Scalable Measures of Magic Resource for Quantum Computers

Tobias Haug* and M.S. Kim
Quantum Optics and Laser Science (QOLS), Blackett Laboratory, Imperial College London, London SW7 2AZ, United Kingdom

(Received 27 April 2022; accepted 6 December 2022; published 3 January 2023)

Nonstabilizerness or magic resource characterizes the amount of non-Clifford operations needed to prepare quantum states. It is a crucial resource for quantum computing and a necessary condition for quantum advantage. However, quantifying magic resource beyond a few qubits has been a major challenge. Here, we introduce efficient measures of magic resource for pure quantum states with a sampling cost that is independent of the number of qubits. Our method uses Bell measurements over two copies of a state, which we implement in experiment together with a cost-free error-mitigation scheme. We show the transition of classically simulable stabilizer states into intractable quantum states on the IonQ quantum computer. For applications, we efficiently distinguish stabilizer and nonstabilizer states with low measurement cost even in the presence of experimental noise. Further, we propose a variational quantum algorithm to maximize our measure via the shift rule. Our algorithm can be free of barren plateaus even for highly expressible variational circuits. Finally, we experimentally demonstrate a Bell-measurement protocol for the stabilizer Rényi entropy as well as the Wallach-Meyer entanglement measure. Our results pave the way to understanding the nonclassical power of quantum computers, quantum simulators, and quantum many-body systems.

DOI: 10.1103/PRXQuantum.4.010301

I. INTRODUCTION

The simulation of quantum states is in general intractable for classical computers. However, particular classes of states can be efficiently simulated classically. An important example concerns stabilizer states, which are generated from Clifford operations [1,2]. The number of non-Clifford operations needed to prepare a state can be quantified by measures of nonstabilizerness or magic resource. Henceforth, we abbreviate the term "magic resource" to "magic." Magic can be related to the difficulty of classical simulation of quantum states [3–9] and to quantum chaos [10–12]. Further, magic is a precious resource required to realize universal unitaries [13,14] in fault-tolerant quantum computers [15–18].

To characterize magic, various measures have been proposed [6,10,19–28]. However, most measures require solving an optimization program, access to the amplitudes of the quantum state, and a computational cost that scales exponentially with the system size. Recently, stabilizer entropy has been proposed as an experimentally accessible measure [10]; however, its randomized measurement protocol scales exponentially with the number of qubits [29].

Rapid progress has been made in experimental demonstrations of noisy intermediate-scale quantum computers [30,31] and fault-tolerant quantum computers [32–35]. A major challenge is to benchmark the power of quantum computers and track their progress [36,37]. For fault-tolerant quantum computers, magic characterizes the capability to implement universal quantum gates [14]. For noisy intermediate-scale quantum computers, an important benchmark is to prepare states that are difficult to simulate classically [38], which can related to particular measures of magic [4–6,11]. The relationship between the complexity of quantum states and magic is also of major interest in quantum many-body physics [39,40].

Here, we introduce Bell magic as an efficiently computable measure of magic for quantum computers. The number of measurements is independent of the number of qubits and the classical postprocessing time scales linearly. Bell magic is a faithful measure of magic for pure states, while for mixed states it is not faithful in general. For noisy quantum computers, we use Bell measurements over two copies of a state and error mitigation to compute Bell magic efficiently. We propose practical applications of Bell magic for state discrimination and to find highly magical states with variational quantum algorithms. Our

*tobias.haug@u.nus.edu

TABLE I. Definitions of symbols.

Name	Symbol
Bell magic	$\mathcal{B}$
Additive Bell magic	$\mathcal{B}_a$
Probability of outcome $\mathbf{r} \in \{0,1\}^{2N}$	$P(\mathbf{r})$
Estimation error	$\Delta\mathcal{B}$
Pauli string	$\sigma_{\mathbf{n}}$
$\frac{1}{\sqrt{2}}(\lvert 0\rangle + e^{-i\frac{\pi}{4}}\lvert 1\rangle)$	$\lvert T\rangle$
$\cos(\frac{\theta}{2})\lvert 0\rangle + e^{-i\frac{\pi}{4}}\sin(\frac{\theta}{2})\lvert 1\rangle, \theta = \arccos(\frac{1}{\sqrt{3}})$	$\lvert R\rangle$
$\cos(\frac{\phi}{2})\lvert 0\rangle + \sin(\frac{\phi}{2})\lvert 1\rangle$	$\lvert A_\phi\rangle$
Number of qubits	N
Number of measurements	N_Q
Resampling steps	N_R
Number of magic states	N_A
Number of T gates	N_T
Depolarizing error	p
Classification error probability	P_E

variational quantum algorithm can have large gradients even for highly expressible ansatz circuits. With the IonQ quantum computer, we study the transition from stabilizer states to intractable quantum states. Further, we experimentally distinguish different types of states using magic. Our results provide an indispensable tool to characterize the magic of quantum computers, quantum simulators, and numerical simulations of quantum many-body systems.

We first define preliminary concepts in Sec. II. Then, we introduce Bell magic in Sec. III, its measurement scheme in Sec. IV, and the method to mitigate errors in Sec. V. We numerically and experimentally demonstrate the measurement of Bell magic in Sec. VI. Then, we show applications of Bell magic for state discrimination in Sec. VII and for finding highly magical states with variational quantum algorithms in Sec. VIII. Finally, the results are discussed in Sec. IX. We give an overview of the definitions of symbols in Table I.

II. PRELIMINARIES

We define the Pauli matrices $\sigma_{00} = I_2$, $\sigma_{01} = \sigma^x$, $\sigma_{10} = \sigma^z$, and $\sigma_{11} = \sigma^y$. The 4^N Pauli strings are N-qubit tensor products of Pauli matrices, which we define as $\sigma_{\mathbf{n}} = \bigotimes_{j=1}^{N} \sigma_{\mathbf{n}_{2j-1}\mathbf{n}_{2j}}$, with $\mathbf{n} \in \{0,1\}^{2N}$. The product of two Pauli strings $\sigma_{\mathbf{r}}$ and $\sigma_{\mathbf{q}}$ can be written as $\sigma_{\mathbf{r}}\sigma_{\mathbf{q}} = \sigma_{\mathbf{r}\oplus\mathbf{q}}$ up to a multiplication with $\{\pm 1, \pm i\}$, where $\oplus$ denotes a bitwise exclusive OR. The Bell states are given by $\lvert\sigma_{00}\rangle = 1/\sqrt{2}\,(\lvert 00\rangle + \lvert 11\rangle)$, $\lvert\sigma_{01}\rangle = 1/\sqrt{2}\,(\lvert 00\rangle - \lvert 11\rangle)$, $\lvert\sigma_{10}\rangle = 1/\sqrt{2}\,(\lvert 01\rangle + \lvert 10\rangle)$, and $\lvert\sigma_{11}\rangle = 1/\sqrt{2}\,(\lvert 01\rangle - \lvert 10\rangle)$ and we define the product of Bell states as $\lvert\sigma_{\mathbf{r}}\rangle = \lvert\sigma_{r_1 r_2}\rangle \otimes \cdots \otimes \lvert\sigma_{r_{2N-1} r_{2N}}\rangle$.

The stabilizer states $\lvert\psi_{\mathrm{STAB}}\rangle$ are defined by a commuting subgroup G of $|G| = 2^N$ Pauli strings σ. We have $\langle\psi_{\mathrm{STAB}}\lvert\sigma\rvert\psi_{\mathrm{STAB}}\rangle = \pm 1$ for $\sigma \in G$ and $\langle\psi_{\mathrm{STAB}}\lvert\sigma'\rvert\psi_{\mathrm{STAB}}\rangle = 0$ for $\sigma' \notin G$ [1]. Any $\sigma_{\mathbf{r}}, \sigma_{\mathbf{r}'} \in G$ commute $[\sigma_{\mathbf{r}}, \sigma_{\mathbf{r}'}] = 0$. The unitaries that transform stabilizer states into other stabilizer states into stabilizer states are the Clifford circuits U_C. They can be generated by combining the Clifford gate set consisting of the S gate ($S = \mathrm{diag}[1, \exp(-i\pi/2)]$), the Hadamard gate H, and the controlled-NOT (CNOT) gate, which can be efficiently simulated on classical computers [1]. Universal unitaries are realized by combining Clifford circuits with non-Clifford resources such as the T gate ($T = \mathrm{diag}[1, \exp(-i\pi/4)]$) [41]. Examples of stabilizer and nonstabilizer single-qubit states are shown in Fig. 1(a).

Measures of magic characterize the distance to the set of stabilizer states or unitaries [6,10,19–24,26,27]. Measures of magic are zero for stabilizer states and greater than zero otherwise. Further, they should be nonincreasing under Clifford operations [22]. Most schemes for fault-tolerant quantum computers are based on stabilizers [15], where universal quantum computation is enabled by consuming magic states [13]. Lower bounds on the number of magic states necessary to generate a state or unitary can be related to (sub)additive measures of magic such as the robustness of magic [20], the stabilizer entropy [10], or "mana" [22]. Further, measures such as the stabilizer rank [5], the robustness of magic [20], or negativity [4] can be related to the computational difficulty of particular simulation algorithms for quantum states. For these algorithms, the simulation cost increases drastically with the number of non-Clifford gates. To understand the quantum and classical cost of simulating and preparing quantum states, one would like to compute measures of magic for large quantum systems. However, the computational cost scales in general exponentially with the qubit number for the aforementioned measures.

We now introduce a measure of magic that can be efficiently computed. We make use of entangled measurements over multiple copies of states, which can reveal information not accessible by single copies [42–45]. In particular, measurements in the basis of Bell states are known to give access to important properties, which for single copies would require exponential resources [45–49]. To realize Bell measurements, we first prepare the tensor product $\rho_A \otimes \rho_B$ of two states ρ_A and ρ_B. Then, we apply the Bell transformation with the unitary $U_{\mathrm{Bell}} = \bigotimes_{n=1}^{N}(H \otimes I_2) \times \mathrm{cnot}$ on all N pairs of qubits [see Fig. 1(b)]. This transformation can also be realized in atomic or photonic systems using a beam splitter [48]. Then, we measure N_Q times in the computational basis and record the outcomes $\mathbf{r}^j \in \{0,1\}^{2N}$ with $j = 1, \ldots, N_Q$. Here, $\mathbf{r}^j_{2n-1}$, $n = 1, \ldots, N$ is the outcome of the nth qubit of subsystem A and $\mathbf{r}^j_{2n}$ of subsystem B. This measurement setting realizes a SWAP test to compute the trace overlap of the two states

$$\mathrm{tr}(\rho_A \rho_B) = 1 - 2P_{\mathrm{odd}}, \tag{1}$$

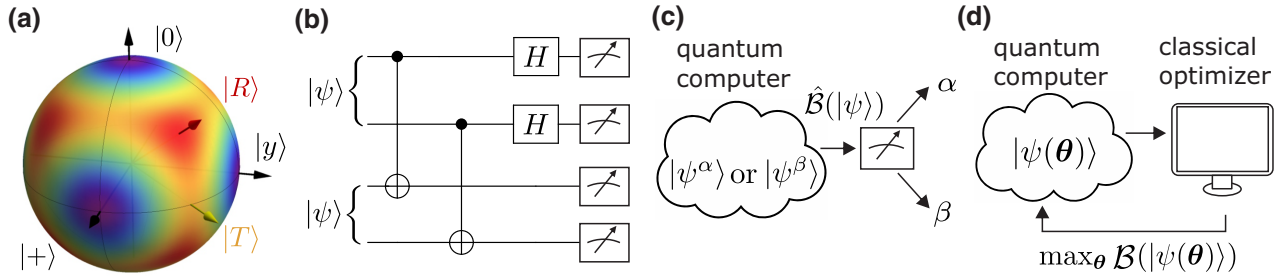

FIG. 1. (a) Bell magic $\mathcal{B}$ [Eq. (3)] is a measure of nonstabilizerness for pure quantum states. The color shows $\mathcal{B}$ for the Bloch sphere of a single qubit with Eq. (6), where the magnitude of $\mathcal{B}$ is increasing from blue to red. Stabilizer states along the main axes, such as $|0\rangle$, $|+\rangle$, and $|y\rangle$, have zero magic, while the magic states $|T\rangle$ and $|R\rangle$ have nonzero magic. (b) The quantum circuit to measure $\mathcal{B}$ of an N-qubit state $|\psi\rangle = U|0\rangle$ with unitary U by preparing two copies $|\psi\rangle \otimes |\psi\rangle$ and measuring in the Bell basis. The number of required measurements N_Q is independent of N. (c) The state-discrimination scheme to determine whether a given state $|\psi\rangle$ belongs to one of two classes α, β, with respective magic $\mathcal{B}^\alpha > \mathcal{B}^\beta$. When the estimated Bell magic $\hat{\mathcal{B}}$ is greater than a threshold $\hat{\mathcal{B}} > \mathcal{B}^*$, we classify the state as class α; otherwise, class β. (d) The variational quantum algorithm to maximize magic by optimizing parameter $\boldsymbol{\theta}$ of the parametrized quantum circuit $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle$. The algorithm runs via a feedback loop between measurements on the quantum computer and a classical optimization routine.

where P_{odd} is the probability that $\mathbf{q}^j \in \{0,1\}^N$ has odd parity, where $\mathbf{q}^j_n = \mathbf{r}^j_{2n-1} \cdot \mathbf{r}^j_{2n}$ is a bit-wise AND of the outcomes of each subsystem [46]. The SWAP test on two copies of the same state $\rho \otimes \rho$ can give us the purity $\text{tr}(\rho^2)$ as well as the entanglement 2-Rényi entropy $\text{tr}(\rho_k^2)$ over a subsystem ρ_k (see Appendix P).

Now, we perform the Bell measurement on two copies of a pure state $|\psi\rangle \otimes |\psi\rangle$. The outcome $\mathbf{r}$ appears with a probability [49]

$$P(\mathbf{r}) = \langle\psi|\langle\psi|O_\mathbf{r}|\psi\rangle|\psi\rangle = 2^{-N}|\langle\psi|\sigma_\mathbf{r}|\psi^*\rangle|^2, \quad (2)$$

where $O_\mathbf{r} = |\sigma_\mathbf{r}\rangle\langle\sigma_\mathbf{r}|$ is the projector onto a product of Bell states and $|\psi^*\rangle$ denotes the complex conjugate of $|\psi\rangle$. For any state, we have $0 \le P(\mathbf{r}) \le 2^{-N}$ and there are between 2^N and 4^N outcomes $\mathbf{r}$ with $P(\mathbf{r}) > 0$. For any set of bit strings $\{\mathbf{r}^j\}_{j=1}^{N_Q}$ sampled in the Bell basis from a pure stabilizer state $|\psi_{\text{STAB}}\rangle \in G$, the Pauli strings of its binary additions must commute, i.e., $[\sigma_{\mathbf{r}^k \oplus \mathbf{r}^l}, \sigma_{\mathbf{r}^n \oplus \mathbf{r}^m}] = 0\ \forall k,l,n,m$ (see Appendix A or Ref. [49]). Conversely, finding at least one noncommuting Pauli string implies that the measured quantum state is not a pure stabilizer state, as the commuting subgroup G contains at most 2^N elements. This motivates the idea that the probability of observing noncommuting Pauli strings is a measure of distance to the set of pure stabilizer states.

III. BELL MAGIC

We now define Bell magic $\mathcal{B}$ as

$$\mathcal{B} = \sum_{\substack{\mathbf{r},\mathbf{r}',\mathbf{q},\mathbf{q}' \\ \in\{0,1\}^{2N}}} P(\mathbf{r})P(\mathbf{r}')P(\mathbf{q})P(\mathbf{q}') \left\|[\sigma_{\mathbf{r}\oplus\mathbf{r}'}, \sigma_{\mathbf{q}\oplus\mathbf{q}'}]\right\|_\infty, \quad (3)$$

where the infinity norm is zero, $\|[\sigma_\mathbf{r}, \sigma_\mathbf{q}]\|_\infty = 0$, when the two Pauli strings commute, $[\sigma_\mathbf{r}, \sigma_\mathbf{q}] = 0$, and $\|[\sigma_\mathbf{r}, \sigma_\mathbf{q}]\|_\infty = 2$ otherwise. As a measure of magic [22], $\mathcal{B}$ is faithful with $\mathcal{B}(|\psi_{\text{STAB}}\rangle) = 0$ only for pure stabilizer states $|\psi_{\text{STAB}}\rangle$ and $\mathcal{B} > 0$ otherwise. As shown in Appendix B, $\mathcal{B}$ is also invariant under Clifford circuits U_C that map stabilizers to stabilizers, i.e., $\mathcal{B}(U_C|\psi\rangle) = \mathcal{B}(|\psi\rangle)$. Further, Bell magic is constant under composition with any stabilizer state $|\psi_{\text{STAB}}\rangle$, i.e., $\mathcal{B}(|\psi\rangle \otimes |\psi_{\text{STAB}}\rangle) = \mathcal{B}(|\psi\rangle)$ (see Appendix D). We numerically test an extensive number of states and find that in all cases, Bell magic is on average nonincreasing under measurements in the computational basis over a set of qubits, i.e., $\mathcal{B}(|\psi\rangle) \ge \sum_n q_n \mathcal{B}(\mathcal{M}_n(|\psi\rangle))$, where $\Pi_n = |n\rangle\langle n| \otimes I_{N-1}$ is a projector on computational basis state $|n\rangle$, I_{N-1} is the identity operator for $N-1$ qubits, $q_n = \langle\psi|\Pi_n|\psi\rangle$ is the measurement probability, and $\mathcal{M}_n(|\psi\rangle) = q_n^{-1/2}\Pi_n|\psi\rangle$ is the projected state.

