A'X \rightarrow B'Y}, \mathcal{M}_{A' \rightarrow B'} \otimes \mathcal{C}_{X \rightarrow Y}) \leq \varepsilon$, then $\Delta(\mathcal{P}_{A' \rightarrow B'}, \mathcal{M}_{A' \rightarrow B'}) \leq \varepsilon$.

Proof. This can be checked as follows:

$$\begin{aligned} & \|\mathcal{P}_{A' \rightarrow B'} - \mathcal{M}_{A' \rightarrow B'}\|_{\diamond} \\ &= \sup_{\rho_{RA'}} \|\mathcal{P}_{A' \rightarrow B'}(\rho_{RA'}) - \mathcal{M}_{A' \rightarrow B'}(\rho_{RA'})\|_1 \end{aligned} \quad (28)$$

$$= \sup_{\rho_{RA'}} \|\text{Tr}_Y \mathcal{P}_{A'X \rightarrow B'Y}(\rho_{RA'} \otimes \pi_X) - \mathcal{M}_{A' \rightarrow B'}(\rho_{RA'})\|_1 \quad (29)$$

$$= \sup_{\rho_{RA'}} \|\text{Tr}_Y \{\mathcal{P}_{A'X \rightarrow B'Y}(\rho_{RA'} \otimes \pi_X) - \mathcal{M}_{A' \rightarrow B'}(\rho_{RA'}) \otimes \mathcal{C}_{X \rightarrow Y}(\pi_X)\}\|_1 \quad (30)$$

$$\leq \sup_{\rho_{RA'}} \|\mathcal{P}_{A'X \rightarrow B'Y}(\rho_{RA'} \otimes \pi_X) - \mathcal{M}_{A' \rightarrow B'}(\rho_{RA'}) \otimes \mathcal{C}_{X \rightarrow Y}(\pi_X)\|_1 \quad (31)$$

$$\leq \sup_{\rho_{RA'X}} \|\mathcal{P}_{A'X \rightarrow B'Y}(\rho_{RA'X}) - \mathcal{M}_{A' \rightarrow B'} \otimes \mathcal{C}_{X \rightarrow Y}(\rho_{RA'X})\|_1 \quad (32)$$

$$= \|\mathcal{P}_{A'X \rightarrow B'Y} - \mathcal{M}_{A' \rightarrow B'} \otimes \mathcal{C}_{X \rightarrow Y}\|_{\diamond}, \quad (33)$$

where the first inequality follows by the data-processing inequality of trace norm and the second inequality follows by relaxing the input state to all quantum states on $RA'X$. $\square$

We are now ready to prove Theorem 10. We first show the “only if” direction. Suppose $\mathcal{N}$ can be catalytically transformed to $\mathcal{M}$. For any $\varepsilon > 0$, there exists a catalyst channel $\mathcal{C}_n$ and a code Π_n such that $\Delta(\Pi_n(\mathcal{N} \otimes \mathcal{C}_n), \mathcal{M} \otimes \mathcal{C}_n) \leq \varepsilon$. By the continuity of the channel mutual information and the monotonicity in Lemma 12 and Lemma 11, we get

$$I(\mathcal{M} \otimes \mathcal{C}_n) \leq I(\Pi_n(\mathcal{N} \otimes \mathcal{C}_n)) + f(2\varepsilon) \leq I(\mathcal{N} \otimes \mathcal{C}_n) + f(2\varepsilon). \quad (34)$$

Then by the additivity of channel mutual information, we get $I(\mathcal{M}) \leq I(\mathcal{N}) + f(2\varepsilon)$. Since the inequality holds for arbitrary ε , we get $I(\mathcal{M}) \leq I(\mathcal{N})$.

