Error Mitigation and Quantum-Assisted Simulation in the Error Corrected Regime

M. Lostaglio [1,2] and A. Ciani [1,3]

[1] *QuTech, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands*
[2] *Korteweg-de Vries Institute for Mathematics and QuSoft, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands*
[3] *JARA Institute for Quantum Information, Forschungszentrum Jülich, D-52425 Jülich, Germany*

(Received 7 April 2021; accepted 28 September 2021; published 12 November 2021)

A standard approach to quantum computing is based on the idea of promoting a classically simulable and fault-tolerant set of operations to a universal set by the addition of "magic" quantum states. In this context, we develop a general framework to discuss the value of the available, nonideal magic resources, relative to those ideally required. We single out a quantity, the quantum-assisted robustness of magic (QROM), which measures the overhead of simulating the ideal resource with the nonideal ones through quasiprobability-based methods. This extends error mitigation techniques, originally developed for noisy intermediate-scale quantum devices, to the case where qubits are logically encoded. The QROM shows how the addition of noisy magic resources allows one to boost classical quasiprobability simulations of a quantum circuit and enables the construction of explicit protocols, interpolating between classical simulation and an ideal quantum computer.

DOI: 10.1103/PhysRevLett.127.200506

Introduction.—Large-scale quantum computing would allow us to solve computational problems that are intractable for classical computers. Because of the fragility of quantum information encoded in physical systems, quantum error correction will be a central component of these machines. However, several restrictions exist to the possibility of achieving a "universal" set of fault-tolerant quantum gates [1,2]. In a standard setting, the "stabilizer operations" (which involve computational basis preparation and measurement, Clifford unitaries, partial trace, and classical randomness [3,4]) are fault tolerant and hence "free." Curiously, the Gottesman-Knill theorem tells us that these operations can be simulated efficiently on a classical computer [3,5]. Stabilizer operations can be promoted to universality by injecting magic states, which are thus a "resource" for quantum computation. Magic states may come at a limited rate, since their fault-tolerant preparation involves complex distillation schemes [6–11]. Also, much of the residual noise in an error corrected computation can originate from these elements.

In this Letter, we go beyond the dichotomy free and resourceful and think in terms of the resource content of an ideal quantum resource *relative* to an available one. We formulate questions such as "How valuable is an ideal magic state, compared to its noisy version, as a function of the noise level?" or "How valuable are the r available magic states, compared to the $t > r$ ideally needed?". Our notion of relative value stems from the overhead of simulating ideal resources with nonideal ones through quasiprobability-based error mitigation [12–16]. As such, it is endowed with a clear operational significance.

Quasiprobability-based error mitigation is a technique that allows one to remove bias from the outcome probabilities of a measurement by expressing the ideal circuit elements as linear combinations of nonideal ones. The coefficients of the decomposition define a quasiprobability, since they are real and sum to one. By sampling from the absolute value of this quasiprobability and performing the corresponding nonideal operations, one can remove the bias at the price of a sampling overhead.

We extend these ideas from noisy intermediate-scale quantum devices [12–16] to fault-tolerant quantum computing. In this context, we single out a measure, the quantum-assisted robustness of magic (QROM), which, loosely speaking, is a distance of the ideal elements relative to the available ones. The QROM provides a unified setting to analyze several central computational and simulation tasks, investigated separately in the literature: (1) Classical simulation: classically sample from the outcome probabilities of an ideal quantum circuit with t magic T states $|T\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$ as input (t is known as the T count). Having no quantum resources at hand, the QROM coincides with the robustness of magic (ROM), a known measure of the overhead of quasiprobability-based classical simulation [4,17–19]. (2) Error mitigation in the quantum error corrected regime: given a quantum circuit involving t ideal T states, obtain the average of measurement outcomes using a circuit with t noisy T states. The overhead is measured by the QROM [20]. (3) Quantum-assisted simulation: obtain the average of measurement outcomes of an ideal quantum circuit involving t ideal T states, given that we can inject only $r < t$ noisy T states. This task can be seen as an intermediate scenario between classical simulation ($r = 0$) and error mitigation ($r = t$), where some quantum resources are available but fewer than what is needed. The residual overhead is quantified by the QROM.