We further define the additive Bell magic:

$$\mathcal{B}_a = -\log_2(1-\mathcal{B}). \quad (4)$$

$\mathcal{B}_a$ shares all properties with $\mathcal{B}$ and, further, is additive with $\mathcal{B}_a(|\psi\rangle \otimes |\phi\rangle) = \mathcal{B}_a(|\psi\rangle) + \mathcal{B}_a(|\phi\rangle)$, which is proven in Appendix C. $\mathcal{B}_a$ has the operational meaning as the number of initial magic states $|T\rangle = T|+\rangle = 1/\sqrt{2}(|0\rangle + e^{-i(\pi)/4}|1\rangle)$ within any Clifford circuit U_C. For a state $|T\rangle^{\otimes k} \otimes |0\rangle^{\otimes(N-k)}$ consisting of a tensor product of k magic states and otherwise the stabilizer state $|0\rangle$, additive Bell magic is given by

$$\mathcal{B}_a(U_C|T\rangle^{\otimes k} \otimes |0\rangle^{\otimes(N-k)}) = k. \quad (5)$$

The additive Bell magic of an N-qubit product state $|\psi_{\text{sp}}(\boldsymbol{\theta},\boldsymbol{\varphi})\rangle = \bigotimes_{n=1}^N (\cos(\theta_n/2)|0\rangle + e^{-i\varphi_n}\sin(\theta_n/2)|1\rangle)$ is

given by

$$\mathcal{B}_a(\boldsymbol{\theta},\boldsymbol{\varphi}) = -\sum_{n=1}^{N} \log_2[1 - \frac{1}{32}\sin^2(\theta_n)(35 + 28\cos(2\theta_n) + \cos(4\theta_n) - 8\cos(4\varphi_n)\sin^4(\theta_n))], \tag{6}$$

which becomes maximal with $B_a^R = N\log_2(27/11) \approx 1.3N$ for the magic state $|R\rangle^{\otimes N} = (\cos(\theta/2)|0\rangle + e^{-i(\pi)/4}\sin(\theta/2)|1\rangle)^{\otimes N}$ with $\theta = \arccos(1/\sqrt{3})$. For $N = 1$, the state of maximal magic is $|R\rangle$ and for $N = 3$ it is the Hoggar state, coinciding with the states of maximal robustness of magic [20]. We report the explicit forms of the pure states of maximal magic up to $N = 4$ in Appendix G. For pure states, we find that Bell magic is upper bounded by

$$\mathcal{B}^{\text{pure}} \leq 4^N \frac{(1 + 2^{-N} - 2\times 4^{-N})^2}{(4^N - 1)(1 + 2^{-N})^2}. \tag{7}$$

This bound is not tight but we find that it is saturated for $N = 1$ and $N = 3$.

Note that as Bell magic is a measure of distance to the set of pure stabilizer states, it is in general nonzero for probabilistic mixtures of stabilizer states and thus not a proper measure of magic for generic mixed states. For example, the maximally mixed state $\rho_m = I_N 2^{-N}$ with the N-qubit identity I_N (which can be written as a probabilistic mixture of pure stabilizer states) has the maximal Bell magic with $\mathcal{B}(\rho_m) = 1 - 4^{-N}$ and $\mathcal{B}_a(\rho_m) = 2N$ (see Appendix E)

However, we can define an extension of Bell magic that is indeed faithful for a class of mixed stabilizer states. We consider N-qubit mixed stabilizer states of the form $\rho_{\text{STAB}} = U_C|\psi_{\text{STAB}}\rangle\langle\psi_{\text{STAB}}| \otimes I_K 2^{-K} U_C^\dagger$, where $|\psi_{\text{STAB}}\rangle$ is a $N - K$ qubit pure stabilizer state and U_C is an arbitrary N-qubit Clifford circuit. These states can be written as $\rho_{\text{STAB}} = 2^{-N}(I + \sum_{\sigma\in G_0}\alpha_\sigma\sigma)$, where $\alpha_\sigma = \pm 1$ and $G_0 \subseteq G/\{I_N\}$ is a subset of a commuting subgroup of Pauli strings (excluding identity I_N) with $|G_0| \leq 2^N - 1$. We define the mixed Bell magic as

$$\mathcal{B}_m(\rho) = 1 - \frac{1 - \mathcal{B}(\rho)}{\text{tr}(\rho^2)^2}. \tag{8}$$

$\mathcal{B}_m$ shares all properties of $\mathcal{B}$ and, additionally, we have $\mathcal{B}_m(\rho_{\text{STAB}}) = 0$ (see Appendix F). We numerically check various states and find that in all cases $\mathcal{B}_m$ is nonincreasing when partially tracing out qubits. We also define the additive mixed Bell magic

$$\mathcal{B}_{\text{a,m}}(\rho) = -\log_2(1 - \mathcal{B}(\rho)) + 2\log_2(\text{tr}(\rho^2)). \tag{9}$$

Note that we can also use error mitigation to extract Bell magic from noisy states, which we show in Sec. V.

Input : Bitstrings $r, q \in \{0,1\}^{2N}$
Output: C
$\sigma_{\boldsymbol{r}} = \bigotimes_{n=1}^{N} \boldsymbol{\sigma}_{\boldsymbol{r}_{2n-1}\boldsymbol{r}_{2n}}$
$\sigma_{\boldsymbol{q}} = \bigotimes_{n=1}^{N} \boldsymbol{\sigma}_{\boldsymbol{q}_{2n-1}\boldsymbol{q}_{2n}}$
if $[\sigma_{\boldsymbol{r}}, \sigma_{\boldsymbol{q}}] = 0$ **then**
  $C \leftarrow 0$
else
  $C \leftarrow 2$
end

Algorithm 1. Check-Commute

IV. THEORY OF MEASURING MAGIC

Bell magic $\mathcal{B}$ can be efficiently estimated from measurements on quantum states. We give an unbiased estimator for $\mathcal{B}$ in Algorithms 1 and 2. We prepare two copies of the state, perform Bell measurements between them, and record the outcome. We repeat this step N_Q times, requiring in total $2N_Q$ copies of the state. Then, we postprocess the outcomes. We randomly draw four bit strings from the outcomes without replacement and check whether their addition commutes. This step is repeated N_R times, where the N_R are the resampling steps and we always draw the bit strings from all N_Q outcomes. $\mathcal{B}$ is then estimated as the probability of obtaining a noncommuting result.

As we show in the next paragraph, the number of measurements N_Q needed to estimate $\mathcal{B}$ with fixed accuracy is independent of the qubit number N, i.e., $N_Q \sim O(1)$. Further, the classical postprocessing scales as $O(N)$. At first glance this seems counterintuitive, as the number of possible outcomes scales exponentially with N. However, to estimate $\mathcal{B}$, we assign the bit strings only two possible values via the Check-Commute routine [corresponding to the norm of the commutator in Eq. (3)]. Bell magic is then estimated as the expectation value of the two values. Thus, the measurement process corresponds to a Bernoulli trial, with the same scaling of errors as estimating the expectation value of a coin flip or a Pauli operator.

Input : $j = 1,\ldots,N_{\text{Q}}$ bitstrings $r^j \in \{0,1\}^{2N}$ sampled from Bell measurement
  Resampling steps N_{R}
Output: Bell magic $\mathcal{B}$
  Additive Bell magic $\mathcal{B}_{\text{a}}$
$\mathcal{B} \leftarrow 0$
for $k = 1,\ldots,N_R$ **do**
  Choose randomly without replacement $n_1, n_2, n_3, n_4 \in \{1,\ldots,N_{\text{Q}}\}$
  $\mathcal{B} \leftarrow \mathcal{B} + \text{Check-Commute}(r^{n_1} \oplus r^{n_2}, r^{n_3} \oplus r^{n_4})$
end
$\mathcal{B} \leftarrow \mathcal{B}/N_{\text{R}}$
$\mathcal{B}_{\text{a}} \leftarrow -\log_2(1 - \mathcal{B})$

Algorithm 2. Bell magic

Now, we give analytic bounds on the estimation error of $\mathcal{B}$ for the case $N_R = N_Q/4$. We slightly modify Algorithm 2 such that the bit strings are not sampled at random but each of the N_Q bit strings is drawn exactly once. As we use each bit string only once, the outcomes of the Check-Commute routine are statistically independent and we can write the estimated Bell magic as a Bernoulli trial $\hat{\mathcal{B}} = 2p_{\rm nc} = 2M_{\rm nc}/M_{\rm total}$, where $p_{\rm nc}$ is the probability that an outcome does not commute, $M_{\rm nc}$ is the number of outcomes that do not commute and $M_{\rm total} = N_Q/4$ is the total number of repetitions. The standard deviation of the Bell magic for such Bernoulli experiments is given by

$$\mathrm{std}(\mathcal{B}) = 2\sqrt{\frac{p_{\rm nc}(1-p_{\rm nc})}{M_{\rm total}}} = \sqrt{\frac{8\mathcal{B}}{N_Q}(1-\frac{\mathcal{B}}{2})}. \tag{10}$$

The number of measurement samples N_Q needed to achieve an estimation error of at most $\Delta\mathcal{B}$ with a failure probability P_F is bounded by Hoeffding's inequality

$$P(|\hat{\mathcal{B}} - \mathcal{B}| \geq \Delta\mathcal{B}) = P_{\rm F} \leq 2\exp\left(-\frac{2\Delta\mathcal{B}^2 M_{\rm total}}{(a_{\max}-a_{\min})^2}\right), \tag{11}$$

where $a_{\max} = 2$ and $a_{\min} = 0$ are the maximal and minimal possible values of each trial of Algorithm 1. The upper bound for the needed samples is given by

$$N_Q \geq \frac{8}{\Delta\mathcal{B}^2}\log(\frac{2}{P_{\rm F}}). \tag{12}$$

In particular, the estimation error scales as $\Delta\mathcal{B} \propto N_Q^{-1/2}$ and is independent of the qubit number N. The above equations are derived for the choice $N_R = N_Q/4$. By increasing the number of postprocessing steps $N_R > N_Q/4$, the accuracy increases further. We numerically find that $N_R = 10N_Q$ provides estimates of the Bell magic close to the maximal possible accuracy. Thus, we find that the classical postprocessing has $O(N)$ complexity in time and memory.

The outcomes of the Bell measurement also allow us to compute the purity $\mathrm{tr}(\rho^2)$ via Eq. (1) at the same time. The purity is estimated as the probability of measuring outcomes of odd parity, which again is a Bernoulli trial. The estimation error of the purity scales as $\propto N_Q^{-1/2}$, which allows us to also estimate mixed Bell magic [Eq. (8)] efficiently.

V. ERROR MITIGATION

Next, we use the purity to mitigate errors from the outcomes of noisy quantum computers [50–52]. Our goal is to determine the Bell magic of the pure state $|\psi\rangle$ by measuring the state $\rho_{\rm dp} = (1-p)|\psi\rangle\langle\psi| + p\,\rho_m$ subject to global depolarizing noise with a probability p [41]. Depolarizing noise has been shown to be a good approximation in experiments on noisy quantum computers [51,52] and coherent errors can be turned into depolarizing errors via randomized compiling [53,54]. As is seen in Sec. VI, our experimental results are well described with a depolarizing model. We prepare two copies $\rho_{\rm dp}\otimes\rho_{\rm dp}$ and apply the Bell transformation, where we assume that the Bell measurement is noise free. From the Bell measurements, we determine the purity $\mathrm{tr}(\rho_{\rm dp}^2)$ via Eq. (1) as well as the Bell magic $\mathcal{B}^{\rm dp}$ of the noise-affected state $\rho_{\rm dp}$. The purity is related to the depolarization error via $\mathrm{tr}(\rho_{\rm dp}^2) = (1-p)^2 + (p(2-p))/2^N$. By inverting, we obtain the depolarization error

$$p = 1 - \frac{\sqrt{(2^N-1)(2^N\mathrm{tr}(\rho_{\rm dp}^2)-1)}}{2^N-1}. \tag{13}$$

The mitigated Bell $\mathcal{B}^{\rm mtg}$ of the noise-free state $|\psi\rangle$ is given by (see Appendix I)

$$\mathcal{B}^{\rm mtg} = \frac{1}{(1-p_c)^2}(\mathcal{B}^{\rm dp} - p_c^2\mathcal{B}(\rho_m) - 2p_c(1-p_c)\mathcal{B}^R), \tag{14}$$

where $p_c = 1-(1-p)^4$ and $\mathcal{B}^R = 1-(1-p_c)^{-1}(\sum_{\mathbf{q}} P_{\rm dp}(\mathbf{q})^2 - 4^{-N}p_c)$. Here, $P_{\rm dp}(\mathbf{q})$ is the probability of measuring bit string $\mathbf{q}$ of the noisy state. In the limit of many qubits N, we approximate $\mathcal{B}^R \approx \mathcal{B}(\rho_m) \approx 1$ and obtain

$$\mathcal{B}^{\rm mtg} \approx \frac{\mathcal{B}^{\rm dp} - p_c(2-p_c)}{(1-p_c)^2}. \tag{15}$$

We now give the scaling of the number of samples N_Q needed to estimate the mitigated magic with an error $\Delta\mathcal{B}$. We define the error $\Delta\mathcal{B}^{\rm dp} = |\hat{\mathcal{B}}^{\rm dp} - \mathcal{B}^{\rm dp}|$ of $\hat{\mathcal{B}}^{\rm dp}$ estimated by measuring the noisy quantum computer. We insert Eq. (15) and obtain the error of the mitigated magic $\Delta\mathcal{B}^{\rm dp} \approx (1-p)^8\Delta\mathcal{B}^{\rm mtg}$, where we use $(1-p_c)^2 = (1-p)^8$. The upper bound of $\Delta\mathcal{B}^{\rm dp}$ is given by Eq. (12), where we insert the mitigated error $\Delta\mathcal{B}^{\rm mtg}$. Note that the upper bound can be slightly violated to our approximations and the error in the estimation of p. However, we argue that the scaling of the error remains the same, which we confirm numerically. Thus, the number of samples needed to estimate the mitigated magic within error $\Delta\mathcal{B}^{\rm mtg}$ scales as

$$N_Q \propto \frac{1}{(1-p)^{16}\Delta\mathcal{B}^{\rm mtg2}}. \tag{16}$$

VI. DEMONSTRATION OF MEASURING BELL MAGIC

Now, we demonstrate the measurement of Bell magic. First, we numerically investigate in Fig. 2 the dependence of the estimation error on various parameters.

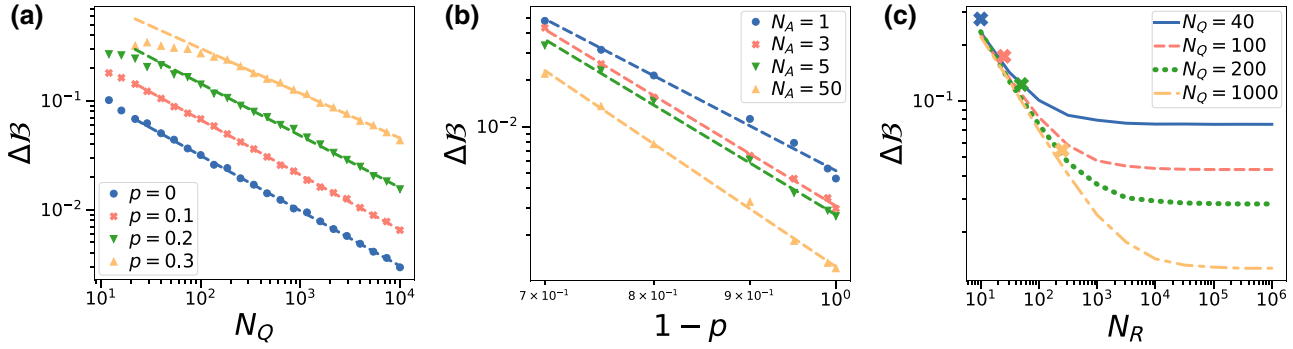

FIG. 2. (a) The simulation of the estimation error of mitigated Bell magic $\Delta\mathcal{B} = \langle|\hat{\mathcal{B}}^{\text{mtg}} - \mathcal{B}^{\text{exact}}|\rangle$ as a function of the number of measurements N_Q for varying depolarizing probability p. The dashed lines are linear fits with a slope of $b = \{-0.507, -0.505, -0.479, -0.412\}$ in descending order of the legend. We use a random Clifford circuit of $N = 50$ qubits applied on an initial state of $N_A = 3$ magic states with $\phi = \pi/4$ and $N_R = 10N_Q$. The error is averaged over 1000 repetitions. (b) A plot of $\Delta\mathcal{B}$ against the depolarizing error p for a varying number of magic states N_A and random U_C. The dashed lines are fits with slope $b = \{-6.32, -7.36, -7.307, -8.18\}$. We use $N = 50$, $N_Q = 10^4$, and $N_R = 10N_Q$. (c) A plot of $\Delta\mathcal{B}$ as a function of the classical resampling steps N_R after performing N_Q measurements on a quantum computer. The crosses are the theoretical values of the standard deviation given in Eq. (10) for $N_R = N_Q/4$. We use a random Clifford circuit of $N = 8$ qubits applied on an initial state of $N_A = 1$ magic states.