Then we prove the “if” direction. That is, if $I(\mathcal{N}) \geq I(\mathcal{M})$ we can find a sequence of Π_n and $\mathcal{C}_n$ that catalytically transform $\mathcal{N}$ to $\mathcal{M}$. By the reversibility of the channel simulation, we know that for any given ε , there exists entanglement-assisted or NS-assisted transformations such that

$$\Pi_n(\mathcal{N}^{\otimes n}) = \mathcal{P}^n \text{ and } \|\mathcal{P}^n - \mathcal{M}^{\otimes n}\|_{\diamond} \leq \varepsilon. \quad (35)$$

Note that $\mathcal{P}^n \in \text{CPTP}(A^n \rightarrow B^n)$. Define the reduced channel on the first i systems as

$$\mathcal{P}_{1:i}^n(\cdot) := \text{Tr}_{B_{i+1:n}} \mathcal{P}_{A^n \rightarrow B^n}^n(\cdot \otimes \pi_{A_{i+1:n}}), \quad (36)$$

where $\mathcal{P}_{1:n}^n := \mathcal{P}^n$. Now consider a catalyst channel

$$\mathcal{C}_n := \frac{1}{n} \sum_{k=1}^n \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k}^n \otimes |k\rangle\langle k|. \quad (37)$$

Then the overall channel is given by

$$\mathcal{N} \otimes \mathcal{C}_n = \frac{1}{n} \sum_{k=1}^n \mathcal{N}^{\otimes k} \otimes \mathcal{P}_{1:n-k}^n \otimes |k\rangle\langle k|. \quad (38)$$

Let $\mathcal{Q}_0(\cdot) := \mathcal{Q}_0(\cdot) \mathcal{Q}_0$ with $\mathcal{Q}_0 := \sum_{i=1}^{n-1} |i\rangle\langle i|$ and $\mathcal{Q}_1(\cdot) := \mathcal{Q}_1(\cdot) \mathcal{Q}_1$ with $\mathcal{Q}_1 := |n\rangle\langle n|$ be two projecting maps on the flag register F . First, performing a classically controlled transformation $\mathcal{I} \otimes \mathcal{Q}_0 + \Pi_n \otimes \mathcal{Q}_1$, we get

$$\mathcal{G}_1 = \frac{1}{n} \sum_{k=1}^{n-1} \mathcal{N}^{\otimes k} \otimes \mathcal{P}_{1:n-k}^n \otimes |k\rangle\langle k| + \frac{1}{n} \mathcal{P}^n \otimes |n\rangle\langle n|. \quad (39)$$

Next, we relabel the classical registers $i \rightarrow i+1$ and $n \rightarrow 1$ and get

$$\mathcal{G}_2 = \frac{1}{n} \sum_{k=1}^n \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k+1}^n \otimes |k\rangle\langle k|. \quad (40)$$

Then performing a cyclic permutation S_{A^n} and S_{B^n} on the input and the output, respectively, such that $i \rightarrow i+1$ and $n \rightarrow 1$, we get

$$\mathcal{G}_3 = \frac{1}{n} \sum_{k=1}^n S_{B^n} \circ (\mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k+1}^n) \circ S_{A^n} \otimes |k\rangle\langle k|. \quad (41)$$

Then we claim that $\mathcal{G}_3$ returns $\mathcal{C}_n$ on its marginal and $\|\mathcal{G}_3 - \mathcal{M} \otimes \mathcal{C}_n\|_{\diamond} \leq 2\varepsilon$. The first claim is equivalent to check the first $n-1$ reduced channel of $\mathcal{G}_2$,

$$\text{Tr}_{B_n} \circ \mathcal{G}_2((\cdot) \otimes \pi_{A_n}) = \frac{1}{n} \sum_{k=1}^n \text{Tr}_{B_n} \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k+1}^n((\cdot) \otimes \pi_{A_n}) \otimes |k\rangle\langle k| \quad (42)$$

$$= \frac{1}{n} \sum_{k=1}^n \mathcal{N}^{\otimes k-1} \otimes \text{Tr}_{B_n} \mathcal{P}_{1:n-k+1}^n((\cdot) \otimes \pi_{A_n}) \otimes |k\rangle\langle k| \quad (43)$$

$$= \frac{1}{n} \sum_{k=1}^n \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k}^n(\cdot) \otimes |k\rangle\langle k| \quad (44)$$