By analytically or numerically solving optimization problems for the QROM, we construct quasiprobability-based error mitigation and quantum-assisted simulation algorithms in the error corrected regime and analyze how the addition of quantum resources gradually interpolates between a classical simulation and an ideal quantum computation.

Quantum-assisted robustness of magic.—The QROM is a generalization of the ROM and formalizes the idea of robustness of an ideal quantum computational resource relative to the (fewer, noisier) available ones. Consider a quantum computer where magic states are injected by means of t ancilla qubits. A resource state σ_r on $r \leq t$ qubits is given, e.g., T states resulting from a number of distillation rounds. Because of practical limitations, these states are, in general, fewer and noisier than required by the ideal circuit.

In broad generality, one can denote by $\mathcal{Q}_t(\sigma_r)$ all the t-qubit states achievable from the resource σ_r by means of stabilizer operations. The relative robustness of a t-qubit ideal resource ρ_t with respect to the available resource σ_r is

$$R(\rho_t|\sigma_r) = \min\left\{ s \geq 0 \middle| \frac{\rho_t + s\eta}{1+s} \in \mathcal{Q}_t(\sigma_r), \eta \in \mathcal{Q}_t(\sigma_r) \right\}. \quad (1)$$

The relative robustness represents the minimum amount of mixing between ρ_t and a state in $\mathcal{Q}_t(\sigma_r)$ such that the resulting state is also in $\mathcal{Q}_t(\sigma_r)$. Since all pure stabilizer states on t qubits can be generated by applying t-qubit Clifford unitaries to $|0\rangle^{\otimes t}$ [3], $R(\rho_t|\sigma_r)$ is the robustness relative to a set given by stabilizer states *augmented* by the available resources, as illustrated in Fig. 1.

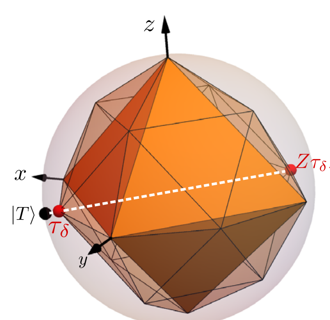

FIG. 1. Relative robustness for a single qubit in the Bloch sphere. The black dot represents the pure magic state $|T\rangle$. The orange, full octahedron represents the convex hull of stabilizer states, while the orange shaded region is the set $\mathcal{Q}_1(\tau_\delta)$ representing all states achievable from the available resource [a noisy T state $\tau_\delta = (1-\delta)|T\rangle\langle T| + \delta I/2$ with $\delta = 0.1$] by stabilizer operations. Geometrically, the relative robustness is the minimum mixing needed to bring the T state inside $\mathcal{Q}_1(\tau_\delta)$ (in this case, mixing with $Z\tau_\delta Z$).

The QROM is the negativity required to decompose the ideal resource ρ_t in terms of the available ones in $\mathcal{Q}_t(\sigma_r)$,

$$\mathcal{R}(\rho_t|\sigma_r) = \min\left\{ \sum_x |q_x| \middle| \rho_t = \sum_x q_x \eta_x, \eta_x \in \mathcal{Q}_t(\sigma_r) \right\}. \quad (2)$$

Note that $q_x \in \mathbb{R}$ and $\sum_x q_x = 1$, so q_x defines a quasiprobability. The two quantities just defined are closely related, as they satisfy $\mathcal{R}(\rho_t|\sigma_r) = 1 + 2R(\rho_t|\sigma_r)$ (see Supplemental Material Sec. I [22]). When σ_r can be written as a mixture of stabilizer states, the QROM reduces to the well-known ROM of Ref. [17] [denoted by $\mathcal{R}(\rho_t)$], which measures the overhead to classically simulate the resource ρ_t via the Gottesman-Knill theorem. We will show that $\mathcal{R}(\rho_t|\sigma_r)$ has an analog interpretation in the presence of the quantum resource σ_r. The QROM has desirable properties of a resource theoretical measure [25], similar to those of the Wigner negativity [26–29]: faithfulness, monotonicity, convexity (see Supplemental Material Sec. II [22]).