We measure the state $U_C|A_\phi\rangle^{\otimes N_A} \otimes |0\rangle^{N-N_A}$, with $N = 50$ qubits, where $|A_\phi\rangle = \cos(\phi/2)|0\rangle + \sin(\phi/2)|1\rangle$ is parameterized with angle ϕ, and U_C are random Clifford circuits realized as hardware-efficient quantum circuits [55] (for details, see Appendix H). We have $\mathcal{B}_a(|A_{\phi=0}\rangle) = 0$ and $\mathcal{B}_a(|A_{\phi=\pi/4}\rangle) = \mathcal{B}_a(|T\rangle) = 1$. We numerically simulate states of $N = 50$ qubits with tensor-network methods [56,57]. In Fig. 2(a), we plot the estimation error $\Delta\mathcal{B} = \langle|\hat{\mathcal{B}}^{\text{mtg}} - \mathcal{B}^{\text{exact}}|\rangle$ between the mitigated magic $\hat{\mathcal{B}}^{\text{mtg}}$ estimated from N_Q measurements and the exact value $\mathcal{B}^{\text{exact}}$. We show the error for a varying number of Bell measurements N_Q and depolarizing error p. We find that the fit matches the prediction $\Delta\mathcal{B} \propto N_Q^{-1/2}$ [Eq. (16)]. We find a slightly different slope for large p, which could be the result of not including estimation errors in the purity in our theory. In Fig. 2(b), we plot $\Delta\mathcal{B}$ against the depolarizing error p for a varying number of magic states N_A within a Clifford circuit. The fit to the data is close to the relation $\Delta\mathcal{B} \propto (1-p)^{-8}$ of Eq. (16). In Fig. 2(c), we plot $\Delta\mathcal{B}$ as a function of the resampling steps N_R for various numbers of measurements N_Q. We find that the theoretical model given in Eq. (10) matches our simulation accurately. Increasing $N_R > N_Q/4$ improves the accuracy, until it converges for large N_R to a constant value. Our numerics suggest that $N_R = 10N_Q$ is a good empirical choice that gives nearly the lowest possible error.

Next, we experimentally measure in Fig. 3 the magic of various states on the 11-qubit IonQ quantum computer [58]. We prepare two instances of the desired quantum state on the quantum computer, then apply the Bell transformation and measure each qubit. We investigate the additive Bell magic of various types of states. In particular, we measure different product states and the state of maximal Bell magic in Fig. 3(a), as well as stabilizer states with a variable number of injected T gates in Fig. 3(b). We note that while product states are not entangled, the other states are substantially entangled, which we confirm experimentally with the Meyer-Wallach measure [59] in Appendix P. The error mitigation with Eq. (14) substantially improves the results, matching the exact simulations quite well. In Fig. 3(a), we observe, for $|T\rangle^{\otimes N}$, a linear increase in $\mathcal{B}_a$ with N, highlighting its additive property. In Fig. 3(b), we study the transition of classically simulable stabilizer states to intractable quantum states [11]. We prepare a state of the form $U_C \prod_{n=1}^{N_T} U_T U_C^n |0\rangle$, where U_C^n is a randomly chosen Clifford circuit and $U_T = I_2 \otimes \cdots \otimes T \otimes \cdots \otimes I_2$ is a T gate placed at a random qubit. For the practical implementation of the circuit on quantum computers, see Appendix H. For $N_T = 0$, we prepare a stabilizer state with $\mathcal{B}_a = 0$. Nonstabilizer states are prepared for $N_T > 0$. We find that $\mathcal{B}_a$ grows nearly monotonously with N_T until it converges. We find that the converged $\mathcal{B}_a$ matches closely the Bell magic averaged over Haar-random states. Agreeing with our observations, recent results have shown that with increasing N_T, circuits will closely approximate unitaries sampled from the distribution of Haar-random states [12,60,61].

VII. STATE DISCRIMINATION

The efficient measurement of magic opens up new applications in state discrimination. When performing quantum computation or communicating over quantum networks, an important task is to verify whether a given unknown state possesses the desired properties such as a sufficient amount of magic [see Fig. 1(c)]. Now, our goal is to

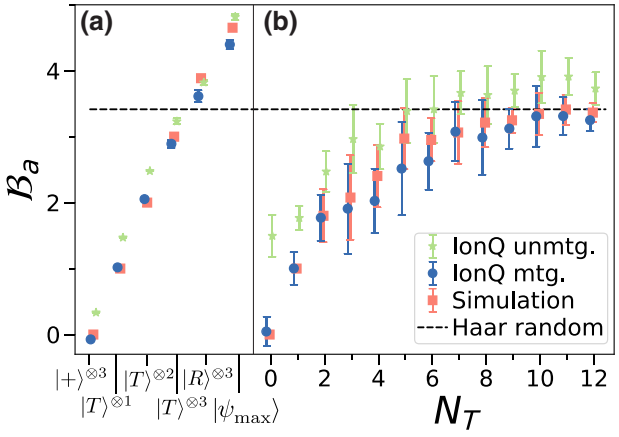

FIG. 3. An experiment to measure the additive Bell magic $\mathcal{B}_a$ on the IonQ quantum computer for various types of states. (a) In the left part of the graph, we show product states of stabilizer states $|+\rangle^{\otimes N}$ and magic states $|T\rangle^{\otimes N}$ and $|R\rangle^{\otimes N}$, as well as the state of maximal magic $|\psi_{\max}\rangle$ for $N = 3$. (b) In the right part of the graph, we show the magic as a function of N_T T gates inserted at random positions in a Clifford circuit. We show the unmitigated and mitigated magic from the IonQ quantum computer and an exact simulation of the quantum states. The mean value and the standard deviation of $\mathcal{B}_a$ are taken over six random instances of the states for $N = 3$ qubits. The dashed line is the additive Bell magic averaged over Haar-random states. The experiment is performed with $N_Q = 10^3$ measurement samples and no further error or readout error mitigation. With the purity measured from the experiment, Eq. (13) gives, on average, a depolarization error of $p \approx 0.1$ for the IonQ quantum computer.

determine the correct class of an unknown state $|\psi\rangle$ sampled from one of two classes α, β with different Bell magic $\mathcal{B}(|\psi\rangle \in \alpha) = \mathcal{B}^\alpha$ and $\mathcal{B}(|\psi\rangle \in \beta) = \mathcal{B}^\beta$, with $\mathcal{B}^\alpha > \mathcal{B}^\beta$. To this end, we perform N_Q repetitions of the Bell measurements and estimate $\hat{\mathcal{B}}$. We choose an appropriate threshold $\mathcal{B}^*$. For $\hat{\mathcal{B}} > \mathcal{B}^*$, we decide that the given state belongs to α, while for $\hat{\mathcal{B}} \leq \mathcal{B}^*$ we say that the state belongs to β. We define P_E as the probability of wrongly classifying a state in the state-discrimination protocol.

We now motivate the scaling of the number of measurements N_Q needed for the classification task with a misclassification probability P_E. If the estimation error $\Delta\mathcal{B}$ of $\mathcal{B}(|\psi\rangle)$ is larger than $\mathcal{B}^\alpha - \mathcal{B}^\beta$, the estimation error is too large to reliably distinguish the two classes. Thus, the estimation error must be smaller than the difference in magic of the two states $\Delta\mathcal{B} < \mathcal{B}^\alpha - \mathcal{B}^\beta$ to reliably distinguish the states. Equation Eq. (16) tells us how many measurements are needed to estimate magic with additive error $\Delta\mathcal{B}$. We argue that, in general, the classification task follows the same scaling as Eq. (16) in the number of samples N_Q:

$$N_Q \propto \frac{1}{(1-p)^{16}(\mathcal{B}^\alpha - \mathcal{B}^\beta)^2}. \tag{17}$$

An important special case is the discrimination of magic states $\mathcal{B}^\alpha > 0$ from stabilizer states with $\mathcal{B}^\beta = 0$. For this case, we can derive the precise number of measurements needed for a threshold $\mathcal{B}^* = 0$ and $p = 0$.

First, we study a state with low magic,

$$|\psi_C(\phi)\rangle = U_C|A_\phi\rangle \otimes |0\rangle^{N-1}, \tag{18}$$

consisting of an arbitrary Clifford circuit U_C and an initial state $|A_\phi\rangle \otimes |0\rangle^{N-1}$ with $N_A = 1$ nonstabilizer qubit $|A_\phi\rangle = \cos(\phi/2)|0\rangle + \sin(\phi/2)|1\rangle$. Here, ϕ controls the amount of magic introduced into the circuit, as seen in Eq. (6). In particular, for $\phi = n\pi/2$, n being an integer, no magic is introduced, whereas for $\phi = \pi/4$ we have $\mathcal{B}_a = 1$. For small ϕ, Bell magic can be approximated as $\mathcal{B}(|\phi| \ll 1) \approx 2\phi^2$. By tuning ϕ, we can create states containing arbitrarily low amounts of magic. We define the error probability $P_E(\phi)$ as the probability of wrongly classifying $|\psi_C(\phi)\rangle$ as a stabilizer state. With the assumption of large N_R, we find (as shown in Appendix N) that

$$P_E(\phi) = 4^{-N_Q}[(3 - \cos(2\phi))^{N_Q} + (3 + \cos(2\phi))^{N_Q}] - 2^{-N_Q}[\sin(\phi)^{2N_Q} + \cos(2\phi)^{2N_Q}]. \tag{19}$$

For $\phi = \pi/4$, we can approximate the error probability as

$$P_E(\frac{\pi}{4}) \approx 2\left(\frac{3}{4}\right)^{N_Q}. \tag{20}$$

We achieve an error $P_E(\pi/4) < 0.01$ when $N_Q > 18$. For near-stabilizer states with small $\phi \ll 1$, we find that

$$N_Q \approx -\frac{2\log(P_E(\phi \ll 1))}{\phi^2}, \tag{21}$$

showing an inverse-quadratic scaling law with ϕ. For example, a modest budget of $N_Q = 375$ samples is needed to classify $\phi = \pi/20$ with an error of $P_E^{\min} < 0.01$. Surprisingly, for small ϕ we find a scaling $N_Q \propto \mathcal{B}^{-1}$. This scaling is better than Eq. (17), which is derived with the additional assumption of a relatively small number of resampling steps $N_R = N_Q/4$.

As second case, we study the limit of highly magical states with $N_A \gg 1$. For these types of states, we can assume that the N_Q Bell measurements yield random bit strings. As we add two bit strings together in Algorithm 2, we have in total $N_Q - 1$ independent bit strings and their associated Pauli strings. The probability that two random Pauli strings commute is approximately 1/2. The misclassification probability is given by the probability that $N_Q - 1$ Pauli strings pairwise commute:

$$P_E(N_A \gg 1) \approx 2^{-(N_Q-1)(N_Q-2)/2}. \tag{22}$$

In particular, we achieve an error probability $P_E(N_A \gg 1) < 0.01$ when $N_Q > 5$.

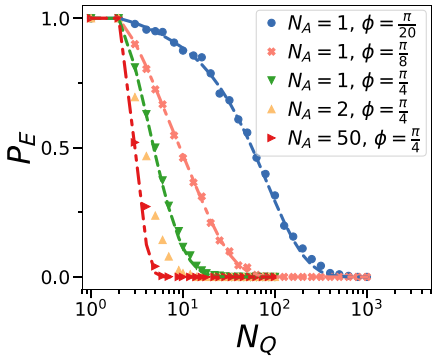

FIG. 4. The simulation of the error probability P_E for wrongly classifying a given magic state as a stabilizer state using N_Q measurements and threshold $\mathcal{B}^* = 0$. The measured states are randomly chosen Clifford circuits of $N = 50$ qubits applied to a product state of N_A magic states, where angle ϕ controls the amount of magic introduced, with $\phi = 0$ zero magic and $\phi = \frac{\pi}{4}$ maximal magic. The dashed lines are Eq. (19) for the $N_A = 1$ cases with $\phi = \frac{\pi}{20}$ (blue), $\phi = \frac{\pi}{8}$ (orange), and $\phi = \frac{\pi}{4}$ (green). The red dashed line Eq. (22) is the analytic formula for highly magical states.

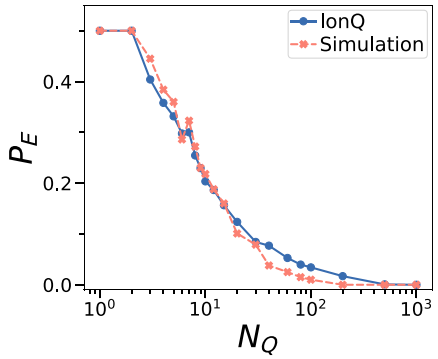

FIG. 5. An experiment to distinguish stabilizer and highly magical states with the IonQ quantum computer using Bell magic. We show classification error P_E as a function of the number of measurements N_Q. The supervised-learning algorithm to determine the best threshold $\mathcal{B}^*$ is shown in Appendix J. We measure 20 randomly chosen instances of stabilizer and magic states, respectively, which are generated with a hardware-efficient circuit of $d = 2$ depth and $N = 3$ qubits. The shown curves are the classification error for the test data set, which consists of 20% of the data and is not seen during training. The error is averaged over ten random distributions of test and training data. The blue dots are the experiment with the IonQ quantum computer, while the orange crosses are a noisy simulation using the average experimentally measured depolarization error $p \approx 0.15$.

Now, we numerically demonstrate our magic state discrimination protocol. In Fig. 4, we simulate the states $|\psi(\phi)\rangle = U_C|A_\phi\rangle^{\otimes N_A} \otimes |0\rangle^{\otimes(N-N_A)}$, where U_C is a randomly chosen Clifford circuit and N_A is the number of nonstabilizer input qubits. We measure the state N_Q times and estimate the magic $\hat{\mathcal{B}}$ with Algorithm 2. For $N_Q \leq 3$, we cannot draw the bit strings without replacement in Algorithm 2 and thus we draw with replacement instead. If we measure $\hat{\mathcal{B}} = \mathcal{B}^* = 0$, we incorrectly say that the state is a stabilizer state; otherwise, for $\hat{\mathcal{B}} > 0$ we classify the state as a nonstabilizer state. We find that our numerical results for the error probability P_E fit very well with our theoretical formulas.

Next, in Fig. 5 we use the IonQ quantum computer to experimentally distinguish stabilizer and nonstabilizer states. We measure stabilizer states and states generated by a hardware-efficient circuit with random parameters that are expected to have a lot of Bell magic (see Appendix H). The measured Bell magic of the states is shown in Appendix K. Due to noise, the prepared stabilizer states are mixed states and have nonzero Bell magic. Thus, we use the supervised-learning algorithm shown in Appendix J to learn the best decision boundary $\mathcal{B}^*$ from the experimental data. When, for a given measured state, $\hat{\mathcal{B}} < \mathcal{B}^*$, we say that the measured state is a stabilizer state; otherwise, we say that it is not a stabilizer state. In Fig. 5, we show the classification error as a function of the number of samples N_Q measured on the quantum computer. We find that the experimental results fit well with a simulation of the protocol with the measured depolarizing noise $p = 0.15$.

Note that with our machine-learning algorithm, the maximal error rate is $P_E^{\max} = 1/2$ due to the trivial strategy of classifying states at random.

VIII. VARIATIONAL BELL MAGIC SOLVER

Variational quantum algorithms find the parameters $\boldsymbol{\theta}$ of a parametrized quantum circuit $|\psi(\boldsymbol{\theta})\rangle$ such that they maximize a cost function $C(\boldsymbol{\theta})$ measured on a quantum computer [31,62,63]. The algorithm runs in a quantum classical feedback loop, where the cost function is measured on the quantum computer and used by a classical optimization routine to find improved parameters. We now propose a variational quantum algorithm to maximize Bell magic [see Fig. 1(d)]. Commonly, the cost function is maximized with gradient descent, where the kth parameter is iteratively updated with the gradient $\boldsymbol{\theta}'_k = \boldsymbol{\theta}_k - \partial_k C(\boldsymbol{\theta})$. The shift rule provides exact gradients when the circuit is composed of parametrized Pauli rotations [64]. For standard measurements on single quantum states, the shift rule is given by $\partial_k\langle C(\boldsymbol{\theta})\rangle = v(\langle C(\boldsymbol{\theta} + \mathbf{e}_k\frac{\pi}{4v})\rangle - \langle C(\boldsymbol{\theta} - \mathbf{e}_k\frac{\pi}{4v})\rangle)$, where $\mathbf{e}_k$ is the kth unit vector and $v > 0$ [65].