$$= \mathcal{C}_n. \quad (45)$$

Note that the second last equality holds by the definition of $\mathcal{P}_i^n$ and here π_{A_n} and Tr_{B_n} are effectively acting on the last systems of the channel $\mathcal{P}_{1:n-k+1}^n$, which is also the $n-k+1$ system in the original definition of $\mathcal{P}_{1:n-k+1}^n$. Since the diamond norm is invariant under system permutation, the second claim is equivalent to check $\|\mathcal{G}_2 - \mathcal{C}_n \otimes \mathcal{M}\|_{\diamond} \leq 2\varepsilon$. This can be shown as follows:

$$\begin{aligned} & \|\mathcal{G}_2 - \mathcal{C}_n \otimes \mathcal{M}\|_{\diamond} \\ &= \left\| \frac{1}{n} \sum_{k=1}^n \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k+1}^n \otimes |k\rangle\langle k| - \frac{1}{n} \sum_{k=1}^n \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k}^n \otimes \mathcal{M} \otimes |k\rangle\langle k| \right\|_{\diamond} \end{aligned} \quad (46)$$

$$= \frac{1}{n} \sum_{k=1}^n \left\| \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k+1}^n - \mathcal{N}^{\otimes k-1} \otimes \mathcal{P}_{1:n-k}^n \otimes \mathcal{M} \right\|_{\diamond}. \quad (47)$$

$$= \frac{1}{n} \sum_{k=1}^n \left\| \mathcal{P}_{1:n-k+1}^n - \mathcal{P}_{1:n-k}^n \otimes \mathcal{M} \right\|_{\diamond}. \quad (48)$$

$$\leq \frac{1}{n} \sum_{k=1}^n \left\| \mathcal{P}_{1:n-k+1}^n - \mathcal{M}^{\otimes n-k+1} \right\|_{\diamond} + \frac{1}{n} \sum_{k=1}^n \left\| \mathcal{P}_{1:n-k}^n \otimes \mathcal{M} - \mathcal{M}^{\otimes n-k+1} \right\|_{\diamond}. \quad (49)$$

$$= \frac{1}{n} \sum_{k=1}^n \left\| \mathcal{P}_{1:n-k+1}^n - \mathcal{M}^{\otimes n-k+1} \right\|_{\diamond} + \frac{1}{n} \sum_{k=1}^n \left\| \mathcal{P}_{1:n-k}^n - \mathcal{M}^{\otimes n-k} \right\|_{\diamond}. \quad (50)$$

$$\leq 2\varepsilon. \quad (51)$$

where the first inequality follows by the triangle inequality of diamond norm and the last inequality follows from the assumption of $\|\mathcal{P}^n - \mathcal{M}^{\otimes n}\|_{\diamond} \leq \varepsilon$ and the monotonicity of diamond norm by taking reduced channel in Lemma 13. As this holds for arbitrary ε , we can take ε to be vanishingly small and conclude the proof.

Data availability

No datasets were generated or analyzed during the current study.

