

Title: No-go theorems for quantum resource purification

Speakers: Zi-Wen Liu

Collection: Symmetry, Phases of Matter, and Resources in Quantum Computing

Date: November 28, 2019 - 2:45 PM

URL: <http://pirsa.org/19110146>

Abstract: The manipulation of quantum "resources" such as entanglement and coherence lies at the heart of quantum advantages and technologies. In practice, a particularly important kind of manipulation is to "purify" the quantum resources, since they are inevitably contaminated by noises and thus often lost their power or become unreliable for direct usage. Here we derive fundamental limitations on how effectively generic noisy resources can be purified enforced by the laws of quantum mechanics, which universally apply to any reasonable kind of quantum resource. Remarkably, it is impossible to achieve perfect resource purification, even probabilistically. Our theorems indicate strong limits on the efficiency of distillation, a widely-used type of resource purification routine that underpins many key applications of quantum information science. In particular, we present explicit lower bounds on the resource cost of magic state distillation, a leading scheme for realizing scalable fault-tolerant quantum computation

No-go theorems for quantum resource purification  
And logarithmic lower bounds on distillation overhead

Zi-Wen Liu  
Perimeter Institute

ArXiv:1909.02540  
Joint work with Kun Fang (Cambridge)

# Outline

- No-go theorems for quantum resource purification
- Lower bounds on distillation overhead
- Magic state distillation

# Motivation

- Quantum technologies (computing, communication, sensing...) potentially provide revolutionary advantages over conventional methods, if reliably scaled up
- Scientific experiments demand more and more precise controls over quantum systems

Noise is a notorious obstacle...

- Quantum systems are highly susceptible to noise effects (undesirable interactions with environments, imperfect controls, unstable memories...)
- Generic effect: making the system **mixed**, thus becoming unreliable for usage or lose its advantage
- Ubiquitous need for q. error correction and noise mitigation routines

Noisy

$\rho$



Pure

$\psi$

Noisy

$\rho$



Pure

$\psi$

- “Purification”
- Example—error correction:  
noisy qubits -> (error correction procedures) -> clean logical qubits

# Limitations induce resources

- Practical scenarios always entail physical limitations

# Limitations induce resources

- Practical scenarios always entail physical limitations
- Examples:
  - For “distant labs”, LOCC is cheap; **Entanglement** emerge as “resource” since it can enable q. communication & cannot be created by LOCC
  - For fault-tolerance, stabilizer circuits are cheap; **Magic states** enable universal q. computation...



# Limitations induce resources

- Practical scenarios always entail physical limitations
- Examples:
  - For “distant labs”, LOCC is cheap; **Entanglement** emerge as “resource” since it can enable q. communication & cannot be created by LOCC
  - For fault-tolerance, stabilizer circuits are cheap; **Magic states** enable universal q. computation...
- Many other types: coherence, asymmetry, athermality, nonlocality, non-Gaussianity...  
Origins of quantum advantages/phenomena in different scenarios

# Limitations induce resources

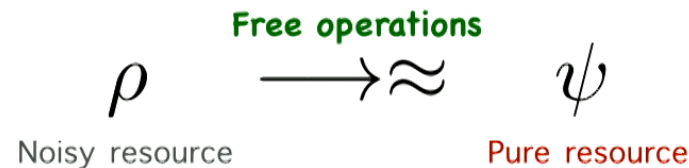
- Practical scenarios always entail physical limitations
- Examples:
  - For “distant labs”, LOCC is cheap; **Entanglement** emerge as “resource” since it can enable q. communication & cannot be created by LOCC
  - For fault-tolerance, stabilizer circuits are cheap; **Magic states** enable universal q. computation...
- Many other types: coherence, asymmetry, athermality, nonlocality, non-Gaussianity... Origins of quantum advantages/phenomena in different scenarios
- “Resource theory” (see also talks by Spekkens, Campbell, Gross..) abstracts out the structure: manipulation of resources under “free” operations