We extend the shift rule to Bell measurements (see Appendix L),

$$\partial_k P(\mathbf{r}) = 2v\langle\psi(\boldsymbol{\theta} + \frac{\pi}{4v}\mathbf{e}_k)|\langle\psi(\boldsymbol{\theta})|O_{\mathbf{r}}|\psi(\boldsymbol{\theta} + \frac{\pi}{4v}\mathbf{e}_k)\rangle|\psi(\boldsymbol{\theta})\rangle - 2v\langle\psi(\boldsymbol{\theta} - \frac{\pi}{4v}\mathbf{e}_k)|\langle\psi(\boldsymbol{\theta})|O_{\mathbf{r}}|\psi(\boldsymbol{\theta} - \frac{\pi}{4v}\mathbf{e}_k)\rangle|\psi(\boldsymbol{\theta})\rangle, \tag{23}$$

Input :
$j = 1, \ldots, 3N_{\mathrm{Q}}$ bitstrings $r^j \in \{0,1\}^{2N}$ sampled from Bell measurement on $|\psi(\boldsymbol{\theta})\rangle \otimes |\psi(\boldsymbol{\theta})\rangle$
$j = 1, \ldots, N_{\mathrm{Q}}$ bitstrings $q_+^j \in \{0,1\}^{2N}$ sampled from Bell measurement on $|\psi(\boldsymbol{\theta} + \frac{\pi}{2}\boldsymbol{e}_k)\rangle \otimes |\psi(\boldsymbol{\theta})\rangle$
$j = 1, \ldots, N_{\mathrm{Q}}$ bitstrings $q_-^j \in \{0,1\}^{2N}$ sampled from Bell measurement on $|\psi(\boldsymbol{\theta} - \frac{\pi}{2}\boldsymbol{e}_k)\rangle \otimes |\psi(\boldsymbol{\theta})\rangle$
Resampling steps N_{R}
Output: Gradient of Bell magic $\partial_k \mathcal{B}$

$\mathcal{B}_+ \leftarrow 0$
$\mathcal{B}_- \leftarrow 0$
for $k = 1, \ldots, N_R$ **do**
  Choose randomly without replacement $n_1, n_2, n_3 \in \{1, \ldots, 3N_{\mathrm{Q}}\}$
  $m \in \{1, \ldots, N_{\mathrm{Q}}\}$
  $\mathcal{B}_+ \leftarrow \mathcal{B}_+ + \text{Check-Commute}(r^{n_1} \oplus r^{n_2}, r^{n_3} \oplus q_+^m)$
  $\mathcal{B}_- \leftarrow \mathcal{B}_- + \text{Check-Commute}(r^{n_1} \oplus r^{n_2}, r^{n_3} \oplus q_-^m)$
end
$\mathcal{B}_+ \leftarrow 4\mathcal{B}_+/N_{\mathrm{R}}$
$\mathcal{B}_- \leftarrow 4\mathcal{B}_-/N_{\mathrm{R}}$
$\partial_k \mathcal{B} \leftarrow \mathcal{B}_+ - \mathcal{B}_-$

Algorithm 3. Gradient of Bell magic

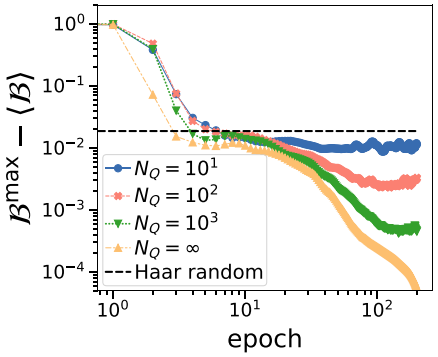

FIG. 6. The simulation of the variational algorithm to maximize Bell magic $\mathcal{B}$ of a parametrized quantum circuit for a varying number of measurement samples N_Q. We plot the difference between the average Bell magic $\langle B \rangle$ found at a training epoch and the maximal possible magic $B^{\max}$ of pure states. The training results are averaged over ten random training instances. The learning rate is $\gamma = 0.1$, the depth of circuit $d = 6$, and the number of qubits $N = 4$.

and the gradient of Bell magic is given by

$$\partial_k \mathcal{B} = 4 \sum_{\substack{\mathbf{r},\mathbf{r}',\mathbf{q},\mathbf{q}' \\ \in \{0,1\}^{2N}}} [\partial_k P(\mathbf{r})] P(\mathbf{r}') P(\mathbf{q}) P(\mathbf{q}') \left\| [\sigma_{\mathbf{r}\oplus\mathbf{r}'}, \sigma_{\mathbf{q}\oplus\mathbf{q}'}] \right\|_\infty . \tag{24}$$

Algorithm 3 depicts how to efficiently measure the gradient on a quantum computer for the case $v = 1/2$.

Conveniently, the Bell measurements also give us access to the diagonal entries $\mathcal{F}_{kk}(\boldsymbol{\theta})$ of the quantum Fisher information metric $\mathcal{F}_{ij}(\boldsymbol{\theta}) = 4[\langle \partial_i \psi | \partial_j \psi \rangle - \langle \partial_i \psi | \psi \rangle \langle \psi | \partial_j \psi \rangle]$ without requiring additional measurements. The quantum Fisher information metric and its diagonal approximation can tremendously speed up the training of variational quantum algorithms with the quantum natural gradient $\mathcal{F}^{-1}(\boldsymbol{\theta}) \nabla C(\boldsymbol{\theta})$ [66–69]. With the shift rule, the diagonal entries of the metric are given by $\mathcal{F}_{kk}(\boldsymbol{\theta}) = 2(1 - |\langle \psi(\boldsymbol{\theta}) | \psi(\boldsymbol{\theta} + \mathbf{e}_k \frac{\pi}{2}) \rangle|^2)$ [70]. For pure states, the fidelity $|\langle \psi(\boldsymbol{\theta}) | \psi(\boldsymbol{\theta} + \mathbf{e}_k \frac{\pi}{2}) \rangle|^2$ is given by the SWAP test in Eq. (1), where we can simply reuse the measurement outcomes for the gradient of magic.

In Fig. 6, we numerically study the variational Bell magic solver to find the pure state with maximal Bell magic $\mathcal{B}^{\max}$. We show the variational magic solver to maximize $\mathcal{B}$ as a function of the training epochs for a varying number of measurement samples N_Q. We use the gradient method *Adam* [71] and the gradients are determined using the shift rule, where we use N_Q measurement samples for each measurement setting. To estimate the gradients, we require in total $N_Q(2K + 3)$ measurement samples, where K is the number of parameters of the circuit. The parametrized quantum circuit is shown in Appendix H and the initial parameters are chosen such that the initial state is close to a stabilizer state. Training is initially fast until reaching the average magic of Haar-random quantum states. Then, optimization continues at a slower pace. With increasing N_Q, our solver finds states that have close to the maximal amount of Bell magic. In general, we find that pure states of high Bell magic are characterized by having a small, but nonzero, expectation value for nearly all Pauli operators. We further study the structure of pure states with high Bell magic in Appendix G.

The performance of variational quantum algorithms is tied to their expressibility and trainability [31]. Expressibility describes how well an ansatz uniformly explores the full Hilbert space, which makes it more likely that the ansatz can express the target solution [72]. A circuit is maximally expressible if it forms a 2-design, i.e., averaging over the ansatz matches an average over Haar-random unitaries up to the second moment. A variational quantum algorithm is trainable when the magnitude of gradients is large. A common issue is the so-called barren-plateau problem, where the magnitude of the gradients vanishes exponentially with the number of qubits [73]. Nearly all variational quantum algorithms use the observable $\mathcal{H} = \sum_{n=1}^{\mathrm{poly}(N)} \gamma_n \sigma_n$, which consists of a sum over a polynomial number of arbitrary Pauli strings σ_n with constant coefficients γ_n. For this general class of cost function, high expressibility leads directly to vanishing gradients and the training becomes impractical [73,74].

Bell magic does not belong to the aforementioned class of cost functions, as it cannot be expressed by a polynomial sum of Pauli strings and it requires two copies of a quantum state to be measured. We illustrate

this fact for a highly expressible ansatz where barren plateaus are absent for Bell magic, while generally used cost functions suffer from barren plateaus. As a simple demonstration, we define the ansatz $|\psi(\theta, U_C)\rangle = U_C \exp(-i(1)/2\theta\sigma_1^y)|0\rangle^{\otimes N}$, where U_C is a Clifford circuit, θ is the parameter of the circuit, and σ_1^y is the y Pauli operator acting on the first qubit. Over randomly sampled Clifford circuits U_C from the Clifford group $\mathcal{C}$, this ansatz is maximally expressible, as it uniformly explores the full Hilbert space and forms a 2-design [75]. Thus, for any cost function $\mathcal{H}$ consisting of a polynomial number of Pauli strings, the magnitude of the gradient decays exponentially with the number N of qubits [73]:

$$\mathrm{Var}[\partial_\theta \langle\psi(\theta, U_C)|\mathcal{H}|\psi(\theta, U_C)\rangle)]_{\theta,\mathcal{C}} = \frac{\mathrm{Tr}(\mathcal{H}^2)}{2(2^{2N}-1)} \propto 2^{-N}, \tag{25}$$

where the variance is taken over θ and the Clifford group $\mathcal{C}$. In contrast, the variance of the gradient in respect to Bell magic is independent of N. As shown in Appendix M, we find that

$$\mathrm{Var}[\partial_\theta \mathcal{B}(|\psi(\theta, U_C)\rangle)]_{\theta,\mathcal{C}} = \frac{1}{2}, \tag{26}$$

ensuring the trainability of this ansatz for any N.

IX. DISCUSSION AND CONCLUSIONS

We show how to measure and learn Bell magic with quantum computers. Our algorithm relies on Bell measurements of two copies of a quantum state, which is straightforward to implement on quantum computers and simulators, as demonstrated in past experiments to measure entanglement [45,48,76]. Bell magic can be measured concurrently with entanglement, which we verify in Appendix P by measuring the Wallach-Meyer entanglement measure. For quantum computers, the Bell transformation can be implemented directly without SWAP gates when all qubits are connected to all other qubits, such as on the IonQ quantum computer [58], or when the two copies of the quantum state can be arranged in a ladder structure [45]. Further, it can be easily implemented in numerical simulations to study the magic of many-body quantum systems [39,40]. In contrast to existing measures of magic, the measurement cost N_Q is independent of the qubit number and has only an inverse-quadratic scaling $N_Q \propto \Delta\mathcal{B}^{-2}$ with the estimation error $\Delta\mathcal{B}$. Noise occurring in experiments can be mitigated with a scaling of $N_Q \propto (1-p)^{-16}$. We derive these bounds assuming a small number of resampling steps $N_R = N_Q/4$ in the classical part of the algorithm, where a better scaling is possible by increasing N_R.

For our study, we assume depolarizing noise, which we find is sufficient to mitigate noise on the IonQ quantum computer. The error-mitigated Bell magic of our experiments closely matches exact simulations, opening up the study of nonstabilizerness on noisy quantum computers [31]. Additional methods that turn nondepolarizing noise into depolarizing noise could be used to further improve error mitigation [53,54].

Fault-tolerant quantum computers are commonly run by state-synthesis protocols, where a resource state $|\phi\rangle_{\mathrm{ini}}$ combined with error-corrected Clifford operations is transformed into a target state $|\phi\rangle_{\mathrm{target}}$ [14]. Magic quantifies a lower bound on the nonstabilizer resources needed to synthesize a particular state or unitary [5,20,25]. As Bell magic is invariant under Clifford unitaries, a necessary condition for state synthesis is that $\mathcal{B}(|\phi\rangle_{\mathrm{target}}) \leq \mathcal{B}(|\phi\rangle_{\mathrm{ini}})$. Thus, Bell magic can experimentally establish lower bounds on the magical resources needed to synthesize states on quantum computers.

Another important task is to verify whether a given state is indeed correct [36,37]. Given a stabilizer state, learning the description of the state requires measurements on $O(N)$ copies [49]. In contrast, discriminating whether a given state is a stabilizer state is comparatively easier, requiring only $O(1)$ copies [77]. Our work allows us to distinguish states with different degrees of nonstabilizerness in the presence of noise. For example, with only $N_Q = 6$ measurements, we can decide with an error of less than 1% whether a given state is a stabilizer state or a highly magical state. To distinguish stabilizer states and near-stabilizer states as defined in Eq. (18), we find a scaling of $N_Q \propto \phi^{-2}$. Our method could be used to reliably certify states for quantum communication [78] and quantum computing tasks [79]. Note that the related task of testing whether a given unitary is Clifford has been studied in Refs. [77,80].

We also find that Bell magic is connected to the recently proposed linear stabilizer entropy $M_{\mathrm{lin}} = 1 - 2^N \sum_{\mathbf{r}} (2^{-N}\langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2)^2$ and the 2-Rényi entropy $M_2 = -\log(2^N \sum_{\mathbf{r}} (2^{-N}\langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2)^2)$ [10]. We find that these measures of magic can be also computed with Bell measurements via $M_{\mathrm{lin}} = 1 - 2^N \sum_{\mathbf{r}} P(\mathbf{r})^2$ and $M_2 = -\log(2^N \sum_{\mathbf{r}} P(\mathbf{r})^2)$ (see Appendix O). Here, one has to explicitly estimate the probabilities $P(\mathbf{r})$, which in general requires an exponential amount of measurements. We experimentally compute the stabilizer entropy with the IonQ quantum computer and compare it with Bell magic in the Appendix O, where we find that both measures behave similarly.

Our results reveal a fundamental relationship between magic and the probability distribution of bit strings in the Bell basis. Statistical tests over these distributions could serve as benchmarks of classical simulation complexity, similar to cross-entropy benchmarking in the computational basis [38].

Bell magic is a faithful measure of magic for pure states, while mixed Bell magic is also faithful for a class of mixed

states [see Eq. (8)]. We prove the invariance under Clifford unitaries and composition and our numerics suggest that Bell magic is nonincreasing under partial trace and measurements in the computational basis. We leave the formal proof of whether mixed Bell magic fulfills these conditions as an open problem [22]. Using convex-roof-type extensions, it may also be possible to construct a faithful Bell magic for arbitrary mixed states [81]. As an extension to unitaries, channel capacity and qudits would be interesting as well [22,24]. It would be also useful to find a quantitative connection between Bell magic and classical simulation complexity [4].

Finally, we show that our variational algorithm for Bell magic can have large gradients even for highly expressible circuits that uniformly sample the full Hilbert space. For commonly used cost functions, expressible circuits must have barren plateaus [73,74,82,83]. However, this rule does not apply to Bell magic, as it cannot be expressed by a polynomial number of Pauli strings. As an example, we demonstrate an ansatz that has a high expressibility and yet the gradient is independent of N. Unbounded cost functions are known to have similar features, although they are not suited for near-term quantum computers [84]. We note that barren plateaus can still appear depending on the choice of ansatz. It would be interesting to search for other cost functions that combine expressibility and large gradients by using entangled measurements over multiple copies [44]. One could also study the trainability of Bell magic in conjunction with the learnability transition that occurs in Clifford circuits combined with T gates [85,86].

The code for this paper is available via Ref. [87].

ACKNOWLEDGMENTS

We thank Kishor Bharti, Ronan Docherty, Nikolaos Koukoulekidis, Ashley Montanaro, Felix Roberts, David Steuerman, and Mark Wilde for insightful comments. We thank IonQ for providing quantum computing resources. This work is supported by a Samsung GRC project and the United Kingdom Hub in Quantum Computing and Simulation, part of the United Kingdom National Quantum Technologies Programme, with funding from the UK Research and Innovation (UKRI) Engineering and Physical Sciences Research Council (EPSRC) under Grant No. EP/T001062/1.

Note added.—Recently, the stabilizer entropy has been experimentally measured with a randomized measurement protocol that scales exponentially with the number of qubits [88].

APPENDIX A: STABILIZER STATES AND BELL MEASUREMENT

Here, we show the connection between Bell measurement and stabilizer states. The transformation into the Bell basis is applied with the unitary $U_{\text{Bell}} = \bigotimes_{n=1}^{N}(H \otimes I_2) \times$ cnot on all N pairs of qubits [see Fig. 1(b)]. The resulting state is sampled N_Q times and we record the jth measurement outcome as the bit string $\mathbf{r}^j \in \{0,1\}^{2N}$. The outcome $\mathbf{r}$ appears with a probability [49],

$$P(\mathbf{r}) = \langle\psi|\langle\psi|O_{\mathbf{r}}|\psi\rangle|\psi\rangle = 2^{-N}|\langle\psi|\sigma_{\mathbf{r}}|\psi^*\rangle|^2\,, \tag{A1}$$

where $O_{\mathbf{r}} = |\sigma_{\mathbf{r}}\rangle\langle\sigma_{\mathbf{r}}|$ is the projector onto the Bell state and $|\psi^*\rangle$ denotes the complex conjugate of $|\psi\rangle$. For any state, there are at least 2^N possible outcomes $\mathbf{r}$ as $|\langle\psi|\sigma_{\mathbf{r}}|\psi^*\rangle|^2 \leq 1$.

The stabilizer states $|\psi_{\text{STAB}}\rangle$ are defined by a commuting subgroup G of $|G| = 2^N$ Pauli strings σ. We have $\langle\psi_{\text{STAB}}|\sigma|\psi_{\text{STAB}}\rangle = \pm 1$ for $\sigma \in G$ and $\langle\psi_{\text{STAB}}|\sigma'|\psi_{\text{STAB}}\rangle = 0$ for $\sigma' \notin G$. Any $\sigma_{\mathbf{r}}, \sigma_{\mathbf{r}'} \in G$ commute $[\sigma_{\mathbf{r}}, \sigma_{\mathbf{r}'}] = 0$. Any stabilizer state can be written as [89]

$$|\psi_{\text{STAB}}\rangle = \frac{1}{\sqrt{|A|}}\sum_{x\in A} i^{\ell(x)}(-1)^{q(x)}|x\rangle\,, \tag{A2}$$

where A is an affine subspace of the Galois field $\mathbb{F}_2^N$ and $\ell, z : \{0,1\}^N \to \{0,1\}$ are linear and quadratic polynomials over $\mathbb{F}_2$. As ℓ is linear, we have $\ell(x) = sx$ for $s \in \{0,1\}^N$ and the complex factor of the stabilizer state can be written as $i^{\ell(x)} = \prod_{k\in S} i^{x_k}$ for some $S \subseteq [N]$. Thus, we can write the complex conjugate as a transformation with the z Pauli operator $|\psi^*\rangle = \sigma_{10}^{\otimes S}|\psi\rangle = \sigma_{\mathbf{g}}|\psi\rangle$, with some $\mathbf{g} = \{s_1, 0, s_2, 0, \ldots, s_N, 0\}$ that characterizes the complex phase of the stabilizer state [49]. Inserting this relation into Eq. (A1), the probability of sampling a bit string $\mathbf{q}$ from a stabilizer state is given as $P(\mathbf{q}) = 2^{-N}|\langle\psi_{\text{STAB}}|\sigma_{\mathbf{q}}\sigma_{\mathbf{g}}|\psi_{\text{STAB}}\rangle|^2 = 2^{-N}|\langle\psi_{\text{STAB}}|\sigma_{\mathbf{q}\oplus\mathbf{g}}|\psi_{\text{STAB}}\rangle|^2$. There are 2^N outcomes with nonzero probability $P(\mathbf{q}) > 0$ with $\sigma_{\mathbf{t}} = \sigma_{\mathbf{q}\oplus\mathbf{g}} \in G$. Any outcome $\mathbf{q}$ can be written as $\mathbf{q} = \mathbf{t} \oplus \mathbf{g}$, where the set of strings $\mathbf{t} \in \{0,1\}^{2N}$ with $\sigma_{\mathbf{t}} \in G$ forms an N-dimensional linear subspace of $\mathbb{F}_2^{2N}$. The addition of two elements of the subspace $\mathbf{t} \oplus \mathbf{t}'$ is again part of the commuting subspace $\sigma_{\mathbf{t}\oplus\mathbf{t}'} \in G$. The addition of two measured outcomes $\mathbf{q}, \mathbf{q}'$ yields $\mathbf{q} \oplus \mathbf{q}' = \mathbf{t} \oplus \mathbf{g} \oplus \mathbf{t}' \oplus \mathbf{g} = \mathbf{t} \oplus \mathbf{t}'$. Thus, the addition of two outcomes from a stabilizer state yields a Pauli string of the commuting subgroup $\sigma_{\mathbf{q}\oplus\mathbf{q}'} \in G$.