In general, $\mathcal{R}(\rho_t|\sigma_r) \leq \mathcal{R}(\rho_t)$. As we shall see, the QROM makes quantitative the intuition that the available quantum resources decrease the negativity and hence the computational overhead. The QROM is an operationally motivated concept suitable to study classical simulation, quantum-assisted simulation, and error mitigation in the error corrected regime under a unified umbrella.

Error mitigation in the quantum error corrected regime.—The first task where the QROM plays a role is error mitigation. We want to perform a quantum computation on n data qubits, which can be assumed to be initialized in state $|0\rangle^{\otimes n}$. The circuit involves Clifford unitaries, typically taken from a fundamental gate set, and T gates,

$$T = |0\rangle\langle 0| + e^{i\pi/4}|1\rangle\langle 1|. \quad (3)$$

The latter promote the computation to universality [30]. Since we assume our qubits to be logically encoded in a suitable quantum error correcting code, we take all Clifford operations to be perfect (this assumption will be relaxed later). Each one of the t T gates is realized via a gate teleportation gadget involving a T state $|T\rangle$ [25]. In practice, T states will be noisy. While our approach is general, for simplicity we focus on noisy magic states with standard form

$$\tau_\delta = (1-\delta)\tau + \delta I/2, \quad (4)$$

where $\tau = |T\rangle\langle T|$, I is the identity matrix, and $\delta \in [0, 1]$ is the noise level [31]. The quantum computation terminates with the measurement of a Pauli operator P. We want to estimate its average $\langle P\rangle$ in the final state of the ideal circuit.

Since T states are noisy, the measured average cannot be expected to be an unbiased estimator of $\langle P\rangle$. In order to cancel the bias via error mitigation, first we find a decomposition of the form in Eq. (2). Ideally we wish to find an optimal decomposition according to Eq. (2) (with $\rho_t = \tau^{\otimes t}$ and $\sigma_t = \tau_\delta^{\otimes t}$), but in practice the decomposition need not be optimal. Setting $\|q\|_1 = \sum_x |q_x|$, the algorithm works as follows: (1) Sample x with probability $|q_x|/\|q\|_1$. (2) Run the quantum circuit with input $\eta_x \otimes |0\rangle\langle 0|^{\otimes n}$ [32]. (3) Measure the Pauli operator P, getting outcome $p = -1$ or 1. (4) Output $o = p\|q\|_1 \mathrm{sign}(q_x)$.

Sampling M times and taking the arithmetic average of the outputs, one obtains an unbiased estimator for $\langle P\rangle$. By Hoeffding's inequality [33], after $M \geq 2\ln(2/\epsilon)\|q\|_1^2/\Delta^2$ runs, the estimate is within additive error Δ of $\langle P\rangle$ with probability $1-\epsilon$ (see Supplemental Material Sec. IV [22]). Hence, this protocol cancels the error on $\langle P\rangle$ from the noisy T states at the price of a sampling overhead. By using an optimal decomposition,

$$M = 2\ln(2/\epsilon)\mathcal{R}(\tau^{\otimes t}|\tau_\delta^{\otimes t})^2/\Delta^2 \tag{5}$$

suffices. Since for an ideal quantum computer one needs $M_{\mathrm{ideal}} = 2\ln(2/\epsilon)/\Delta^2$ runs to obtain the same guarantee, the QROM squared quantifies the overhead.

Finding the optimal decomposition in Eq. (2) is a convex optimization problem [34] whose size scales superexponentially with the T count [35]. While such an approach is unscalable, we obtain upper and lower bounds

$$\frac{\mathcal{M}(\tau^{\otimes t})}{\mathcal{M}(\tau_\delta^{\otimes t})} \leq \mathcal{R}(\tau^{\otimes t}|\tau_\delta^{\otimes t}) \leq [\mathcal{R}(\tau^{\otimes k}|\tau_\delta^{\otimes k})]^{t/k}. \tag{6}$$