# Limitations induce resources

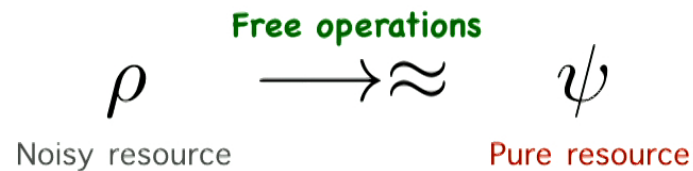
- Practical scenarios always entail physical limitations
- Examples:
  - For “distant labs”, LOCC is cheap; **Entanglement** emerge as “resource” since it can enable q. communication & cannot be created by LOCC
  - For fault-tolerance, stabilizer circuits are cheap; **Magic states** enable universal q. computation...
- Many other types: coherence, asymmetry, athermality, nonlocality, non-Gaussianity...  
Origins of quantum advantages/phenomena in different scenarios
- “Resource theory” (see also talks by Spekkens, Campbell, Gross..) abstracts out the structure: manipulation of resources under “free” operations
- **Golden rule** of free operation: Maps free state to free state
  - Resource “non-generating” operations
  - Includes all possible sets of free operations (minimal condition of nontrivial theory)  
⇒ What is impossible for this set is impossible for all theories

# Limitations induce resources

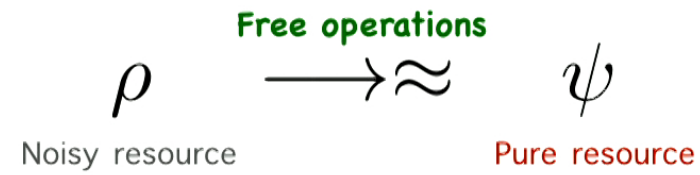
- Practical scenarios always entail physical limitations
- Examples:
  - For “distant labs”, LOCC is cheap; **Entanglement** emerge as “resource” since it can enable q. communication & cannot be created by LOCC
  - For fault-tolerance, stabilizer circuits are cheap; **Magic states** enable universal q. computation...
- Many other types: coherence, asymmetry, athermality, nonlocality, non-Gaussianity... Origins of quantum advantages/phenomena in different scenarios
- “Resource theory” (see also talks by Spekkens, Campbell, Gross..) abstracts out the structure: manipulation of resources under “free” operations
- **Golden rule** of free operation: Maps free state to free state
  - Resource “non-generating” operations
  - Includes all possible sets of free operations (minimal condition of nontrivial theory)  
⇒ What is impossible for this set is impossible for all theories



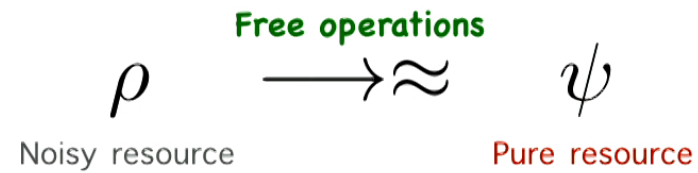
# One-shot



- In practice, we only have access to a finite amount of resources — “One/ single-shot” setting
- Some error could be allowed
- In contrast to the “asymptotic” assumption conventional in information theory, where we consider the limit of infinite i.i.d. instances for simplification
- E.g. Input and target can be finite copies of some “unit” state, codes of finite size...
- A prequel to this work: Unified theory of resource monotones that quantify the values of resources in one-shot settings [ZWL/Bu/Takagi, PRL '19]



One mathematical assumption:  $\rho$  is full-rank (e.g.  $\sim$ depolarizing)



One mathematical assumption:  $\rho$  is full-rank (e.g.  $\sim$ depolarizing)

Results apply to:

- Any kind of resource (entanglement, coherence, [magic states](#)...)
- Any pure non-free state as target (however weak it is)
- Any free operation (minimal constraint on legitimate procedure)

# Deterministic transformation

For intuition, first consider the case of q. channels/CPTP maps, representing deterministic protocols that do not use postselection

Theorem (No-go for deterministic purification)