In summary, for any set of bit strings $\{\mathbf{q}_n\}_{n=1}^{N_Q}$ sampled in the Bell basis from a pure stabilizer state, the Pauli strings of its binary additions must commute, i.e., $[\sigma_{\mathbf{q}_k\oplus\mathbf{q}_l}, \sigma_{\mathbf{q}_n\oplus\mathbf{q}_m}] = 0\ \forall k, l, n, m$. Conversely, finding at least one noncommuting Pauli string implies that the measured quantum state is not a pure stabilizer state, as the commuting subgroup G contains at most 2^N elements.

APPENDIX B: INVARIANCE UNDER CLIFFORD CIRCUITS

We now show that Bell magic is invariant $\mathcal{B}(|\psi\rangle) = \mathcal{B}(U_C|\psi\rangle)$ under arbitrary Clifford circuits U_C, i.e., unitaries that map stabilizer states into stabilizer states.

To this end, we start by proving that the second and third lines of Eq. (B1) are indeed equal:

$$\begin{aligned} Q(\mathbf{n}) &= \sum_{\mathbf{r}\in\{0,1\}^{2N}} P(\mathbf{r})P(\mathbf{r}\oplus\mathbf{n}) \\ &\equiv \sum_{\mathbf{r}} \langle\psi|\langle\psi|O_{\mathbf{r}}|\psi\rangle|\psi\rangle\langle\psi|\langle\psi|O_{\mathbf{r}\oplus\mathbf{n}}|\psi\rangle|\psi\rangle \\ &= 4^{-N}\sum_{\mathbf{r}} \langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2\langle\psi|\sigma_{\mathbf{r}\oplus\mathbf{n}}|\psi\rangle^2, \end{aligned} \tag{B1}$$

where we have an arbitrary $\mathbf{n}\in\{0,1\}^{2N}$. We note that by using four copies of $|\psi\rangle$, we can rewrite the bottom line of Eq. (B1) as

$$\sum_{\mathbf{r}} \langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2\langle\psi|\sigma_{\mathbf{r}\oplus\mathbf{n}}|\psi\rangle^2 = \langle\psi|^{\otimes 4}\sum_{\mathbf{r}}\sigma_{\mathbf{r}}^{\otimes 2}\otimes\sigma_{\mathbf{r}\oplus\mathbf{n}}^{\otimes 2}|\psi\rangle^{\otimes 4}. \tag{B2}$$

Similarly, the second line of Eq. (B1) can be rewritten as

$$\begin{aligned} &\sum_{\mathbf{r}} \langle\psi|\langle\psi|O_{\mathbf{r}}|\psi\rangle|\psi\rangle\langle\psi|\langle\psi|O_{\mathbf{r}\oplus\mathbf{n}}|\psi\rangle|\psi\rangle \\ &= \langle\psi|^{\otimes 4}\sum_{\mathbf{r}} O_{\mathbf{r}}\otimes O_{\mathbf{r}\oplus\mathbf{n}}|\psi\rangle^{\otimes 4}. \end{aligned} \tag{B3}$$

Now, we note that both the Bell operator $O_{\mathbf{r}}$ and the Pauli string $\sigma_{\mathbf{r}}$ can be written as tensor products acting on two and one qubits, respectively. In particular,

$$O_{\mathbf{r}} = |\sigma_{r_1r_2}\rangle\langle\sigma_{r_1r_2}|\otimes\cdots\otimes|\sigma_{r_{2N-1}r_{2N}}\rangle\langle\sigma_{r_{2N-1}r_{2N}}| \tag{B4}$$

and

$$\sigma_{\mathbf{r}} = \sigma_{r_1r_2}\otimes\cdots\otimes\sigma_{r_{2N-1}r_{2N}}. \tag{B5}$$

At the same time, the sum can be decomposed as $\sum_{\mathbf{r}} = \sum_{r_1}\sum_{r_2}\cdots\sum_{r_{2N}}$.

The operators can be written as tensor products of N operators acting on four-qubit subspaces:

$$\begin{aligned} \sum_{\mathbf{r}}\sigma_{\mathbf{r}}^{\otimes 2}\otimes\sigma_{\mathbf{r}\oplus\mathbf{n}}^{\otimes 2} &= \sum_{r_1,r_2}\sigma_{r_1r_2}^{\otimes 2}\otimes\sigma_{r_1\oplus n_1 r_2\oplus n_2}^{\otimes 2}\otimes\cdots\otimes \\ &\sum_{r_{2N-1}r_{2N}}\sigma_{r_{2N-1}r_{2N}}^{\otimes 2}\otimes\sigma_{r_{2N-1}\oplus n_{2N-1}r_{2N}\oplus n_{2N}}^{\otimes 2} \end{aligned}$$

and

$$\begin{aligned} &\sum_{\mathbf{r}} O_{\mathbf{r}}\otimes O_{\mathbf{r}\oplus\mathbf{n}} \\ &= \sum_{r_1,r_2} |\sigma_{r_1r_2}\rangle\langle\sigma_{r_1r_2}|\otimes|\sigma_{r_1\oplus n_1 r_2\oplus n_2}\rangle\langle\sigma_{r_1\oplus n_1 r_2\oplus n_2}|\otimes\cdots\otimes \\ &\sum_{r_{2N-1}r_{2N}} |\sigma_{r_{2N-1}r_{2N}}\rangle\langle\sigma_{r_{2N-1}r_{2N}}|\otimes \\ &|\sigma_{r_{2N-1}\oplus n_{2N-1}r_{2N}\oplus n_{2N}}\rangle\langle\sigma_{r_{2N-1}\oplus n_{2N-1}r_{2N}\oplus n_{2N}}|. \end{aligned}$$

Now, one can explicitly calculate the operators on the four-qubit Hilbert space by hand and we find for all $n_1, n_2 \in \{0,1\}$ the following equivalence:

$$\begin{aligned} &\sum_{r_1,r_2} |\sigma_{r_1r_2}\rangle\langle\sigma_{r_1r_2}|\otimes|\sigma_{r_1\oplus n_1 r_2\oplus n_2}\rangle\langle\sigma_{r_1\oplus n_1 r_2\oplus n_2}| \\ &\quad = \frac{1}{4}\sum_{r_1,r_2}\sigma_{r_1r_2}^{\otimes 2}\otimes\sigma_{r_1\oplus n_1 r_2\oplus n_2}^{\otimes 2}. \end{aligned}$$

This equivalence also holds for any tensor product of the four-qubit operator. Thus we find, for any $\mathbf{n}\in\{0,1\}^{2N}$,

$$\sum_{\mathbf{r}} O_{\mathbf{r}}\otimes O_{\mathbf{r}\oplus\mathbf{n}} = \frac{1}{4^N}\sum_{\mathbf{r}}\sigma_{\mathbf{r}}^{\otimes 2}\otimes\sigma_{\mathbf{r}\oplus\mathbf{n}}^{\otimes 2}. \tag{B6}$$

Together with Eqs. Eq. (B2) and Eq. (B3), this proves that Eq. (B1) is indeed correct.

As a corollary, combining Eqs. Eq. (B1) and Eq. (A1) yields the surprising relation

$$\begin{aligned} &\sum_{\mathbf{r}\in\{0,1\}^{2N}} \langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2\langle\psi|\sigma_{\mathbf{r}\oplus\mathbf{n}}|\psi\rangle^2 \\ &\quad = \sum_{\mathbf{r}\in\{0,1\}^{2N}} |\langle\psi|\sigma_{\mathbf{r}}|\psi^*\rangle|^2|\langle\psi|\sigma_{\mathbf{r}\oplus\mathbf{n}}|\psi^*\rangle|^2. \end{aligned} \tag{B7}$$

Next, we proceed to prove the invariance of Bell magic. As reminder, Bell magic is defined as

$$\mathcal{B} = \sum_{\substack{\mathbf{r},\mathbf{r}',\mathbf{q},\mathbf{q}'\\ \in\{0,1\}^{2N}}} P(\mathbf{r})P(\mathbf{r}')P(\mathbf{q})P(\mathbf{q}')\left\|[\sigma_{\mathbf{r}\oplus\mathbf{r}'},\sigma_{\mathbf{q}\oplus\mathbf{q}'}]\right\|_\infty. \tag{B8}$$

Also, recall that $\sigma_{\mathbf{r}\oplus\mathbf{r}'} = \sigma_{\mathbf{r}}\sigma_{\mathbf{r}'}$ up to a prefactor $\{1,-1,i,-i\}$. Now, we equivalently rewrite Bell magic into

$$\mathcal{B} = \sum_{\mathbf{n},\mathbf{q}\in\{0,1\}^{2N}} Q(\mathbf{n})Q(\mathbf{q})\left\|[\sigma_{\mathbf{n}},\sigma_{\mathbf{q}}]\right\|_\infty. \tag{B9}$$

where we define $Q(\mathbf{n}) = \sum_{\mathbf{r}} P(\mathbf{r})P(\mathbf{r}\oplus\mathbf{n})$, with $P(\mathbf{r})$ as defined in Eq. (A1). Now, we transform $|\psi\rangle$ with a random Clifford circuit U_C into $U_C|\psi\rangle$. Note that U_C transforms a

Pauli string $\sigma_{\mathbf{n}}$ into another Pauli string $\sigma_{\mathbf{q}}$ with $U_C\sigma_{\mathbf{n}}U_C^\dagger = \sigma_{\mathbf{q}}$. This transformation is bijective, i.e., each Pauli string is mapped to another unique Pauli string. Using Eq. (B1), we have $Q(\mathbf{n}) = 4^{-N}\sum_{\mathbf{r}}\langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2\langle\psi|\sigma_{\mathbf{r}}\sigma_{\mathbf{n}}|\psi\rangle^2$. The transformed probability $Q'(\mathbf{n})$ is given by

$$\begin{aligned}Q'(\mathbf{n}) &= 4^{-N}\sum_{\mathbf{r}}\langle\psi|U_C^\dagger\sigma_{\mathbf{r}}U_C|\psi\rangle^2\langle\psi|U_C^\dagger\sigma_{\mathbf{r}}\sigma_{\mathbf{n}}U_C|\psi\rangle^2\\ &= 4^{-N}\sum_{\mathbf{r}}\langle\psi|U_C^\dagger\sigma_{\mathbf{r}}U_C|\psi\rangle^2\langle\psi|U_C^\dagger\sigma_{\mathbf{r}}U_CU_C^\dagger\sigma_{\mathbf{n}}U_C|\psi\rangle^2\\ &= 4^{-N}\sum_{\mathbf{r}}\langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2\langle\psi|\sigma_{\mathbf{r}}\sigma_{\mathbf{m}}|\psi\rangle^2 \equiv Q(\mathbf{m}),\end{aligned}$$

where we use that the sum over all Pauli strings remains invariant due to the bijective property and we define $\mathbf{m}$ as the transformed Pauli string $\sigma_{\mathbf{m}} = U_C^\dagger\sigma_{\mathbf{n}}U_C$. This means that a transformation with U_C simply permutes the distribution $Q(\mathbf{n})$. The Bell magic $\mathcal{B}'$ after transformation is given by

$$\begin{aligned}\mathcal{B}' &= \sum_{\mathbf{n},\mathbf{q}\in\{0,1\}^{2N}} Q'(\mathbf{n})Q'(\mathbf{q})\left\|[\sigma_{\mathbf{n}},\sigma_{\mathbf{q}}]\right\|_\infty\\ &= \sum_{\mathbf{n},\mathbf{q}\in\{0,1\}^{2N}} Q(\mathbf{n})Q(\mathbf{q})\left\|[U_C\sigma_{\mathbf{n}}U_C^\dagger, U_C\sigma_{\mathbf{q}}U_C^\dagger]\right\|_\infty\\ &= \sum_{\mathbf{n},\mathbf{q}\in\{0,1\}^{2N}} Q(\mathbf{n})Q(\mathbf{q})\left\|U_C[\sigma_{\mathbf{n}},\sigma_{\mathbf{q}}]U_C^\dagger\right\|_\infty\\ &= \sum_{\mathbf{n},\mathbf{q}\in\{0,1\}^{2N}} Q(\mathbf{n})Q(\mathbf{q})\left\|[\sigma_{\mathbf{n}},\sigma_{\mathbf{q}}]\right\|_\infty \equiv \mathcal{B},\end{aligned}$$

where in the final step we use that the commutator of two Pauli strings is either 0 or another Pauli string and therefore the norm is left invariant under transformation with U_C.

APPENDIX C: ADDITIVE BELL MAGIC

We now show that the additive Bell magic $\mathcal{B}_a = -\log_2(1-\mathcal{B})$ is additive. We consider a product $|\psi\rangle = |A\rangle\otimes|B\rangle$ of two arbitrary states $|A\rangle$ and $|B\rangle$. We now want to show that $\mathcal{B}_a(|A\rangle\otimes|B\rangle) = \mathcal{B}_a(|A\rangle) + \mathcal{B}_a(|B\rangle)$.

First, we set out some preliminary considerations.

As the Bell transformation is a tensor product, the outcomes appearing on the qubits of $|A\rangle$ and $|B\rangle$ are independent of each other. We define the outcomes for the qubits of $|A\rangle$ as $\mathbf{r}_A$ and for $|B\rangle$ as $\mathbf{r}_B$.

The product of two Pauli strings $\sigma_{\mathbf{r}}$ and $\sigma_{\mathbf{q}}$ is given by $\sigma_{\mathbf{r}}\sigma_{\mathbf{q}} = \pm i^{C_{\mathbf{r},\mathbf{q}}}\sigma_{\mathbf{r}\oplus\mathbf{q}}$, where $C_{\mathbf{r},\mathbf{q}} = 0$ when $[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}] = 0$ and $C_{\mathbf{r},\mathbf{q}} = 1$ otherwise. We can use this to write the commutator for the tensor product as

$$\begin{aligned}[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}] &= [\sigma_{\mathbf{r}_A}\otimes\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_A}\otimes\sigma_{\mathbf{q}_B}]\\ &= (\sigma_{\mathbf{r}_A}\sigma_{\mathbf{q}_A})\otimes(\sigma_{\mathbf{r}_B}\sigma_{\mathbf{q}_B}) - (\sigma_{\mathbf{q}_A}\sigma_{\mathbf{r}_A})\otimes(\sigma_{\mathbf{q}_B}\sigma_{\mathbf{r}_B})\\ &= \pm\sigma_{\mathbf{r}_A\oplus\mathbf{q}_A}\otimes\sigma_{\mathbf{r}_B\oplus\mathbf{q}_B}((-1)^{C_{\mathbf{r}_A,\mathbf{q}_A}+C_{\mathbf{r}_B,\mathbf{q}_B}}\\ &\quad - (-1)^{C_{\mathbf{r}_A,\mathbf{q}_A}+C_{\mathbf{r}_B,\mathbf{q}_B}}).\end{aligned}$$

This implies that the commutator is nonzero only when $C_{\mathbf{r}_A,\mathbf{q}_A} + C_{\mathbf{r}_B,\mathbf{q}_B} = 1$, i.e., when we have $[\sigma_{\mathbf{r}_A},\sigma_{\mathbf{q}_A}] \neq 0$ and $[\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_B}] = 0$, or $[\sigma_{\mathbf{r}_A},\sigma_{\mathbf{q}_A}] = 0$ and $[\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_B}] \neq 0$. This is the case when the Pauli strings of A commute but not those of B, as well as the reverse case. With this result, we can now write the infinity norm of the commutator as $\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty = \left\|[\sigma_{\mathbf{r}_A},\sigma_{\mathbf{q}_A}]\right\|_\infty + \left\|[\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_B}]\right\|_\infty - \left\|[\sigma_{\mathbf{r}_A},\sigma_{\mathbf{q}_A}]\right\|_\infty\left\|[\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_B}]\right\|_\infty$. This expression is zero only when both A and B do not commute or both commute. Recall that $\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty = 2$ when $\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}$ do not commute and zero otherwise.

To simplify the notation for the Bell magic, we define $Q(\mathbf{r}) = \sum_{\mathbf{q}}P(\mathbf{q})P(\mathbf{q}\oplus\mathbf{r})$. Note that $\sigma_{\mathbf{q}}\sigma_{\mathbf{r}} = \sigma_{\mathbf{q}\oplus\mathbf{r}}$ up to a multiplication with $\{1,-1,i,-i\}$, where we can ignore this factor for the calculation of the Bell magic since we take the norm of the commutator. With the above fact, we can write the Bell magic as

$$\mathcal{B} = \sum_{\mathbf{r},\mathbf{q}\in\{0,1\}^{2N}} Q(\mathbf{r})Q(\mathbf{q})\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty . \tag{C1}$$

As the Bell transformation is a tensor product and the underlying state is a product state, the probability for the outcomes are independent and we can write

$$Q(\mathbf{r}) = Q_A(\mathbf{r}_A)Q_B(\mathbf{r}_B). \tag{C2}$$

We note that $\sum_{\mathbf{r}_A,\mathbf{q}_A}Q_A(\mathbf{r}_A)Q_A(\mathbf{q}_A) = 1$ and $\sum_{\mathbf{r}_B,\mathbf{q}_B}Q_B(\mathbf{r}_B)Q_B(\mathbf{q}_B) = 1$.