References

- Horodecki, R., Horodecki, P., Horodecki, M. & Horodecki, K. Quantum entanglement. *Rev. Mod. Phys.* **81**, 865–942 (2009).
- Streltsov, A., Adesso, G. & Plenio, M. B. Colloquium: quantum coherence as a resource. *Rev. Mod. Phys.* **89**, 041003 (2017).
- Bravyi, S. & Kitaev, A. Universal quantum computation with ideal clifford gates and noisy ancillas. *Phys. Rev. A* **71**, 022316 (2005).
- Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
- Bennett, C. H. et al. Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722 (1996).
- Campbell, E. T. & Browne, D. E. Bound states for magic state distillation in fault-tolerant quantum computation. *Phys. Rev. Lett.* **104**, 030503 (2010).
- Bravyi, S. & Haah, J. Magic-state distillation with low overhead. *Phys. Rev. A* **86**, 052329 (2012).
- Hastings, M. B. & Haah, J. Distillation with sublogarithmic overhead. *Phys. Rev. Lett.* **120**, 050504 (2018).
- Krishna, A. & Tillich, J.-P. Towards low overhead magic state distillation. *Phys. Rev. Lett.* **123**, 070507 (2019).
- Liu, Z.-W., Bu, K. & Takagi, R. One-shot operational quantum resource theory. *Phys. Rev. Lett.* **123**, 020401 (2019).
- Fang, K. & Liu, Z.-W. No-go theorems for quantum resource purification. *Phys. Rev. Lett.* **125**, 060405 (2020).
- Fang, K. & Liu, Z.-W. No-go theorems for quantum resource purification: New approach and channel theory. *PRX Quantum* **3**, 010337 (2022).
- Liu, Z.-W. & Winter, A. Many-body quantum magic. *PRX Quantum* **3**, 020333 (2022).
- Wills, A., Hsieh, M.-H. & Yamasaki, H. Constant-overhead magic state distillation. *Nat. Phys.* **21**, 1842–1846 (2025).
- Golowich, L. & Guruswami, V. Asymptotically good quantum codes with transversal non-clifford gates. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2025).
- Nguyen, Q. T. Good binary quantum codes with transversal ccz gate. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2025).
- Lee, S.-H., Thomsen, F., Fazio, N., Brown, B. J. & Bartlett, S. D. Low-overhead magic state distillation with color codes. *PRX Quantum* **6**, 030317 (2025).
- Golowich, L. & Lin, T.-C. Quantum LDPC codes with transversal non-clifford gates via products of algebraic codes. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, USA, 2025).
- Bennett, C. H., Bernstein, H. J., Popescu, S. & Schumacher, B. Concentrating partial entanglement by local operations. *Phys. Rev. A* **53**, 2046–2052 (1996).
- Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824–3851 (1996).
- Fang, K., Wang, X., Tomamichel, M. & Duan, R. Non-asymptotic entanglement distillation. *IEEE Trans. Inf. Theory* **65**, 6454–6465 (2019).
- Winter, A. & Yang, D. Operational resource theory of coherence. *Phys. Rev. Lett.* **116**, 120404 (2016).
- Fang, K., Wang, X., Lami, L., Regula, B. & Adesso, G. Probabilistic distillation of quantum coherence. *Phys. Rev. Lett.* **121**, 070404 (2018).
- Regula, B., Fang, K., Wang, X. & Adesso, G. One-shot coherence distillation. *Phys. Rev. Lett.* **121**, 010401 (2018).
- Hayashi, M., Fang, K. & Wang, K. Finite block length analysis on quantum coherence distillation and incoherent randomness extraction. *IEEE Trans. Inf. Theory* **67**, 3926–3944 (2021).