Lower bounds can be derived from any convex magic monotone $\mathcal{M}$, see Supplemental Material Sec. V [22]. Taking $\mathcal{M}$ to be the dyadic negativity [36], we show that $\mathcal{R}(\tau^{\otimes t}|\tau_\delta^{\otimes t})$ grows exponentially with the T count t for any $\delta \in (0,1]$,

$$\mathcal{R}(\tau^{\otimes t}|\tau_\delta^{\otimes t}) \geq \min\left\{\frac{1}{(1-\delta/2)^t}, [2(2-\sqrt{2})]^t\right\}.$$

The upper bounds are based on the submultiplicativity of the QROM. Taking t to be divisible by k, $\tau^{\otimes t} = \tau^{\otimes k} \otimes \ldots \otimes \tau^{\otimes k}$, decompositions for $\tau^{\otimes k}$ give (suboptimal) block decompositions for $\tau^{\otimes t}$. These can be used in practical error mitigation protocols.

Let us start with $k = 1$. The QROM is

$$\mathcal{R}(\tau|\tau_\delta) = \begin{cases} \sqrt{2} = \mathcal{R}(\tau), & \delta > \delta_{\mathrm{th}}, \\ \frac{1}{1-\delta}, & \delta \leq \delta_{\mathrm{th}}. \end{cases} \tag{7}$$

Above a noise threshold $\delta_{\mathrm{th}} = 1 - \sqrt{2}/2 \approx 0.293$ the ROM is recovered [17]. The optimal decomposition for $\delta \leq \delta_{\mathrm{th}}$ is instead

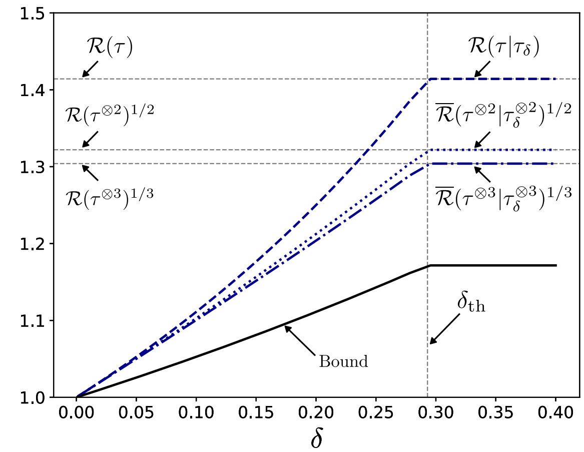

FIG. 2. QROM as a function of noise. kth root of the QROM of $\tau^{\otimes k}$ with respect to $\tau_\delta^{\otimes k}$, as a function of the noise level δ. $\bar{\mathcal{R}}(\tau^{\otimes k}|\tau_\delta^{\otimes k})$ are upper bounds for the corresponding QROM, based on analytical decompositions. At $\delta = \delta_{\mathrm{th}}$, the $k = 2$ ($k = 3$) upper bound recovers exactly (approximately) the ROM $\mathcal{R}(\tau^{\otimes k})$ of Ref. [17]. The black solid line represents the bound $\mathcal{R}_{\mathrm{bound}}(\tau^{\otimes t}|\tau_\delta^{\otimes t})^{1/t}$, which for small δ goes as $1 + \delta/2$. While $k = 2, 3$ outperform $k = 1$, the improvement is modest for small level of noise.

$$\tau = \frac{1-\delta/2}{1-\delta}\tau_\delta - \frac{\delta/2}{1-\delta}Z\tau_\delta Z, \tag{8}$$

where Z is the Pauli z matrix. The threshold corresponds to the point where the noisy magic state τ_δ enters the set of stabilizer states (inner octahedron in Fig. 1). By tuning the noise, the QROM interpolates between error mitigation and classical simulation. The latter is recovered at $\delta = \delta_{\mathrm{th}}$, since at that point the noise is so large that the quantum resources can be classically simulated efficiently.

For $k = 2, 3$ the situation is similar, but computing the QROM exactly is hard. In Supplemental Material Sec. VI [22] we obtain explicit decompositions whose performance is presented in Fig. 2.

In the small error regime, realistic for quantum error corrected setups ($\delta \lesssim 10^{-2}$), global decompositions only marginally outperform the single-qubit decomposition of Eq. (8) [37]. Already at $\delta = 10^{-2}$ and moderate overheads ($\sim 10^2$) one can error mitigate in regimes ($t \sim 230$) in which classical simulation is currently unfeasible even with state-of-the-art algorithms [36,38,39].