We now combine our considerations to calculate the magic of the product state. We find that

$$\begin{aligned}&\mathcal{B}(|A\rangle\otimes|B\rangle)\\ &\quad= \sum_{\mathbf{r}_A,\mathbf{q}_A}\sum_{\mathbf{r}_B,\mathbf{q}_B} Q_A(\mathbf{r}_A)Q_A(\mathbf{q}_A)Q_B(\mathbf{r}_B)Q_B(\mathbf{q}_B))\\ &\quad\left\|[\sigma_{\mathbf{r}_A}\otimes\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_A}\otimes\sigma_{\mathbf{q}_B}]\right\|_\infty\\ &\quad= \sum_{\mathbf{r}_A,\mathbf{q}_A} Q_A(\mathbf{r}_A)Q_A(\mathbf{q}_A)\left\|[\sigma_{\mathbf{r}_A},\sigma_{\mathbf{q}_A}]\right\|_\infty\end{aligned}$$

$$+\sum_{\mathbf{r}_B,\mathbf{q}_B} Q_B(\mathbf{r}_B)Q_B(\mathbf{q}_B)\left\|[\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_B}]\right\|_\infty - \left(\sum_{\mathbf{r}_A,\mathbf{q}_A} Q_A(\mathbf{r}_A)Q_A(\mathbf{q}_A)\left\|[\sigma_{\mathbf{r}_A},\sigma_{\mathbf{q}_A}]\right\|_\infty\right)\cdot \times\left(\sum_{\mathbf{r}_B,\mathbf{q}_B} Q_B(\mathbf{r}_B)Q_B(\mathbf{q}_B)\left\|[\sigma_{\mathbf{r}_B},\sigma_{\mathbf{q}_B}]\right\|_\infty\right) = \mathcal{B}(|A\rangle)+\mathcal{B}(|B\rangle)-\mathcal{B}(|A\rangle)\mathcal{B}(|B\rangle).$$

We now find, for the additive magic,

$$\mathcal{B}_a(|A\rangle\otimes|B\rangle) = -\log_2[1-\mathcal{B}(|A\rangle)-\mathcal{B}(|B\rangle) + \mathcal{B}(|A\rangle)\mathcal{B}(|B\rangle)]. \tag{C3}$$

Finally, the additive magic of the individual states $|A\rangle$ and $|B\rangle$ is given by

$$\begin{aligned}\mathcal{B}_a(|A\rangle)+\mathcal{B}_a(|B\rangle) &= -\log_2[(1-\mathcal{B}(|A\rangle))(1-\mathcal{B}(|B\rangle))]\\ &= -\log_2[1-\mathcal{B}(|A\rangle)-\mathcal{B}(|B\rangle)+\mathcal{B}(|A\rangle)\mathcal{B}(|B\rangle)]\\ &\equiv \mathcal{B}_a(|A\rangle\otimes|B\rangle),\end{aligned}$$

which concludes our proof.

APPENDIX D: COMPOSITION OF BELL MAGIC AND STABILIZER STATES

We now show that Bell magic is invariant under composition of a state $|\psi\rangle$ with a stabilizer state $|\psi_{\text{STAB}}\rangle$, i.e., $\mathcal{B}(|\psi\rangle\otimes|\psi_{\text{STAB}}\rangle)=\mathcal{B}(|\psi\rangle)$. For additive Bell magic, we have

$$\mathcal{B}_a(|\psi\rangle\otimes|\psi_{\text{STAB}}\rangle) = \mathcal{B}_a(|\psi\rangle) + \mathcal{B}_a(|\psi_{\text{STAB}}\rangle) = \mathcal{B}_a(|\psi\rangle), \tag{D1}$$

where we use the properties of additivity and faithfulness. Thus, the property of composition holds for $\mathcal{B}_a$. Now, we apply the definition of additive Bell magic $\mathcal{B}=1-2^{-\mathcal{B}_a}$ to Eq. (D1), from which it immediately follows that composition with a stabilizer state leaves $\mathcal{B}$ invariant as well.

APPENDIX E: BELL MAGIC OF MAXIMALLY MIXED STATE

Here, we derive the Bell magic of the maximally mixed state $\mathcal{B}(\rho_m)$. The Bell measurement applied to the maximally mixed state $\rho_m = I/2^N$ with identity matrix I produces every possible bit string $\mathbf{r}\in\{0,1\}^{2N}$ with equal probability $P(\mathbf{r})=4^{-N}$. Now, we define the probability of all binary additions that yield $\mathbf{r}$ as $Q(\mathbf{r})=\sum_{\mathbf{q}}P(\mathbf{q})P(\mathbf{q}\oplus\mathbf{r})=4^{-N}$. Now, we can write the Bell magic as

$$\mathcal{B}=\sum_{\mathbf{r},\mathbf{q}} Q(\mathbf{r})Q(\mathbf{q})\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty = 4^{-2N}\sum_{\mathbf{r},\mathbf{q}}\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty. \tag{E1}$$

We split the equation into the two cases, $\mathbf{r}=\mathbf{0}$ and $\mathbf{r}\neq\mathbf{0}$, and obtain $\mathcal{B}=4^{-2N}\sum_{\mathbf{q}}\left\|[\sigma_{\mathbf{0}},\sigma_{\mathbf{q}}]\right\|_\infty + 4^{-2N}\sum_{\mathbf{r}\neq\mathbf{0}}\sum_{\mathbf{q}}\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty$. There are 4^N Pauli strings in total. The identity Pauli string $\sigma_{\mathbf{0}}=I$ commutes with every other Pauli string, i.e., $[\sigma_{\mathbf{0}},\sigma_{\mathbf{q}}]=0$. All other Pauli strings commute with half of the Pauli strings and do not commute with the other half, yielding $\sum_{\mathbf{q}}\left\|[\sigma_{\mathbf{r}\neq\mathbf{0}},\sigma_{\mathbf{q}}]\right\|_\infty = 4^N$. Thus,

$$\mathcal{B}(\rho_m)=4^{-2N}\sum_{\mathbf{r}\neq\mathbf{0}}4^N = 1-4^{-N}. \tag{E2}$$

APPENDIX F: MIXED BELL MAGIC

We now show that $\mathcal{B}_m(\rho)=1-(1-\mathcal{B}(\rho))/(\text{tr}(\rho^2)^2)$ is faithful for mixed stabilizer states of the form $\rho_{\text{STAB}}=U_C|\psi_{\text{STAB}}\rangle\langle\psi_{\text{STAB}}|\otimes\rho_m U_C^\dagger$, i.e., $\mathcal{B}_m(\rho_{\text{STAB}})=0$. Here, $\rho_m=I_K 2^{-K}$ is the maximally mixed state over K qubits and $|\psi_{\text{STAB}}\rangle$ is a stabilizer state of $N-K$ qubits. First, we consider the case $U_C=I$. Then, from the invariance under composition and Appendix E, it follows that

$$\mathcal{B}(\rho_{\text{STAB}})=\mathcal{B}(\rho_m)=1-4^{-N}=1-\text{tr}(\rho_m)^2, \tag{F1}$$

where in the final step we use the fact that $\text{tr}(\rho_m)=2^{-K}$. Inserting the above equation into the definition of mixed Bell magic, we find that $\mathcal{B}_m(\rho_{\text{STAB}})=0$. Faithfulness for ρ_{STAB} with $U_C\neq I$ follows from the invariance of $\mathcal{B}$ and purity $\text{tr}(\rho^2)$ under Clifford unitaries.

APPENDIX G: MAXIMAL MAGIC OF PURE STATES

We can give an upper bound $\mathcal{B}_{\text{pure}}\leq\mathcal{B}_{\text{max}}^{\text{pure}}$ on the Bell magic of pure states. Bell magic is the average of the commutator norm over the probability distribution $P(\mathbf{r})$. It becomes minimal for stabilizer states, where $P(\mathbf{r})$ is only nonzero for a small subset of $\mathbf{r}$. In contrast, it is maximal when the probability distribution is uniformly spread over as many $\mathbf{r}$ as possible, such as for the maximally mixed state. For pure states, the distribution has the constraint that $P(\mathbf{r})$ is zero for $\mathbf{r}$ with odd parity, i.e., $\text{tr}(\rho^2)=1-2P_{\text{odd}}=1$. This follows from the SWAP test in Eq. (1), where odd-parity outcomes are forbidden for pure states. We have $P_{\text{odd}}\leq 1/2(1-2^{-N})$. Now, the Bell distribution $P_{\text{max}}(\mathbf{r}$ with the highest Bell magic for pure states will have zero value for odd $\mathbf{r}_{\text{odd}}$ and a constant value for even $\mathbf{r}_{\text{even}}$. Thus, by setting all even $P(\mathbf{r}_{\text{even}})=P_{\text{uni}}$ and all odd $P(\mathbf{r}_{\text{odd}})=0$, we find from the normalization

of probability distribution that

$$P_{\text{uni}} = \frac{2 \times 4^{-N}}{1 + 2^{-N}}. \tag{G1}$$

Now, following the calculation in Sec. E, we compute

$$\mathcal{B} = \sum_{\mathbf{r},\mathbf{q}} Q(\mathbf{r}) Q(\mathbf{q}) \left\| [\sigma_{\mathbf{r}}, \sigma_{\mathbf{q}}] \right\|_{\infty} \tag{G2}$$

for $Q(\mathbf{r}) = \sum_{\mathbf{q}} P_{\max}(\mathbf{q}) P_{\max}(\mathbf{q} \oplus \mathbf{r})$. We find $Q(\mathbf{0}) = P_{\text{uni}}$ and $Q(\mathbf{r} \neq \mathbf{0}) = (1 - P_{\text{uni}})/(4^N - 1) = (1 + 2^{-N} - 2 \times 4^{-N})/((4^N - 1)(1 + 2^{-N}))$. Finally, we obtain

$$\begin{aligned} \mathcal{B}^{\text{pure}} \leq \mathcal{B}^{\text{pure}}_{\max} &= Q(\mathbf{r} \neq \mathbf{0})^2 \times 4^N (4^N - 1) \\ &= 4^N \frac{(1 + 2^{-N} - 2 \times 4^{-N})^2}{(4^N - 1)(1 + 2^{-N})^2}. \end{aligned}$$

Note that $\mathcal{B}^{\text{pure}}_{\max}$ is an upper bound on the Bell magic of pure states and in general is not saturated, as no state with the corresponding Bell-measurement distribution may exist. However, we numerically find that for $N = 1$ and $N = 3$, the bound is saturated.

We now give explicit pure states that have the maximal amount of Bell magic, which we find by using our variational quantum algorithm. For $N = 1$, this is the magic state $B^{(1)}_{\text{a}} = \log_2(27/11) \approx 1.29545588$ for magic state $|R\rangle$ with $\theta = \arccos(1/\sqrt{3})$ and $\varphi = \pi/4$. For $N = 2$, the maximal state is

$$|\psi^{(2)}\rangle = \frac{1}{2}\{1, 1, 1, i\} \tag{G3}$$

with $\mathcal{B}^{(2)}_a \approx 2.67807$. For $N = 3$, the state of maximal Bell magic states is the Hoggar state,

$$|\psi^{(3)}\rangle = \frac{1}{6}\{1 + i, 0, -1, 1, -i, 1, 0, 0\}, \tag{G4}$$

with $\mathcal{B}^{(3)}_a \approx 4.651794$, which is also the maximal state for the robustness of magic [20]. For $N = 4$, we report the maximal Bell magic of $\mathcal{B}^{(4)}_a \approx 6.221364$. While we do not find an exact form for this state, we report here a nearly maximal state with a simple description,

$$\begin{aligned} |\psi^{(4)}\rangle = &\frac{1}{8\sqrt{2}}\{4, 1 + i, 4i, -1 + i, 4i, 3(1 + i), 2i, -1 - i, \\ &- 1 + i, 4i, 3(1 - i), -2i, -1 - i, 2i, -1 + i, 2\} \end{aligned}$$

where $\mathcal{B}_a = 6.221239$. These reported states have substantially more Bell magic than corresponding Haar-random states of the same N and may be useful for state preparation.

To get a better understanding what constitutes a state of high Bell magic, we write states in the form $\rho = 2^{-N}(\sigma_{\mathbf{0}} + \sum_{\mathbf{r} \neq \mathbf{0}} \alpha_{\mathbf{r}} \sigma_{\mathbf{r}})$, where for pure states we demand that $\sum_{\mathbf{r}} |\alpha_{\mathbf{r}}|^2 = 2^N + 1$. Note that for a valid state, we additionally demand that $\alpha_{\mathbf{r}}$ are chosen such that ρ is positive semidefinite. Stabilizer states have exactly 2^N terms with $|\alpha_{\mathbf{r}}| = 1$ and zero otherwise. In contrast, we find that for states of high Bell magic all $\alpha_{\mathbf{r}}$ have nearly the same value over all 4^N Pauli operators. For example for $N = 1$ we have $|R\rangle\langle R| = 1/2(I + 1/\sqrt{3}(\sigma_x + \sigma_y + \sigma_z))$ or, for the Hoggar state for $N = 3$, we have $|\alpha_{\mathbf{r}}| = 1/3$. These states also saturate the upper bound of pure-state Bell magic. Note that for $N = 2$, the maximal pure state does not saturate the bound, as it has a nonequal distribution of $\alpha_{\mathbf{r}}$.

APPENDIX H: CONSTRUCTION OF QUANTUM CIRCUITS

To prepare quantum states in an experimentally friendly way on the IonQ quantum computer, we use parametrized quantum circuits as shown in Fig. 7 [55]. The state

$$|\psi(\boldsymbol{\theta})\rangle = \prod_{l=1}^{d} W[\bigotimes_{n=1}^{N} R_z(\boldsymbol{\theta}^z_{l,n})][\bigotimes_{n=1}^{N} R_y(\boldsymbol{\theta}^y_{l,n})]|0\rangle^{\otimes N} \tag{H1}$$

is generated by d layers of parametrized single-qubit rotations $R_\alpha(\theta) = \exp(-i(\theta)/(2)\sigma^\alpha)$, where $\alpha \in \{x, y, z\}$, and a set of fixed entangling gates W which are CNOT gates arranged in a nearest-neighbor chain configuration. We can choose the K parameters $\boldsymbol{\theta}$ for the parametrized quantum circuit in two fashions. First, we sample them uniformly $\boldsymbol{\theta}^{\text{rand}} \in [0, 2\pi)^K$. This set of parameters generates highly random quantum states $|\psi(\boldsymbol{\theta}^{\text{rand}})\rangle$ that approximate Haar-random states for sufficiently deep circuits [73]. This circuit is used to create the highly magical states for the experimental state-discrimination task on the IonQ quantum computer.

Next, to prepare stabilizer states as well as to measure the transition of stabilizer into intractable quantum states on the IonQ quantum computer, we use the same circuit with a different set of parameters. We choose $K - N_T$ parameters as $n\pi/2$, where n is an integer, and N_T parameters as multiples of $\pi/2$ shifted by $\pi/4$ with $\{n(\pi)/2 + \pi/4\}$, yielding $\boldsymbol{\theta}^{N_T} \in \{n(\pi)/2\}^{K-N_T} \otimes \{n(\pi)/2 + \pi/4\}^{N_T}$. For $N_T = 0$, all single-qubit rotations are Clifford gates and do not introduce any magic into the circuit. The entangling gates W composed of CNOT gates are Clifford gates as well and thus for $N_T = 0$, we obtain random stabilizer states. For $N_T > 0$, N_T non-Clifford gates are introduced into the circuit, which yield an increasing amount of magic. As one can easily check, the shift in parameter by $\pi/4$ is equivalent to adding a T gate (for z rotations) or a stabilizer-transformed version of the T gate in the y basis (for y rotations). For large d, our approach is equivalent to adding N_T T gates at random positions

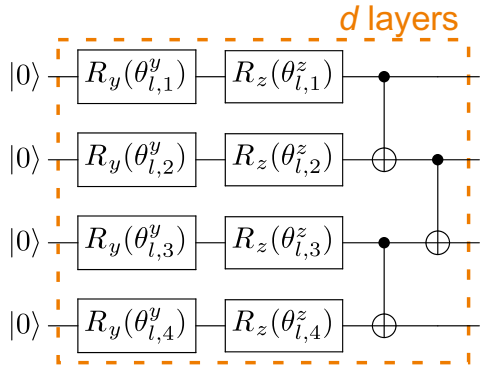

FIG. 7. A parametrized quantum circuit $|\psi_C(\boldsymbol{\theta})\rangle$ of N qubits. It is composed of d layers of single-qubit rotations around the y and z axes parametrized with $\boldsymbol{\theta}$ and entangling CNOT gates arranged in a nearest-neighbor chain configuration.

into a Clifford circuit sampled randomly from the Clifford group.

For the numerical simulation of the state-discrimination and magic estimation task of $N = 50$ qubits, we use a modified version of Fig. 7, where the initial state $|0\rangle$ is replaced by N_A magic states. In particular, the circuit consists of an initial state of N_A magic states and a Clifford circuit of depth d:

$$|\psi(\boldsymbol{\theta})\rangle = \prod_{l=1}^{d} W[\bigotimes_{n=1}^{N} R_z(\theta^z_{l,n})][\bigotimes_{n=1}^{N} R_y(\theta^y_{l,n})]|T\rangle^{N_A}|0\rangle^{\otimes N-N_A}. \tag{H2}$$

Here, the position of the N_A magic states $|T\rangle$ is randomly permuted within the N qubits. The layered unitaries are chosen as random Clifford circuits by choosing random parameters $\boldsymbol{\theta} \in \{n(\pi)/2\}^K$, where n is an integer, such that all parametrized gates are single-qubit Clifford gates. W is an entangling layer consisting of CNOT gates arranged in a nearest-neighbor chain configuration. For our simulations, we use a depth of $d = 4$. This circuit has a Bell magic $\mathcal{B}_a = N_A$.