Given **any** full-rank state  $\rho$  and **any** pure target resource state  $\psi$ , the transformation error (1-fidelity)  $\epsilon$  of **any** free channel must satisfy

$$\epsilon \geq \varepsilon(\rho, \psi) = \lambda_{\min}(\rho) (1 - f_{\psi})$$

Smallest eigenvalue

Max overlap with  
free states, <1



# Proof idea

- An information-theoretic tool: Quantum hypothesis testing relative entropy (QHTRE) [Buscemi/Datta TIT '10, Wang/Renner PRL '12]

## Definition (QHTRE)

$$D_H^\epsilon(\rho\|\sigma) := \max_{0 \leq P \leq I, \text{Tr}\{P\rho\} \geq 1-\epsilon} (-\log \text{Tr}\{P\sigma\})$$

- A divergence based on the error of distinguishing the two states (HT)
- “Operator-smoothing” of the “min-relative entropy”:

$$D_H^{\epsilon=0}(\rho\|\sigma) = D_{\min}(\rho\|\sigma) = -\log \text{Tr}\{\Pi_\rho\sigma\}$$

0-order limit of the (non-sandwiched) q. Renyi divergence (smallest!)

- Minimizing over free  $\sigma$  gives a **monotone**, due to the data processing inequality of QHTRE [Wang/Renner PRL '12]
- Modified versions known to upper bound one-shot distillation rate [ZWL/Bu/Takagi PRL '19]

## Proof idea

- Consider  $D_H^{\epsilon=0}(\rho\|\sigma) = D_{\min}(\rho\|\sigma) = -\log \text{Tr} \{ \Pi_{\rho} \sigma \}$

Observation: When  $\rho$  is full-rank, it vanishes

- $\epsilon$  characterizes transformation error. To account for approximate transformation, we need the following continuity bound

### Lemma

Given **any** full-rank state  $\rho$  and **any**  $\sigma$ , their QHTRE is continuous around  $\epsilon=0$ .  
For  $0 \leq \epsilon < \lambda_{\min}(\rho)$

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

- So: the QHTRE monotone can be made arbitrarily small by decreasing  $\epsilon$  for all full-rank states, but for pure states  $\geq$  constant...  
By suitably combining the above continuity bound and monotonicity, we prove the claimed bound

## An alternative proof

See more discussions on this “overlap formalism” of distillation in [Regula/  
Bu/Takagi/ZWL 1909.11677]

# An alternative proof

See more discussions on this “overlap formalism” of distillation in [Regula/Bu/Takagi/ZWL 1909.11677]

Theorem (Maximum fidelity [1909.11677])

$$F_O(\rho \rightarrow \psi) := \sup_{\Lambda \in \mathcal{O}} \text{Tr}[\Lambda(\rho)\psi] \leq G\left(\rho; f_\psi^{-1}\right)$$

$$G(\rho; k) := \sup \left\{ \text{Tr}[\rho W] \mid 0 \leq W \leq \mathbb{I}, \text{Tr}[\sigma W] \leq \frac{1}{k}, \forall \sigma \in \mathcal{F} \right\}$$

# An alternative proof

See more discussions on this “overlap formalism” of distillation in [Regula/Bu/Takagi/ZWL 1909.11677]

Theorem (Maximum fidelity [1909.11677])

$$F_O(\rho \rightarrow \psi) := \sup_{\Lambda \in \mathcal{O}} \text{Tr}[\Lambda(\rho)\psi] \leq G\left(\rho; f_\psi^{-1}\right)$$

$$G(\rho; k) := \sup \left\{ \text{Tr}[\rho W] \mid 0 \leq W \leq \mathbb{I}, \text{Tr}[\sigma W] \leq \frac{1}{k}, \forall \sigma \in \mathcal{F} \right\}$$

- Dual optimization problem: (strong duality holds for finite dimensions)

$$\tilde{G}(\rho; k) := \frac{1}{k} + \frac{1}{k} \inf \{ \text{Tr}[A] + (k-1) \text{Tr}[B] \mid A, B \geq 0, \rho + A - B = X, X \in \text{cone}(\mathcal{F}) \}$$