Error mitigation with noisy Cliffords.—So far, we assumed that Clifford unitaries are ideal. In reality, they have a residual noise. In this situation it is more natural to perform error mitigation at the level of channels rather than states. Consider a simple error model where each $k = 1$, 2-qubit Clifford $\mathcal{U}^{(k)}$ is independently affected by depolarizing noise,

$$\mathcal{U}_{\delta_c}^{(k)} = (1-\delta_c)\mathcal{U}^{(k)} + \delta_c\mathcal{G}_k, \tag{9}$$

where $\mathcal{G}_k(\rho) = I/2^k$ for all ρ and $\delta_c \in [0, 1]$ is the noise level on the Cliffords. The noise affects not only the Clifford unitaries on the data qubits, but also those involved in the gate teleportation gadget. The implementable T gates, denoted by $\mathcal{T}_{\delta,\delta_c}$, are then noisy both due to $\delta > 0$ and $\delta_c > 0$. Consider an ideal circuit realizing a unitary $\mathcal{U}$ by a sequence of $n_c^{(1)}$ single-qubit Cliffords, $n_c^{(2)}$ 2-qubit Cliffords and t T gates. To reproduce the expectation values $\langle P \rangle$ of this ideal circuit, find a quasiprobability decomposition $\mathcal{U} = \sum_x q_x \mathcal{A}_x$, with $\mathcal{A}_x$ available noisy channels (in this case, $\mathcal{A}_x = \mathcal{T}_{\delta,\delta_c}, \mathcal{U}_{\delta_c}^{(k)}$ and compositions thereof). The minimal negativity over all such decomposition is the "channel QROM" $\mathcal{R}(\mathcal{U}|\{\mathcal{A}_x\})$.

Given any decomposition, a protocol canceling the bias on $\langle P \rangle$ due to noisy gates is obtained following steps 1, 3, and 4 of the previous algorithm, but replacing step 2 with (2′) Run the noisy quantum circuit $\mathcal{A}_x$.

As before, after $M \geq 2\ln(2/\epsilon)\|q\|_1^2/\Delta^2$ runs, the estimate is within error Δ of $\langle P \rangle$ with probability $1 - \epsilon$. $\mathcal{R}(\mathcal{U}|\{\mathcal{A}_x\})^2$ is then the minimal sampling overhead. In Supplemental Material Sec. VII [22] we obtain a (block) decomposition of $\mathcal{U}$ by separately finding a quasiprobability decomposition for T gates, 1- and 2-qubit Clifford gates. For $\delta < \delta_{\rm th}$, the corresponding sampling overhead [which upper bounds $\mathcal{R}(\mathcal{U}|\{\mathcal{A}_x\})$] is

$$\left(\frac{2-\delta}{(1-\delta)(1-\delta_c)^2} - 1\right)^{2t} \left(\frac{1+\frac{\delta_c}{2}}{1-\delta_c}\right)^{2n_c^{(1)}} \left(\frac{1+\frac{7\delta_c}{8}}{1-\delta_c}\right)^{2n_c^{(2)}}.$$

For $\delta_c = 0$ we recover $1/(1-\delta)^{2t}$, as expected from Eq. (7) [40]. The T-gate overhead is rather close to the one found for ideal Clifford gates.

Magic state distillation protocols [1] can be combined with error mitigation, in order to decrease the initial T-state error. In Fig. 3, we illustrate this by presenting the maximum number of Cliffords and T gates whose noise can be mitigated with moderate overhead ($\leq 10^2$) after zero, one or two rounds of the Bravyi-Haah $14 \mapsto 2$ magic state distillation protocol [7]. To highlight the performance of the error mitigation stage, we assumed an ideal distillation protocol. However, we show in Supplemental Material Sec. VIII [22] that also the noise in the Clifford unitaries required in the distillation round can be error mitigated.