APPENDIX I: ERROR MITIGATION

We want to determine the Bell magic of a pure state $|\psi\rangle$ subject to depolarizing noise,

$$\rho = (1-p)|\psi\rangle + pI2^{-N}, \tag{I1}$$

by measuring the depolarized state ρ. From the measurement of the purity $\mathrm{tr}(\rho_{\mathrm{dp}}^2)$, we can calculate

$$p = 1 - \frac{\sqrt{(2^N-1)(2^N\mathrm{tr}(\rho_{\mathrm{dp}}^2)-1)}}{2^N-1}. \tag{I2}$$

Now, we derive the error-mitigation method. First, the projector onto a Bell state can be written as

$$\Pi_{r_1r_2} = |\sigma_{r_1r_2}\rangle\langle\sigma_{r_1r_2}| = \frac{1}{4}(I\otimes I + E^x_{r_1r_2}\sigma^x\otimes\sigma^x + E^y_{r_1r_2}\sigma^y\otimes\sigma^y + E^z_{r_1r_2}\sigma^z\otimes\sigma^z), \tag{I3}$$

with factors $E^\alpha_{r_1r_2} = \pm 1$. The projector onto a product of Bell states is then given by $O_{\mathbf{r}} = \otimes_{n=1}^{N}\Pi_{r_{2n-1},r_{2n}}$.

Now, the state affected by depolarizing noise is given by $\rho = (1-p)|\psi\rangle\langle\psi| + pI/2^N$. The probability of measuring bit string $\mathbf{r}$ via Bell measurement on the noisy state is given by

$$\begin{aligned} P_{\mathrm{dp}}(\mathbf{r}) &= \mathrm{Tr}(\rho\otimes\rho O_{\mathbf{r}}) \\ &= (1-p)^2P_0(\mathbf{r}) + p^2 4^{-N}\mathrm{Tr}(I\otimes IO_{\mathbf{r}}) \\ &\quad + p(1-p)2^{-N}(\mathrm{Tr}(|\psi\rangle\langle\psi|\otimes IO_{\mathbf{r}}) \\ &\quad + \mathrm{Tr}(I\otimes|\psi\rangle\langle\psi|O_{\mathbf{r}})), \end{aligned}$$

with $P_0(\mathbf{r}) = \mathrm{Tr}(|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|O_{\mathbf{r}})$. With the decomposition of the projector into Pauli strings [see Eq. (I3)], $\mathrm{Tr}(I\otimes|\psi\rangle\langle\psi|\sigma^\alpha\otimes\sigma^\alpha) = 0$, $\mathrm{Tr}(I\otimes I) = 4^N$, and $\mathrm{Tr}(|\psi\rangle\langle\psi|\otimes I) = 2^N$, we obtain

$$P_{\mathrm{dp}}(\mathbf{r}) = (1-p)^2P_0(\mathbf{r}) + p(2-p)4^{-N}. \tag{I4}$$

This means that depolarization occurring on one of the copies results in the same measured probability distribution as global depolarization acting on all copies. Thus, the probability of no error occurring is given by $(1-p)^2$. Now, we define the probability

$$Q_{\mathrm{dp}}(\mathbf{r}) = \sum_{\mathbf{q}\oplus\mathbf{q}'=\mathbf{r}} P_{\mathrm{dp}}(\mathbf{q})P_{\mathrm{dp}}(\mathbf{q}') \tag{I5}$$

of getting the binary added bit string $\mathbf{r}$. The Bell magic can be then written as

$$\mathcal{B}_{\mathrm{dp}} = \sum_{\mathbf{r},\mathbf{q}\in\{0,1\}^{2N}} Q_{\mathrm{dp}}(\mathbf{r})Q_{\mathrm{dp}}(\mathbf{q})\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty. \tag{I6}$$

Here, we use that $\sigma_{\mathbf{r}} = \sigma_{\mathbf{q}}\sigma_{\mathbf{q}'} = \sigma_{\mathbf{q}\oplus\mathbf{q}'}$ up to a multiplication with $\{1,-1,i,-i\}$. Inserting $P_{\mathrm{dp}}(\mathbf{r})$ with the depolarizing noise, we obtain $Q_{\mathrm{dp}}(\mathbf{r}) = (1-p)^4Q_0(\mathbf{r}) + (1-(1-p)^4)4^{-N}$, where $Q_0(\mathbf{r}) = \sum_{\mathbf{q}\oplus\mathbf{q}'}P_0(\mathbf{q})P_0(\mathbf{q})$ is the probability for the pure state. Here, we use the fact that $\sum_{\mathbf{r}\in\{0,1\}^{2N}}P_0(\mathbf{r}) = 1$ and $\sum_{\mathbf{r}\in\{0,1\}^{2N}} = 4^N$. We now define

$$p_c = 1-(1-p)^4 \tag{I7}$$

to simplify to $Q_{\mathrm{dp}}(\mathbf{r}) = p_cQ_0(\mathbf{r}) + (1-p_c)4^{-N}$. We can now write out the Bell magic in terms of the probability

Q_0 for the noise-free state

$$\begin{aligned}\mathcal{B}_{\rm dp} &= \sum_{\mathbf{r},\mathbf{q}\in\{0,1\}^{2N}} Q_{\rm dp}(\mathbf{r})Q_{\rm dp}(\mathbf{q})\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty \\ &= \sum_{\mathbf{r},\mathbf{q}} \left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty (p_c^2 Q_0(\mathbf{r})Q_0(\mathbf{q}) \\ &\quad + (1-p_c)^2 4^{-2N} + 2p_c(1-p_c)4^{-N}Q_0(\mathbf{r}) \\ &= (1-p_c)^2\mathcal{B}^{\rm mtg} + p_c^2\mathcal{B}(\rho_m) + 2p_c(1-p_c)\mathcal{B}^R.\end{aligned}$$

Here, $\mathcal{B}(\rho_m) = \sum_{\mathbf{r},\mathbf{q}} 4^{-2N}\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty = 1-4^{-N}$ is the Bell magic of the maximally mixed state, $\mathcal{B}^{\rm mtg} = \sum_{\mathbf{r},\mathbf{q}} Q_0(\mathbf{r})Q_0(\mathbf{q})\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty$ is the mitigated Bell magic of the noise-free state, and we define

$$\mathcal{B}^R = 4^{-N}\sum_{\mathbf{r},\mathbf{q}} Q_0(\mathbf{r})\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty. \tag{I8}$$

We now show how to calculate $\mathcal{B}^R$. First, we split this term into the cases $\mathbf{r}=\mathbf{0}$ and $\mathbf{r}\neq\mathbf{0}$:

$$\begin{aligned}\mathcal{B}^R &= 4^{-N}\sum_{\mathbf{q}} Q_0(\mathbf{0})\left\|[\sigma_{\mathbf{0}},\sigma_{\mathbf{q}}]\right\|_\infty \\ &\quad + 4^{-N}\sum_{\mathbf{r}\neq\mathbf{0}} Q_0(\mathbf{r})\sum_{\mathbf{q}}\left\|[\sigma_{\mathbf{r}},\sigma_{\mathbf{q}}]\right\|_\infty.\end{aligned}$$

A given Pauli string $\sigma_{\mathbf{r}}$ with $\mathbf{r}\neq\mathbf{0}$ commutes with half of all 4^N Pauli strings, while it does not commute with the other half. Thus, $\sum_{\mathbf{q}}\left\|[\sigma_{\mathbf{r}\neq\mathbf{0}},\sigma_{\mathbf{q}}]\right\|_\infty = 4^N$. For the case $\mathbf{r}=\mathbf{0}$, the Pauli string $\sigma_{\mathbf{0}} = I$ is the identity and thus it always commutes $\left\|[\sigma_{\mathbf{0}},\sigma_{\mathbf{q}}]\right\|_\infty = 0$. Using $4^{-N}\sum_{\mathbf{r}\neq\mathbf{0}} 1 = 1 - Q_0(\mathbf{0})$, we obtain

$$\begin{aligned}\mathcal{B}^R &= 1 - Q_0(\mathbf{0}) = 1 - \sum_{\mathbf{q}\oplus\mathbf{q}'=\mathbf{0}} P_0(\mathbf{q})P_0(\mathbf{q}') \\ &= 1 - \sum_{\mathbf{q}} P_0(\mathbf{q})^2.\end{aligned} \tag{I9}$$

For a stabilizer state $\sum_{\mathbf{q}} P_0(\mathbf{q})^2 = 2^{-N}$, as $P(\mathbf{q}) = 2^{-N}|\langle\psi|\sigma_{\mathbf{r}}|\psi^*\rangle|^2$ and there are 2^N Pauli strings with nonzero expectation values. On the other hand, for a maximally mixed state, we find that $\sum_{\mathbf{q}} P_0(\mathbf{q})^2 = 4^{-N}$, as every bit string $\mathbf{q}$ appears with equal probability. Thus, we can bound $1-2^{-N} \le \mathcal{B}^R \le 1-4^{-N}$. Now we want to calculate $\sum_{\mathbf{q}} P_0(\mathbf{q})^2$ by measuring the depolarized state. The sum of the squares of the probabilities of the bit strings is given by $\sum_{\mathbf{q}} P_{\rm dp}(\mathbf{q})^2$. Inserting Eq. (I4), we obtain $\sum_{\mathbf{q}} P_{\rm dp}(\mathbf{q})^2 = (1-p_c)\sum_{\mathbf{q}} P_0(\mathbf{q})^2 + p_c 4^{-N}$. We invert this equation to obtain

$$\sum_{\mathbf{q}} P_0(\mathbf{q})^2 = \frac{\sum_{\mathbf{q}} P_{\rm dp}(\mathbf{q})^2 - 4^{-N}p_c}{1-p_c}. \tag{I10}$$

Putting all our results together, the mitigated Bell magic is given by

$$\mathcal{B}^{\rm mtg} = \frac{1}{(1-p_c)^2}(\mathcal{B}^{\rm dp} - p_c^2\mathcal{B}(\rho_m) - 2p_c(1-p_c)\mathcal{B}^R), \tag{I11}$$

where $\mathcal{B}^R = 1 - (1-p_c)^{-1}(\sum_{\mathbf{q}} P_{\rm dp}(\mathbf{q})^2 - 4^{-N}p_c)$. For a large number of qubits N, the sum of probabilities becomes exponentially small—$4^{-N} \le \sum_{\mathbf{q}} P(\mathbf{q})^2 \le 2^{-N}$—and becomes challenging to measure. In this limit, we can approximate $\mathcal{B}^R \approx \mathcal{B}(\rho_m) \approx 1$ and finally obtain

$$\mathcal{B}^{\rm mtg} \approx \frac{\mathcal{B}^{\rm dp} - p_c(2-p_c)}{(1-p_c)^2}. \tag{I12}$$

APPENDIX J: SUPERVISED LEARNING FOR DECISION BOUNDARIES

We want to learn to classify unknown states using Bell magic measured on (noisy) states with a finite number of measurement samples N_Q. We have two classes of states with different amounts of Bell magic, i.e., class β with stabilizer states with low Bell magic and class α with random states with high Bell magic. To train the classifier, we are given a training set of $N_{\rm train}$ states, where we know to which class the states belong by virtue of the label $y_i \in \{-1,1\}$. The label $y_i = -1$ indicates class β and $y_i = 1$ class α. We now measure Bell magic using N_Q measurement samples for each state of the training set and estimate the Bell magic for each state $\hat{\mathcal{B}}^{(i)}$. Now, we want to find the best threshold $\mathcal{B}^*$ that separates the two classes such that $\hat{\mathcal{B}}^{(i)} \le \mathcal{B}^*$ is correctly assigned as $\hat{y}_i = -1$ and $\hat{\mathcal{B}}_2^{(i)} > \mathcal{B}^*$ as $\hat{y}_i = 1$. To find the best threshold, we maximize

$$\mathcal{B}^*_{\rm opt} = \max_{\mathcal{B}^*}\sum_{i=1}^{N_{\rm train}} {\rm sign}(\hat{\mathcal{B}}^{(i)} - \mathcal{B}^*)y_i. \tag{J1}$$

To evaluate the performance of the classifier, we test the threshold $\mathcal{B}^*_{\rm opt}$ on an unlabeled test data set of $N_{\rm test}$ states that have not been used during training. We define $P_{\rm error} = N_{\rm test}^{\rm wrong}/N_{\rm test}$ as the probability of wrongly classifying $N_{\rm test}^{\rm wrong}$ states. The trivial strategy of guessing at random would achieve an error probability of $P_{\rm error} = 1/2$.

APPENDIX K: DATA FOR LEARNING-STATE DISCRIMINATION

In Fig. 8, we show the data that we use for our experimental demonstration of the state-discrimination protocol

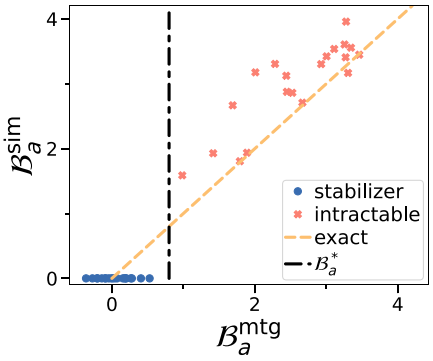

FIG. 8. The simulated $\mathcal{B}_a^{\text{sim}}$ plotted against the experimentally measured mitigated additive Bell magic $\mathcal{B}_a^{\text{mtg}}$. These data points are used for the state-discrimination task in the main text. We show the result for $N_Q = 1000$ measurements and $N = 3$ qubits.

on the IonQ quantum computer. We show the simulated $\mathcal{B}_a^{\text{sim}}$ and the mitigated additive Bell magic $\mathcal{B}_a^{\text{mtg}}$. The blue shows the stabilizer data, while the orange shows the highly magical states generated by a hardware-efficient circuit with random parameters as defined in Appendix H. The linear curve is the relationship expected for perfect results. We observe that the Bell magic is slightly underestimated in experiment. The vertical dashed line is the optimal threshold $\mathcal{B}_a^*$ to discriminate stabilizer and non-stabilizer states from the experiment for $N_Q = 1000$. For smaller N_Q, the measurement error increases and no $\mathcal{B}_a^*$ to perfectly distinguish the two classes can be found, yielding a finite classification error.

APPENDIX L: SHIFT RULE FOR BELL MAGIC

We now derive the shift rule for Bell measurements and the Bell magic. The shift rule provides exact gradients when the circuit is composed of parametrized Pauli rotations [64]. For standard measurements on single quantum states, the shift rule is given by $\partial_k\langle C(\boldsymbol{\theta})\rangle = v(\langle C(\boldsymbol{\theta} + \mathbf{e}_k(\pi)/4v)\rangle - \langle C(\boldsymbol{\theta} - \mathbf{e}_k(\pi)/4v)\rangle)$, where $\mathbf{e}_k$ is the kth unit vector and r is an arbitrary number [65].

For any operators U, V, and O, we can write [64]

$$\langle\psi|U^\dagger OV|\psi\rangle + \text{h.c.} = \frac{1}{2}[\langle\psi|(U+V)^\dagger O(U+V)|\psi\rangle - \langle\psi|(U-V)^\dagger O(U-V)|\psi\rangle], \quad \text{(L1)}$$

where h.c. indicates the Hermitian conjugate of the preceding terms.

For Bell magic, we have to estimate the probability $P(\mathbf{r}) = \langle\psi|\langle\psi|O_\mathbf{r}|\psi\rangle|\psi\rangle$ of measuring a Bell state $|\sigma_\mathbf{r}\rangle$, where $O_\mathbf{r} = |\sigma_\mathbf{r}\rangle\langle\sigma_\mathbf{r}|$ is the projector onto the Bell state. The derivate of the parametrized quantum circuit $|\psi(\boldsymbol{\theta})\rangle = \prod_{n=1}^d V_n(\boldsymbol{\theta}_n)W_n|0\rangle$ with d layers, entangling gates W_n, parameters $\boldsymbol{\theta}$, and parametrized rotations $V_n(\boldsymbol{\theta}_n) = e^{-i\frac{\boldsymbol{\theta}_k}{2}\sigma_n}$ is given by some Pauli string σ_n, the derivative on the quantum state $|\psi\rangle$ can be written as

$$\begin{aligned}\partial_k|\psi(\boldsymbol{\theta})\rangle &= \prod_{n=k+1}^{d}[V_n(\boldsymbol{\theta}_n)W_n](-i\frac{1}{2}\sigma_k)\prod_{n=1}^{k}[V_n(\boldsymbol{\theta}_n)W_n]|0\rangle \\ &\equiv U_k(-i\frac{1}{2}\sigma_k)|\phi_k\rangle,\end{aligned}$$

where in the last step we define $|\phi_k\rangle = \prod_{n=1}^k V_n(\boldsymbol{\theta}_n)W_n|0\rangle$ and $U_k = \prod_{n=k+1}^d V_n(\boldsymbol{\theta}_n)W_n$. Now, the derivative of $P(\mathbf{r})$ using the product rule is given by

$$\begin{aligned}\partial_k P(\mathbf{r}) =& 2\langle\phi_k|\langle\phi_k|(U_k\otimes U_k)^\dagger O_\mathbf{r}(U_k\otimes U_k) \\ & [(-i\frac{1}{2}\sigma_k)\otimes I]|\phi_k\rangle|\phi_k\rangle + \text{h.c.}\end{aligned}$$

We now define, for simplicity, $O'_\mathbf{r} = (U_k\otimes U_k)^\dagger O_\mathbf{r}(U_k\otimes U_k)$, introduce an arbitrary factor $v > 0$, and apply Eq. (L1):

$$\begin{aligned}\partial_k P(\mathbf{r}) &= 2v\langle\phi_k|\langle\phi_k|O'_\mathbf{r}[(-i\frac{1}{2v}\sigma_k)\otimes I]|\phi_k\rangle|\phi_k\rangle + \text{h.c.} \\ &= v\langle\phi_k|\langle\phi_k|[(I - i\frac{1}{2v}\sigma_k)^\dagger\otimes I]O'_\mathbf{r} \\ &\quad\times[(I - i\frac{1}{2v}\sigma_k)\otimes I]|\phi_k\rangle|\phi_k\rangle \\ &\quad - v\langle\phi_k|\langle\phi_k|[(I + i\frac{1}{2v}\sigma_k)^\dagger\otimes I]O'_\mathbf{r} \\ &\quad\times[(I + i\frac{1}{2v}\sigma_k)\otimes I]|\phi_k\rangle|\phi_k\rangle.\end{aligned}$$