- For any  $X$ , the optimum is achieved by  $A = (\rho - X)_-, B = (\rho - X)_+$ , so

$$\tilde{G}(\rho; k) := \frac{1}{k} + \frac{1}{k} \inf \{ \text{Tr}(\rho - X)_- + (k-1) \text{Tr}(\rho - X)_+ \mid X \in \text{cone}(\mathcal{F}) \}$$

# Probabilistic transformation

- E.g. error correction routines, succeed probabilistically upon passing syndrome measurements
- Setting: A free probabilistic protocol with succ. prob.  $p$  and error  $\epsilon$  is modeled as the following “flagged” transformation

$$\mathcal{E}_{A \rightarrow XB}(\rho_A) = |0\rangle\langle 0|_X \otimes \mathcal{L}_{A \rightarrow B}(\rho_A) + |1\rangle\langle 1|_X \otimes \mathcal{G}_{A \rightarrow B}(\rho_A)$$

- Flag register  $X$  records whether the protocol succeeds or not
- $\mathcal{L}, \mathcal{G}$  are free **sub**-operations (CP trace non-increasing maps) under golden rule  $\forall \rho \in \mathcal{F}, \exists t \geq 0, \sigma \in \mathcal{F}, \text{ s.t. } \mathcal{L}(\rho) = t \cdot \sigma$

# Probabilistic transformation

- E.g. error correction routines, succeed probabilistically upon passing syndrome measurements
- Setting: A free probabilistic protocol with succ. prob.  $p$  and error  $\epsilon$  is modeled as the following “flagged” transformation

$$\mathcal{E}_{A \rightarrow XB}(\rho_A) = |0\rangle\langle 0|_X \otimes \mathcal{L}_{A \rightarrow B}(\rho_A) + |1\rangle\langle 1|_X \otimes \mathcal{G}_{A \rightarrow B}(\rho_A)$$

Desired sub-transformation

Garbage

$$\mathcal{L}_{A \rightarrow B}(\rho_A) = p\tau_B$$

$$p = \text{Tr } \mathcal{L}(\rho)$$

$$F(\tau, \psi) \geq 1 - \epsilon$$

- Flag register  $X$  records whether the protocol succeeds or not
- $\mathcal{L}, \mathcal{G}$  are free **sub-operations** (CP trace non-increasing maps) under golden rule  $\forall \rho \in \mathcal{F}, \exists t \geq 0, \sigma \in \mathcal{F}, \text{ s.t. } \mathcal{L}(\rho) = t \cdot \sigma$

# Probabilistic transformation

Theorem (No-go for probabilistic purification)

Given **any** full-rank state  $\rho$  and **any** pure target resource state  $\psi$ , the succ. prob.  $p$  and error  $\epsilon$  of **any** free probabilistic transformation must satisfy

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

Generalized robustness

$$R_G(\rho) := \min\{s \mid \exists \sigma, s \geq 0, \text{ s.t. } (\rho + s\sigma)/(1 + s) \in \mathcal{F}\}$$



# Probabilistic transformation

## Theorem (No-go for probabilistic purification)

Given **any** full-rank state  $\rho$  and **any** pure target resource state  $\psi$ , the succ. prob.  $p$  and error  $\epsilon$  of **any** free probabilistic transformation must satisfy

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

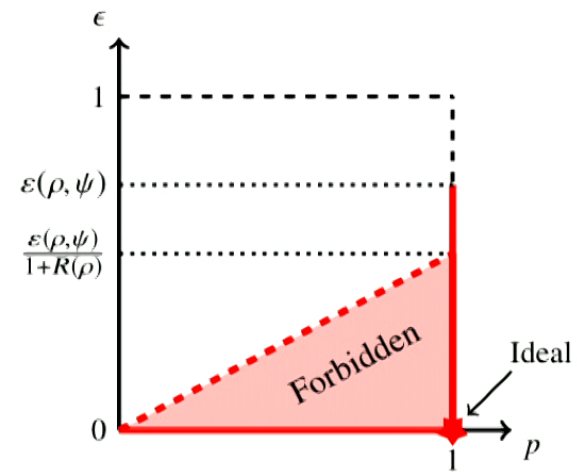
- Proof needs another property of QHTRE:

For flagged states  $\rho = p|0\rangle\langle 0| \otimes \rho_1 + (1 - p)|1\rangle\langle 1| \otimes \rho_2$  it holds that  
 $\sigma = q|0\rangle\langle 0| \otimes \sigma_1 + (1 - q)|1\rangle\langle 1| \otimes \sigma_2$

$$\beta_\epsilon(\rho \parallel \sigma) \leq q\beta_\epsilon(\rho_1 \parallel \sigma_1) + (1 - q)\beta_\epsilon(\rho_2 \parallel \sigma_2), \quad \beta_\epsilon = 2^{-D_H^\epsilon}$$

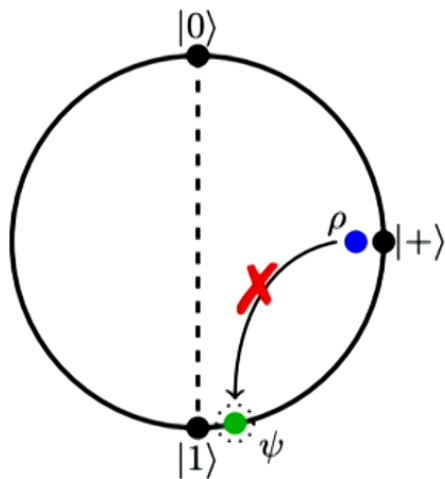
- A technical fact  $\exists \omega \in \mathcal{F}, \text{Tr } \mathcal{L}(\omega) \geq (1 + R_G(\rho))^{-1} \text{Tr } \mathcal{L}(\rho)$
- The rest follows similar idea as the deterministic case

There is a fundamental trade-off between accuracy and success probability—an “uncertainty principle”



#### Corollary

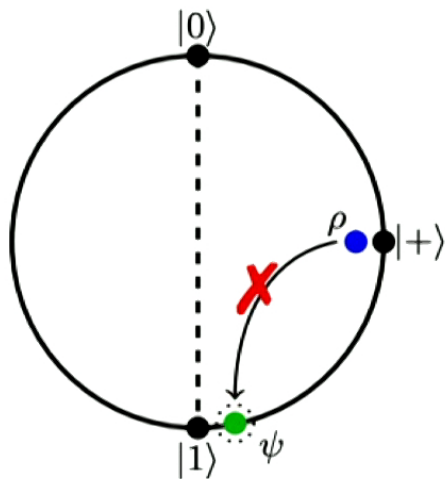
Perfect resource purification is impossible, even probabilistically.



$\rho$  : a slightly dirty powerful resource state  
(could even live in higher dim)

$\psi$  : a very weak pure resource state

- $\rho$  could be much more useful in generic tasks, take much larger value in terms of other resource monotones...
- But,  $\rho$  cannot be used to trade for a nearly perfect  $\psi$  (even probabilistically)



$\rho$  : a slightly dirty powerful resource state  
(could even live in higher dim)

$\psi$  : a very weak pure resource state

- $\rho$  could be much more useful in generic tasks, take much larger value in terms of other resource monotones...
- But,  $\rho$  cannot be used to trade for a nearly perfect  $\psi$  (even probabilistically)

Highlights the special role of QHTRE monotone — Other monotones do not rule out the  $\rho > \psi$  partial order



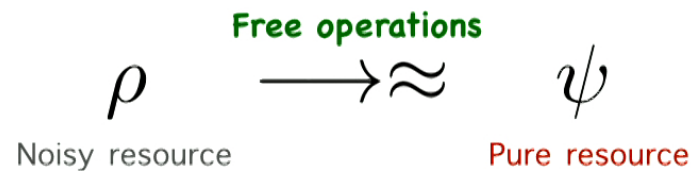
**No**  
“~~We must find~~ one magic monotone to rule  
them all” - ~~Not Spekkens~~

*Spekkens, Not Campbell*

# Resource overhead of distillation

Practically, we often use a supply of raw resources to “distill” some high-quality target (entanglement, magic state, coherence...)