Quantum-assisted simulation.—What happens when the T count t of the ideal quantum circuit exceeds the available resources? That is indeed a generic situation. Specifically, suppose we have at our disposal only $r < t$ noisy T states and ideal Cliffords to simulate the circuit. We call this task "quantum-assisted simulation," as it interpolates between classical simulation (where $r = 0$) and error mitigation (where $r = t$).

Injection of insufficient T states, as well as noise on each T, both lead to bias on the expectation value $\langle P \rangle$, which can

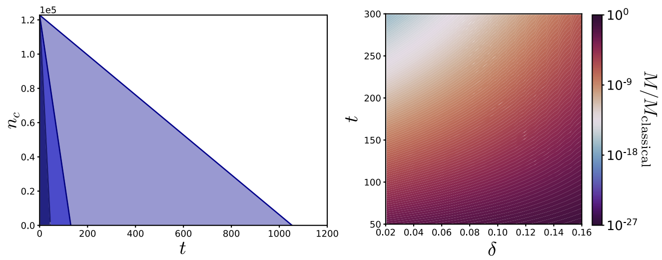

FIG. 3. Left: error mitigation and distillation. Maximum number of T gates (t) and Cliffords (n_c) that can be error mitigated with moderate overhead (≤ 100) after 0, 1, and 2 rounds of ideal Bravyi-Haah $14 \mapsto 2$ distillation (from dark to lighter). The initial noise parameters are $\delta_c = 10^{-5}$ and $\delta = 5 \times 10^{-2}$. We took the worst-case scenario where n_c are all 2-qubit Cliffords. Right: classical vs quantum-assisted simulation. Ratio between the quantum-assisted and classical number of samples $M/M_{\rm classical}$ needed to estimate the average value of a Pauli observable with given precision, as a function of the noise δ and the T count t of the circuit, for $t/r = 3$ (inject a third of the required T states).

be corrected by quasiprobability methods. To do so, find a decomposition of the form in Eq. (2) with $\rho_t = \tau^{\otimes t}$ and $\sigma_r = \tau_\delta^{\otimes r}$, with r a fraction of t. Then, apply the algorithm steps 1–4 above. By construction, in step 2 of the protocol at most r noisy magic states are injected, rather than the t ideally required. Nevertheless, after taking

$$M = 2\ln(2/\epsilon)\mathcal{R}(\tau^{\otimes t}|\tau_\delta^{\otimes r})^2/\Delta^2 \tag{10}$$

samples, the output average will be within Δ of the ideal average with probability $1 - \epsilon$. From

$$\mathcal{R}(\tau^{\otimes t}|\tau_\delta^{\otimes r}) \leq [\mathcal{R}(\tau^{\otimes t/r}|\tau_\delta)]^r, \tag{11}$$

explicit protocols can obtained by decomposing each t/r T state using a single noisy T state.

We find analytical decompositions upper bounding the QROM for $t/r = 2$ and $t/r = 3$ (see Supplemental Material Sec. VI [22]). The performance of the corresponding protocols is compared to the best-known classical algorithms based on the ROM in Fig. 3. For $\delta = 10^{-2}$ and $t/r = 3$ (inject a third of the required T states), we get a scaling overhead 1.45715^t, compared to 1.667^t of the best classical algorithm given in Ref. [18]. Even injecting a fraction of the required quantum resources leads to orders of magnitude improvements over the classical protocol.

The previous protocol injects fewer magic states at the price of a sampling overhead. Nevertheless, it still requires an n-qubit quantum computer. However, suppose we only have $r < n$ noisy qubits and we want to use them to assist a classical computer to estimate outcome probabilities of an ideal circuit with T count $t = kr$ ($k = 2, 3, \ldots$). We can proceed as follows. If $P = \Pi_1 - \Pi_{-1}$, in steps 2-3 of the previous algorithm we need to sample the measurement outcome $p = \pm 1$ with probability

$$\mathrm{Tr}[\Pi_p \mathcal{A}(\eta_x \otimes |0\rangle\langle 0|^{\otimes n})], \quad (12)$$