26. Jonathan, D. & Plenio, M. B. Entanglement-assisted local manipulation of pure quantum states. *Phys. Rev. Lett.* **83**, 3566 (1999).
27. Nielsen, M. A. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.* **83**, 436–439 (1999).
28. Nielsen, M. A. et al. *An Introduction to Majorization and Its Applications to Quantum Mechanics*. (2002).
29. Aberg, J. Catalytic coherence. *Phys. Rev. Lett.* **113**, 150402 (2014).
30. Takagi, R. & Shiraishi, N. Correlation in catalysts enables arbitrary manipulation of quantum coherence. *Phys. Rev. Lett.* **128**, 240501 (2022).
31. Müller, M. P. Correlating thermal machines and the second law at the nanoscale. *Phys. Rev. X* **8**, 041051 (2018).
32. Shiraishi, N. & Sagawa, T. Quantum thermodynamics of correlated catalytic state conversion at small scale. *Phys. Rev. Lett.* **126**, 150502 (2021).
33. Wilming, H. Entropy and reversible catalysis. *Phys. Rev. Lett.* **127**, 260402 (2021).
34. Boes, P., Eisert, J., Gallego, R., Müller, M. P. & Wilming, H. Von neumann entropy from unitarity. *Phys. Rev. Lett.* **122**, 210402 (2019).
35. Campbell, E. T. Catalysis and activation of magic states in fault-tolerant architectures. *Phys. Rev. A* **83**, 032317 (2011).
36. Gidney, C. Halving the cost of quantum addition. *Quantum* **2**, 74 (2018).
37. Beverland, M., Campbell, E., Howard, M. & Kliuchnikov, V. Lower bounds on the non-clifford resources for quantum computations. *Quantum Sci. Technol.* **5**, 035009 (2020).
38. Char, P., Chakraborty, D., Bhar, A., Chattopadhyay, I. & Sarkar, D. Catalytic transformations in coherence theory. *Phys. Rev. A* **107**, 012404 (2023).
39. Datta, C., Ganardi, R., Kondra, T. V. & Streltsov, A. Is there a finite complete set of monotones in any quantum resource theory? *Phys. Rev. Lett.* **130**, 240204 (2023).
40. Kondra, T. V., Datta, C. & Streltsov, A. Catalytic transformations of pure entangled states. *Phys. Rev. Lett.* **127**, 150503 (2021).
41. Lipka-Bartosik, P. & Skrzypczyk, P. Catalytic quantum teleportation. *Phys. Rev. Lett.* **127**, 080502 (2021).
42. Bavaresco, J., Brunner, N., Girardin, A., Lipka-Bartosik, P. & Sekatski, P. Catalytic activation of bell nonlocality. *Phys. Rev. Lett.* – (2025).
43. Datta, C., Kondra, T. V., Miller, M. & Streltsov, A. Catalysis of entanglement and other quantum resources. *Rep. Prog. Phys.* **135**, 220203 (2023).
44. Lipka-Bartosik, P., Wilming, H. & Ng, N. H. Catalysis in quantum information theory. *Rev. Mod. Phys.* **96**, 025005 (2024).
45. Lami, L., Regula, B. & Streltsov, A. No-go theorem for entanglement distillation using catalysis. *Phys. Rev. A* **109**, <https://doi.org/10.1103/PhysRevA.109.L050401> (2024).
46. Ganardi, R., Kondra, T. V. & Streltsov, A. Catalytic and asymptotic equivalence for quantum entanglement. *Phys. Rev. Lett.* **133**, 250201 (2024).
47. Karvonen, M. Neither contextuality nor nonlocality admits catalysts. *Phys. Rev. Lett.* **127**, 160402 (2021).
48. Chitambar, E. & Gour, G. Quantum resource theories. *Rev. Mod. Phys.* **91**, 025001 (2019).
49. Takagi, R. & Regula, B. General resource theories in quantum mechanics and beyond: Operational characterization via discrimination tasks. *Phys. Rev. X* **9**, 031053 (2019).
50. Duan, R., Feng, Y., Li, X. & Ying, M. Multiple-copy entanglement transformation and entanglement catalysis. *Phys. Rev. A* **71**, 042319 (2005).
51. Heußen, S. Magic state distillation without measurements and post-selection. Preprint at <https://arxiv.org/abs/2504.17509> (2025).
52. Haah, J. & Hastings, M. B. Codes and protocols for distilling T , controlled- S , and toffoli gates. *Quantum* **2**, 71 (2018).
53. Meier, A. M., Eastin, B. & Knill, E. Magic-state distillation with the four-qubit code. *Quantum Inf. Comput.* **13**, 195–209 (2013).
54. Campbell, E. T., Anwar, H. & Browne, D. E. Magic-state distillation in all prime dimensions using quantum reed-muller codes. *Phys. Rev. X* **2**, 041021 (2012).
55. Jones, C. Multilevel distillation of magic states for quantum computing. *Phys. Rev. A* **87**, 042305 (2013).
56. Rubboli, R. & Tomamichel, M. Fundamental limits on correlated catalytic state transformations. *Phys. Rev. Lett.* **129**, 120506 (2022).
57. Bravyi, S. et al. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum* **3**, 181 (2019).