Aim to minimize the cost/overhead of the task, i.e. the “amount” of raw resources needed.



# Resource overhead of distillation

Practically, we often use a supply of raw resources to “distill” some high-quality target (entanglement, magic state, coherence...)

Aim to minimize the cost/overhead of the task, i.e. the “amount” of raw resources needed.

$$\hat{\rho}^{\otimes n} \xrightarrow{\text{Free operations}} \approx \psi$$

Noisy resource Pure resource

## Definition (Total overhead)

Given  $n$  copies of unit noisy primitive states  $\rho$ , output a state that  $\varepsilon$ -approximates some target pure state  $\psi$ .

Total overhead/cost :=  $\min n$ .

# Resource overhead of distillation

Theorem (Lower bounds on total overhead)

For **any** full-rank  $\rho$  and sufficiently small  $\epsilon$ :

$$n \geq \log_{\frac{1+R_G(\phi)}{\lambda_{\min}(\hat{\rho})}} \frac{(1 - f_\psi) p}{\epsilon}$$

For deterministic case ( $p=1$ ) the bound can be improved to

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

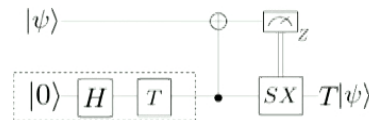
Total overhead scales as  $\Omega(\log(1/\epsilon))$



# Magic state distillation

A leading scheme of fault-tolerant quantum computing

- Clifford gates are low-cost, but not universal (Gottesman-Knill)...
- Need to add “magic”/non-Clifford gates, e.g. T-gates, to make it universal, but they are hard to protect
- A compelling method: use the “injection” gadget to consume T-states to emulate T-gates



⇒ Produce high-quality T-states in a separate “factory”!

The huge cost for achieving fault-tolerant universality is a major obstacle for practical q. computing... MSD is estimated to be a dominant source of resource cost...

Understanding the achievable efficiency of MSD is of great theoretical and practical importance

# Magic state distillation

Common MSD protocols: encode in some  $[n,k,d]$  stabilizer code  $\rightarrow$  syndrome measurement  $\rightarrow$  postselection  $\rightarrow$  decode (omitting Clifford processings), to reduce noise. Iterate until reaching target error  $\epsilon$ .

# Magic state distillation

Common MSD protocols: encode in some  $[n,k,d]$  stabilizer code  $\rightarrow$  syndrome measurement  $\rightarrow$  postselection  $\rightarrow$  decode (omitting Clifford processings), to reduce noise. Iterate until reaching target error  $\epsilon$ .

General formulation of the task: Given  $n$  full rank noisy magic states  $\rho$ , output with probability  $p$  an  $m$ -qubit state  $\tau$  s.t.

$$\text{Tr } \tau_i T = \langle T | \tau_i | T \rangle \geq 1 - \epsilon, \forall i = 1, \dots, m$$

# Magic state distillation

Common MSD protocols: encode in some  $[n,k,d]$  stabilizer code  $\rightarrow$  syndrome measurement  $\rightarrow$  postselection  $\rightarrow$  decode (omitting Clifford processings), to reduce noise. Iterate until reaching target error  $\epsilon$ .

General formulation of the task: Given  $n$  full rank noisy magic states  $\rho$ , output with probability  $p$  an  $m$ -qubit state  $\tau$  s.t.

$$\text{Tr } \tau_i T = \langle T | \tau_i | T \rangle \geq 1 - \epsilon, \forall i = 1, \dots, m$$

A widely considered figure of merit is the average overhead  $n/m$ , especially the exponent  $\gamma$  in its scaling  $O(\log^\gamma(1/\epsilon))$ . Assuming non-vanishing succ. prob.,

$$\gamma = \log(n/k)/\log d.$$

# Magic state distillation

Theorem (Lower bound on T-state distillation average overhead)

$$n/m \geq \frac{1}{m} \log_{\frac{1+R_G(\sigma)}{\lambda_{\min}(\sigma)}} \frac{((4 - 2\sqrt{2})^m - 1) p}{(4 - 2\sqrt{2})^m m \epsilon}$$