where $\mathcal{A}$ is some (adaptive) Clifford circuit and $\eta_x \in \mathcal{Q}_t(\tau_\delta^{\otimes s})$ with $s \leq r$. To sample from the above, we make use of the extended Gottesman-Knill theorem [41] (also see [38,39,42]). Given a description of $\mathcal{A}$ we classically obtain in poly$(n+t)$ time a description of a new adaptive Clifford circuit $\bar{\mathcal{A}}$ involving at most s measurements. Then, from a sample of $\mathrm{Tr}[\Pi^{(p)}\bar{\mathcal{A}}(\tau_\delta^{\otimes s})]$ and a poly$(n+t)$ computation, one can obtain a sample of Eq. (12). Steps 2 and 3 have now been reduced to a task that can be performed on the s-qubit noisy quantum computer. The number of required samples is again quantified by the QROM $\mathcal{R}(\tau^{\otimes t/r}|\tau_\delta)^r$.

Discussion.—Our analysis led to practical protocols for error mitigation in currently classical intractable regimes and quantum-assisted simulation outperforming the best-known quasiprobability-based algorithms. These proposals appear relevant for a regime in which full magic state distillation is unavailable. Alternative quantum-assisted robustness of magic can also be defined, e.g., by restricting stabilizer operations to account for specific architecture limitations. In this regard it should be also noted that our work, like all quasiprobability-based error cancellation methods, is limited by the requirement of well-characterized noise. We suggest that similar approaches can be envisioned to boost alternative simulation methods based on the stabilizer rank [36,38,39,43,44] or on generalized Wigner functions [45,46].

Our work was supported by ERC Grant EQEC No. 682726. A.C. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 390534769. We thank B. Terhal for useful discussions.

Note added.—Recently, we became aware of an independent effort to use error mitigation for universal quantum computing via encoded Clifford + T circuits [47].

[1] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature (London) **549**, 172 (2017).
[2] B. Eastin and E. Knill, Restrictions on Transversal Encoded Quantum Gate Sets, Phys. Rev. Lett. **102**, 110502 (2009).
[3] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).
[4] J. R. Seddon and E. T. Campbell, Quantifying magic for multi-qubit operations, Proc. R. Soc. A **475**, 20190251 (2019).
[5] D. Gottesman, The Heisenberg representation of quantum computers, in *Proceedings of the 22nd International Colloquium on Group Theoretical Methods in Physics* (International Press, Cambridge, 1998), pp. 32–43.
[6] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, Phys. Rev. A **71**, 022316 (2005).
[7] S. Bravyi and J. Haah, Magic-state distillation with low overhead, Phys. Rev. A **86**, 052329 (2012).
[8] A. G. Fowler, S. J. Devitt, and C. Jones, Surface code implementation of block code state distillation, Sci. Rep. **3**, 1939 (2013).
[9] A. M. Meier, B. Eastin, and E. Knill, Magic-state distillation with the four-qubit code, Quantum Inf. Comput. **13**, 195 (2013).
[10] D. Litinski, Magic state distillation: Not as costly as you think, Quantum **3**, 205 (2019).
[11] J. O'Gorman and E. T. Campbell, Quantum computation with realistic magic-state factories, Phys. Rev. A **95**, 032338 (2017).
[12] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).
[13] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature (London) **567**, 491 (2019).
[14] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).
[15] R. Takagi, Optimal resource cost for error mitigation, Phys. Rev. Research **3**, 033178 (2021).
[16] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, J. Phys. Soc. Jpn. **90**, 032001 (2021).
[17] M. Howard and E. Campbell, Application of a Resource Theory for Magic States to Fault-Tolerant Quantum Computing, Phys. Rev. Lett. **118**, 090501 (2017).
[18] M. Heinrich and D. Gross, Robustness of magic and symmetries of the stabiliser polytope, Quantum **3**, 132 (2019).
[19] S. Hakkaku and K. Fujii, Comparative Study of Sampling-Based Simulation Costs of Noisy Quantum Circuits, Phys. Rev. Applied **15**, 064027 (2021).
[20] A complementary approach analyzing error mitigation of decoding and approximating errors in the fault-tolerant regime has been discussed in Ref. [21].
[21] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation for fault-tolerant quantum computing, arXiv:2010.03887.
[22] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.127.200506, which includes Refs. [23,24], for the properties of the QROM and further analytical and numerical results.
[23] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, CalTech, 1997.
[24] X. Zhou, D. W. Leung, and I. L. Chuang, Methodology for quantum logic gate construction, Phys. Rev. A **62**, 052316 (2000).
[25] See Ref. [1]. For a reminder, see Supplemental Material Sec. III [22].
[26] R. Takagi and Q. Zhuang, Convex resource theory of non-Gaussianity, Phys. Rev. A **97**, 062337 (2018).