58. Regula, B. Probabilistic transformations of quantum resources. *Phys. Rev. Lett.* **128**, 110505 (2022).
59. Devetak, I. & Winter, A. Distillation of secret key and entanglement from quantum states. *Proc. R. Soc. A* **461**, 207–235 (2005).
60. Bennett, C. H., Devetak, I., Harrow, A. W., Shor, P. W. & Winter, A. The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels. *IEEE Trans. Inf. Theory* **60**, 2926–2959 (2014).
61. Gour, G. Comparison of quantum channels by superchannels. *IEEE Trans. Inf. Theory* **65**, 5880–5904 (2019).
62. Liu, Z.-W. & Winter, A. Resource theories of quantum channels and the universal role of resource erasure. Preprint at <https://doi.org/10.48550/arXiv.1904.04201> (2019).
63. Liu, Y. & Yuan, X. Operational resource theory of quantum channels. *Phys. Rev. Res.* **2**, 012035 (2020).
64. Fang, K., Wang, X., Tomamichel, M. & Berta, M. Quantum channel simulation and the channel’s smooth max-information. In *2018 IEEE International Symposium on Information Theory (ISIT)*, 2326–2330 (IEEE, 2018).
65. Fang, K., Wang, X., Tomamichel, M. & Berta, M. Quantum channel simulation and the channel’s smooth max-information. *IEEE Trans. Inf. Theory* **66**, 2129–2140 (2019).
66. Gour, G. & Scandolo, C. M. Dynamical entanglement. *Phys. Rev. Lett.* **125**, 180505 (2020).
67. Regula, B. & Takagi, R. Fundamental limitations on distillation of quantum channel resources. *Nat. Commun.* **12**, <https://doi.org/10.1038/s41467-021-24699-0> (2021).
68. Berta, M., Borderi, F., Fawzi, O. & Scholz, V. B. Semidefinite programming hierarchies for constrained bilinear optimization. *Math. Prog.* 1–49 <https://link.springer.com/article/10.1007/s10107-021-01650-1> (2022).
69. Kitaev, A. Y. Quantum computations: algorithms and error correction. *Russ. Math. Surv.* **52**, 1191 (1997).
70. Bluvstein, D. et al. Logical quantum processor based on reconfigurable atom arrays. *Nature* **626**, 58–65 (2023).
71. Wang, Y. et al. Efficient fault-tolerant implementations of non-clifford gates with reconfigurable atom arrays. *Npj Quantum Inf.* **10**, <https://doi.org/10.1038/s41534-024-00945-3> (2024).
72. van Dam, W. & Hayden, P. Universal entanglement transformations without communication. *Phys. Rev. A* **67**, 060302 (2003).
73. Leung, D., Toner, B. & Watrous, J. Coherent state exchange in multiprover quantum interactive proof systems. *Chicago J. Theoret. Comput. Sci.* **11**, 1 (2013).
74. Leung, D. & Wang, B. Characteristics of universal embezzling families. *Phys. Rev. A* **90**, 042331 (2014).
75. Zanon, E., Theurer, T. & Gour, G. Complete characterization of entanglement embezzlement. *Quantum* **8**, 1368 (2024).
76. Brun, T., Devetak, I. & Hsieh, M.-H. Correcting quantum errors with entanglement. *Science* **314**, 436–439 (2006).
77. Brun, T. A., Devetak, I. & Hsieh, M.-H. Catalytic quantum error correction. *IEEE Trans. Inf. Theory* **60**, 3073–3089 (2014).
78. Chitambar, E., Leung, D., Mančinska, L., Ozols, M. & Winter, A. Everything you always wanted to know about locc (but were afraid to ask). *Commun. Math. Phys.* **328**, 303–326 (2014).
79. Uhlmann, A. The “transition probability” in the state space of a^* -algebra. *Rep. Math. Phys.* **9**, 273–279 (1976).

Acknowledgements

We thank Hayata Yamasaki for pointing out the deterministic nature of their magic state distillation protocols described in ref. 14. K.F. is supported in part by the National Natural Science Foundation of China (Grant No. 92470113 and 12404569), the Shenzhen Science and Technology Program (Grant No. QNXMB20250701091826036 and JCYJ20240813113519025), the Shenzhen Fundamental Research Program (Grant No. JCYJ20241202124023031), the General R&D Projects of 1+1+1 CUHK-CUHK(SZ)-GDST Joint Collaboration Fund (Grant No. GRDP2025-022), and the University Development Fund (Grant No. UDF01003565). Z.-W.L. is supported in part by NSFC under Grant No. 12475023, Dushi program, and startup funding from YMSC.

Author contributions

K.F. developed the main idea and technical results. K.F. and Z.-W.L. jointly contributed to the conception, discussion and interpretation of the results and writing of the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-72686-0>.

Correspondence and requests for materials should be addressed to Kun Fang or Zi-Wen Liu.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026