(For other target states of interest, just plug in corresponding parameters)

# Magic state distillation

Theorem (Lower bound on T-state distillation average overhead)

$$n/m \geq \frac{1}{m} \log_{\frac{1+R_G(\sigma)}{\lambda_{\min}(\sigma)}} \frac{((4 - 2\sqrt{2})^m - 1) p}{(4 - 2\sqrt{2})^m m \epsilon}$$

(For other target states of interest, just plug in corresponding parameters)

⇒ Any “sublogarithmic” protocol [Hastings/Haah PRL '18, Krishna/Tillich PRL '19] must have rapidly diverging output size — although avg. overhead is considered low, the overall cost could easily blow up

⇒ For any constant-size-output protocol (with non-vanishing succ. prob. under concatenation), e.g. concatenation of  $k=1$  codes, the avg. overhead is  $\Omega(\log(1/\epsilon))$  ( $\gamma \geq 1$  lower bound; best known  $\gamma \rightarrow 2$ , still a gap)

# Channel simulation

Channels (gates, dynamical processes...) are also important resource objects

General theory: ZWL/Winter 1904.04201, Spekkens...

# Channel simulation

Channels (gates, dynamical processes...) are also important resource objects

General theory: ZWL/Winter 1904.04201, Spekkens...

Theorem (No-go for unitary channel simulation)

Given any noisy channel  $\mathcal{N}_{A \rightarrow B}$  and any target unitary resource channel  $\mathcal{U}_{C \rightarrow D}$   
No free superchannel can transform  $\mathcal{N}_{A \rightarrow B}$  to  $\mathcal{U}_{C \rightarrow D}$  with zero error.



# Channel simulation

Channels (gates, dynamical processes...) are also important resource objects

General theory: ZWL/Winter 1904.04201, Spekkens...

Theorem (No-go for unitary channel simulation)

Given any noisy channel  $\mathcal{N}_{A \rightarrow B}$  and any target unitary resource channel  $\mathcal{U}_{C \rightarrow D}$   
No free superchannel can transform  $\mathcal{N}_{A \rightarrow B}$  to  $\mathcal{U}_{C \rightarrow D}$  with zero error.

Similar golden rule

# Channel simulation

Channels (gates, dynamical processes...) are also important resource objects

General theory: ZWL/Winter 1904.04201, Spekkens...

Theorem (No-go for unitary channel simulation)

Given any noisy channel  $\mathcal{N}_{A \rightarrow B}$  and any target unitary resource channel  $\mathcal{U}_{C \rightarrow D}$   
No free superchannel can transform  $\mathcal{N}_{A \rightarrow B}$  to  $\mathcal{U}_{C \rightarrow D}$  with zero error.

Similar golden rule

Based on a channel analogue of QHTRE

Corollary: Zero-error quantum capacity of generic noisy channels (e.g. depolarizing) is zero

## Summary & outlook

- Nearly perfect purification of **any** resource by **any** free procedure is never possible for generic noisy systems; draws practical boundary for general error correction
- $\log(1/\varepsilon)$  lower bound on the total overhead of distillation-type tasks

# Summary & outlook

- Nearly perfect purification of **any** resource by **any** free procedure is never possible for generic noisy systems; draws practical boundary for general error correction
- $\log(1/\varepsilon)$  lower bound on the total overhead of distillation-type tasks
- Achievability of fundamental limits (good answers in specific theories)
- Approximate and probabilistic cases for channel theory
- Continuous variable (Work in progress). Can transplant no-go theorem when the system is “gapped”. When “gapless”: no-go techniques do not carry over, and no distillation method that keeps improving the fidelity w/ e.g. GKP is known—no lower & upper bound known!

Thanks for your attention!

arXiv:1909.02540