[27] F. Albarelli, M. G. Genoni, M. G. A. Paris, and A. Ferraro, Resource theory of quantum non-Gaussianity and Wigner negativity, Phys. Rev. A **98**, 052350 (2018).

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

[29] A. Mari and J. Eisert, Positive Wigner Functions Render Classical Simulation of Quantum Computation Efficient, Phys. Rev. Lett. **109**, 230503 (2012).

[30] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, 2011).

[31] If $|T\rangle$ is affected by a different noise, it can be brought into this standard form by applying the identity or the Clifford unitary $\exp[-i\pi/4]SX$ with probability $1/2$, as also remarked in Ref. [7]. Here S is the phase gate and X the Pauli x unitary.

[32] Perhaps one cannot directly sample η_x. In fact, according to the definition of $\mathcal{Q}_t(\sigma_r)$, we may be able to only prepare some states ϕ_i such that $\eta_x = \sum_i h_i^{(x)}\phi_i$ for a probability $\{h_i^{(x)}\}_i$. If that is the case, sample ϕ_i with probability $h_i^{(x)}$. An application of Hoeffding's inequality shows that this sampling has the same overhead as the original one.

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

[34] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, New York, 2004).

[35] It is possible to alleviate this problem by exploiting the symmetries of the state $\tau^{\otimes t}$ [18].

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

[37] For example, the $k = 3$ block decomposition error mitigates a quantum circuit with depolarizing noise $\delta = 10^{-2}$ on each T state with sampling overhead 1.0201^t, while $k = 1$ achieves 1.0203^t. Formally, we conjecture $\mathcal{R}(\tau^{\otimes t}|\tau_\delta^{\otimes t}) \approx (1+\delta)^t$ as $\delta \to 0$.

[38] S. Bravyi and D. Gosset, Improved Classical Simulation of Quantum Circuits Dominated by Clifford Gates, Phys. Rev. Lett. **116**, 250501 (2016).

[39] H. Pashayan, O. Reardon-Smith, K. Korzekwa, and S. D. Bartlett, Fast estimation of outcome probabilities for quantum circuits, arXiv:2101.12223.

[40] For the sake of comparison, setting $\delta = 10^{-2}$ and $\delta_c = 10^{-3}$ we get an overhead $1.02845^t \times 1.00301^{n_c^{(1)}} \times 1.00376^{n_c^{(2)}}$.

[41] M. Yoganathan, R. Jozsa, and S. Strelchuk, Quantum advantage of unitary Clifford circuits with magic state inputs, Proc. R. Soc. A **475**, 20180427 (2019).

[42] S. Bravyi, G. Smith, and J. A. Smolin, Trading Classical and Quantum Computational Resources, Phys. Rev. X **6**, 021043 (2016).

[43] S. Bravyi, G. Smith, and J. A. Smolin, Trading Classical and Quantum Computational Resources, Phys. Rev. X **6**, 021043 (2016).

[44] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, Simulation of quantum circuits by low-rank stabilizer decompositions, Quantum **3**, 181 (2019).

[45] R. Raussendorf, J. Bermejo-Vega, E. Tyhurst, C. Okay, and M. Zurel, Phase-space-simulation method for quantum computation with magic states on qubits, Phys. Rev. A **101**, 012350 (2020).

[46] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating Outcome Probabilities of Quantum Circuits Using Quasiprobabilities, Phys. Rev. Lett. **115**, 070501 (2015).

[47] C. Piveteau, D. Sutter, S. Bravyi, J. M. Gambetta, and K. Temme, preceding Letter, Error Mitigation for Universal Gates on Encoded Qubits, Phys. Rev. Lett. **127**, 200505 (2021).