For any Pauli strings σ, we can rewrite the generators into a unitary as follows [64]:

$$e^{-i\frac{\pi}{4v}\frac{1}{2}\sigma} = \frac{1}{\sqrt{2}}(I - i\frac{1}{2v}\sigma)\,. \quad \text{(L2)}$$

We now find that

$$\begin{aligned}&\partial_k P(\mathbf{r}) \\ &= 2v\langle\phi_k|\langle\phi_k|[e^{-i\frac{\pi}{4v}\frac{1}{2}\sigma_n}\otimes I]^\dagger O'_\mathbf{r}[e^{-i\frac{\pi}{4v}\frac{1}{2}\sigma_n}\otimes I]|\phi_k\rangle|\phi_k\rangle \\ &- 2v\langle\phi_k|\langle\phi_k|[e^{i\frac{\pi}{4v}\frac{1}{2}\sigma_n}\otimes I]^\dagger O'_\mathbf{r}[e^{i\frac{\pi}{4v}\frac{1}{2}\sigma_n}\otimes I]|\phi_k\rangle|\phi_k\rangle \\ &= 2v\langle\psi(\boldsymbol{\theta}+\frac{\pi}{4v}\mathbf{e}_k)|\langle\psi(\boldsymbol{\theta})|O_\mathbf{r}|\psi(\boldsymbol{\theta}+\frac{\pi}{4v}\mathbf{e}_k)\rangle|\psi(\boldsymbol{\theta})\rangle \\ &- 2v\langle\psi(\boldsymbol{\theta}-\frac{\pi}{4v}\mathbf{e}_k)|\langle\psi(\boldsymbol{\theta})|O_\mathbf{r}|\psi(\boldsymbol{\theta}-\frac{\pi}{4v}\mathbf{e}_k)\rangle|\psi(\boldsymbol{\theta})\rangle,\end{aligned} \quad \text{(L3)}$$

where in the last step we introduce the kth unit vector $\mathbf{e}_k$ and we absorb $e^{-i(\pi)/(4v)(1)/(2)\sigma_n}$ into the definition of the kth parametrized rotation, i.e., $U_k e^{-i(\pi)/(4v)(1)/(2)\sigma_n}|\phi_k\rangle$

$= |\psi(\boldsymbol{\theta} + (\pi)/(4v)\mathbf{e}_k)\rangle$. Finally, with the product rule, the gradient of Bell magic is given by

$$\partial_k \mathcal{B} = 4 \sum_{\substack{\mathbf{r},\mathbf{r}',\mathbf{q},\mathbf{q}' \\ \in\{0,1\}^{2N}}} [\partial_k P(\mathbf{r})] P(\mathbf{r}') P(\mathbf{q}) P(\mathbf{q}') \left\| [\sigma_{\mathbf{r}}\sigma_{\mathbf{r}'}, \sigma_{\mathbf{q}}\sigma_{\mathbf{q}'}] \right\|_\infty . \tag{L4}$$

APPENDIX M: TRAINABILITY OF BELL MAGIC

Here, we present the calculation of the gradient of the Bell magic for the ansatz $|\psi(\theta, U_C\rangle = U_C \exp(-i\frac{1}{2}\theta\sigma_1^y)|0\rangle^{\otimes N}$. A straightforward calculation yields $\mathcal{B}(\theta, U_C) = \frac{1}{2}\sin(2\theta)^2$. Here, we use that Bell magic is invariant under the choice of Clifford circuit U_C. For the gradient, we find that $\partial_\theta \mathcal{B}(\theta, U_C) = \sin(4\theta)$. We calculate the variance of the gradient by integrating over θ, which yields

$$\mathrm{Var}(\partial_\theta \mathcal{B}(\theta, U_C))_{\theta, U_C} = \frac{1}{2\pi}\left(\int_0^{2\pi} [\partial_\theta \mathcal{B}(\theta, U_C)]^2 \mathrm{d}\theta - \left(\int_0^{2\pi} \partial_\theta \mathcal{B}(\theta, U_C)\mathrm{d}\theta\right)^2\right) = \frac{1}{2}. \tag{M1}$$

APPENDIX N: STATE DISCRIMINATION FOR A SINGLE MAGIC STATE

Here, we derive the error probability of classifying a Clifford circuit with exactly one magic state as input. The state $|\psi_C(\phi)\rangle = U_C|\phi\rangle \otimes |0\rangle^{N-1}$ consists of an arbitrary Clifford circuit U_C and an initial state $|A_\phi\rangle \otimes |0\rangle^{N-1}$ with exactly one nonstabilizer qubit $|A_\phi\rangle = \cos(\phi/2)|0\rangle + \sin(\phi/2)|1\rangle$. ϕ controls the amount of Bell magic introduced into the circuit. In particular, for $\phi = n\pi/2$, where n is an integer, no magic is introduced, whereas for $\phi = \pi/4$, the Bell magic introduced in the circuit is equivalent to the $|T\rangle$ state.

We now derive the error for the case $U_C = I$ and $N = 1$ with $|\psi_C(\phi)\rangle = |A_\phi\rangle$. General N and U_C follow from faithfulness, composition, and invariance under transformation with U_C for the Bell magic. First, we apply the Bell transformation and obtain

$$U_{\mathrm{Bell}}|A_\phi\rangle \otimes |A_\phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle - \cos(\phi)|10\rangle + \sin(\phi)|01\rangle). \tag{N1}$$

The probabilities of measuring the respective bit strings are $p_{00} = 1/2$, $p_{10} = 1/(2)\cos(\phi)^2$ and $p_{10} = 1/(2)\sin(\phi)^2$. Now, we sample this state N_Q times and apply the algorithm for the Bell magic. We assume the case of a large number of resampling steps N_R such that the algorithm adds all possible combination pairs of bit strings together, then checks whether any of those pairs correspond to noncommuting Pauli strings. One can easily check that such a noncommuting pair of Pauli strings is only found when one samples each possible bit string {00}, {01}, and {10} at least once. Thus, the error probability of finding only commuting Pauli strings and thus wrongly estimating $\mathcal{B} = 0$ is given by

$$P_E(N_Q) = \sum_{k=0}^{N_Q} \binom{N_Q}{k} [p_{00}^{N-k}p_{01}^k + p_{00}^{N-k}p_{10}^k + p_{10}^{N-k}p_{01}^k] - p_{00}^N - p_{10}^N - p_{01}^N.$$

Here, we sum over all possible combinations of measuring only two kinds of bit strings. The last three terms are subtracted as these terms appear twice in the sums. After inserting the probabilities and simplifying, we obtain

$$P_E(\phi) = 4^{-N_Q}[(3 - \cos(2\phi))^{N_Q} + (3 + \cos(2\phi))^{N_Q}] - 2^{-N_Q}[\sin(\phi)^{2N_Q} + \cos(2\phi)^{2N_Q}].$$

APPENDIX O: STABILIZER ENTROPY AND BELL MEASUREMENT

Stabilizer entropy is a class of measures of magic that have been recently introduced [10]. They can be measured using a randomized-measurement approach. As an alternative approach, we show how to measure the stabilizer entropy with Bell measurements. The stabilizer 2-Rényi entropy is given by

$$M_2 = -\log(2^N \sum_{\mathbf{r}} (2^{-N}\langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2)^2) \tag{O1}$$

and the linear stabilizer entropy by

$$M_{\mathrm{lin}} = 1 - 2^N \sum_{\mathbf{r}} (2^{-N}\langle\psi|\sigma_{\mathbf{r}}|\psi\rangle^2)^2. \tag{O2}$$

With Eq. (B1) and setting $\mathbf{n} = \mathbf{0}$, we can write the stabilizer 2-Rényi entropy with the outcome probability $P(\mathbf{r})$ of Bell measurements as

$$M_2 = -\log(2^N \sum_{\mathbf{r}} P(\mathbf{r})^2) \tag{O3}$$

and the linear stabilizer entropy

$$M_{\mathrm{lin}} = 1 - 2^N \sum_{\mathbf{r}} P(\mathbf{r})^2. \tag{O4}$$

Note that the explicit estimation of $P(\mathbf{r})$ requires a number of measurement samples that scale exponentially with the number of qubits N. Using our shift rule, given in Eq. (L3), one could also maximize M_2 in a variational quantum algorithm.

We can also mitigate errors of M_2 on noisy quantum computers. Assuming depolarizing noise p as outlined in

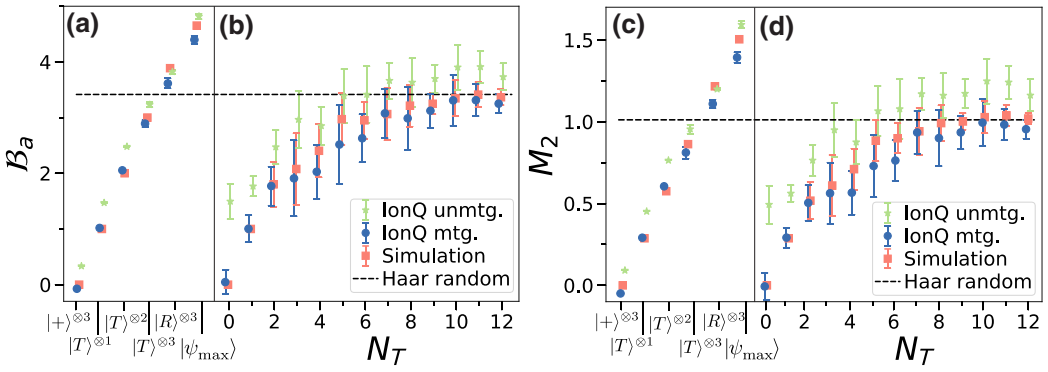

FIG. 9. An experiment to measure (a),(b) additive Bell magic $\mathcal{B}_a$ and (c),(d) stabilizer 2-Rényi entropy M_2 on the IonQ quantum computer for various types of states. (a) In the left part of the graph, we show product states of stabilizer states $|+\rangle^{\otimes N}$, magic states $|T\rangle^{\otimes N}$ and $|R\rangle^{\otimes N}$ as well as the state of maximal Bell magic, $|\psi_{\max}\rangle$ for $N = 3$. (b) In the right part of the graph, we show magic as a function of N_T T gates inserted at random positions in a Clifford circuit. For all measures of magic, we show the unmitigated and mitigated magic from the IonQ quantum computer as well as an exact simulation of the quantum states. The mean value and the standard deviation are taken over six random instances of the state for $N = 3$ qubits. The dashed line is the Bell magic averaged over Haar-random states. The experiment is performed with $N_Q = 10^3$ measurement samples and no further error or readout error mitigation. The purity measured on the IonQ quantum computers gives us a depolarization error of $p \approx 0.1$.

Sec. I, we measure the probabilities $P_{\rm dp}(\mathbf{r})$ of the noisy state. Using the purity estimated from the SWAP test of the Bell measurements in Eq. (I2), we can estimate p. Then, the mitigated probabilities $P_0(\mathbf{r})$ can be computed via Eq. (I4) as

$$P_0(\mathbf{r}) = \frac{P_{\rm dp}(\mathbf{r}) - p(2-p)4^{-N}}{(1-p)^2}. \tag{O5}$$

Note that negative probabilities can appear due to shot noise or the noise not being perfectly depolarizing. In this case, we set all $P_0(\mathbf{r}) < 0$ to zero.

In Fig. 9, we compare the experimental results for Bell measurements on the IonQ quantum computer for stabilizer entropy and Bell magic. For all measures, we use the same experimental data. We show additive Bell magic $\mathcal{B}_a$ in Figs. 9(a) and 9(b) and stabilizer 2-Rényi entropy M_2 in Figs. 9(c) and 9(d). We find that error mitigation improves the result for all measures. For the exact simulation, all measures produce similar behavior; in particular, we observe that both $\mathcal{B}_a$ and M_2 are additive for the various $|T\rangle$ states with N. Both measures share the same state of maximal Bell magic. Further, we see that with increasing N_T, the average magic of both measures converges to the respective value found for Haar-random states.

APPENDIX P: ENTANGLEMENT AND BELL MEASUREMENTS

We now investigate the entanglement of the states that we studied in the main text for Bell magic. The Meyer-Wallach measure $\mathcal{E}$ [59] has been proposed as a measure to characterize the entanglement of states prepared on quantum computers [72]. First, one defines a mapping $\iota_j(e)$ that acts on the computational basis states $\iota_j(b)|b_1\cdots b_n\rangle = \delta_{bb_j}|b_1\cdots\tilde{b}_j\cdots b_n\rangle$, where $b_j \in \{0,1\}$ and $\tilde{b}_j$ denotes the absence of the jth qubit. The Meyer-Wallach measure is then defined as

$$\mathcal{E}(|\psi\rangle) \equiv \frac{4}{n}\sum_{j=1}^{n} D\big(\iota_j(0)|\psi\rangle, \iota_j(1)|\psi\rangle\big), \tag{P1}$$

where D is the generalized distance of the coefficients of two states $|u\rangle = \sum u_i|e_i\rangle$ and $|v\rangle = \sum v_i|e_i\rangle$,

$$D(|u\rangle,|v\rangle) = \frac{1}{2}\sum_{i,j}|u_i v_j - u_j v_i|^2. \tag{P2}$$

It can be rewritten as [90]

$$\mathcal{E}(\psi) = 2\Big(1 - \frac{1}{N}\sum_{i=k}^{N}\mathrm{tr}(\rho_k^2)\Big), \tag{P3}$$

where $\rho_k = \mathrm{tr}_k(\rho)$ is the partial trace of $|\psi\rangle$ over all qubits except for the kth qubit. It is zero for pure product states and can maximally reach $\mathcal{E} = 1$ for classes of highly entangled states such as the Greenberger-Horne-Zeilinger (GHZ) state. For Haar-random states and random stabilizer states, we find that $\mathcal{E}(|\psi_{\rm Haar}\rangle) = (2^N-2)/(2^N+1)$ [91]. We now assume that a pure state $|\psi\rangle$ is subject to depolarizing error p, resulting in the noisy state $\rho_{\rm dp}$ as defined in Eq. (I1). By measuring the Meyer-Wallach measure $\mathcal{E}_{\rm dp}$ on $\rho_{\rm dp}$, we want to compute the mitigated measure $\mathcal{E}_{\rm mtg}$ for the corresponding pure state $|\psi\rangle$. We now apply the

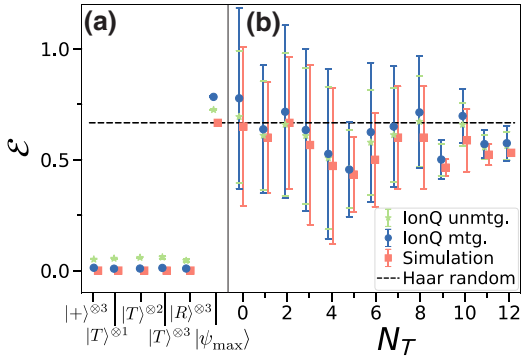

FIG. 10. The measurement of the Wallach-Meyer measure $\mathcal{E}$ for entanglement on the IonQ quantum computer. We compute $\mathcal{E}$ for various type of states. We use the same states and parameters as in Fig. 9. The dashed line is $\mathcal{E}$ for Haar-random states.

error-mitigation method in Appendix I with

$$\rho_{k,\mathrm{dp}} = \rho_k(1-p) + \frac{1}{2}I_k p. \tag{P4}$$

After squaring and taking the trace over $\rho_{k,\mathrm{dp}}$, we find that

$$\mathcal{E}_{\mathrm{mtg}} = \frac{\mathcal{E}_{\mathrm{dp}} - (2-p)\frac{p}{2}}{(1-p)^2}. \tag{P5}$$

Note that the $\mathrm{tr}(\rho_k^2)$ can be efficiently computed via Bell measurements [46]. In particular, $\mathrm{tr}(\rho_k^2) = 1 - 2P_{\mathrm{odd,k}}$, where $P_{\mathrm{odd,k}}$ is the probability of odd parity of the outcomes measured on the kth qubits of the two copies.

We experimentally measure $\mathcal{E}$ on the IonQ quantum computer in Fig. 10. We use the same states and parameters as used for computing Bell magic in Fig. 9. As expected, we find experimentally that product states have $\mathcal{E} \approx 0$. In contrast, entangled states such as $|\psi\rangle_{\max}$ and the random Clifford states with T gates have high $\mathcal{E}$. We find that $\mathcal{E}$ is nearly independent of N_T and close to the average value expected for Haar-random states. We observe some variance in our result, as we only consider states with a short circuit depth and we measure only a small number of states. Our results highlight the complementary properties of Bell magic and entanglement of different types of states, which can be easily measured with Bell measurements on noisy quantum computers.

[1] D. Gottesman, The Heisenberg representation of quantum computers, arXiv:quant-ph/9807006 (1998).

[2] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).

[3] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the "magic" for quantum computation, Nature **510**, 351 (2014).

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

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

[6] 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).

[7] 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).

[8] J. R. Seddon and E. T. Campbell, Quantifying magic for multi-qubit operations, Proc. R. Soc. A **475**, 20190251 (2019).

[9] N. Koukoulekidis, H. Kwon, H. H. Jee, D. Jennings, and M. Kim, Born probability estimation via gate merging and dynamic frame optimisation, arXiv:2202.12114 (2022).

[10] L. Leone, S. F. Oliviero, and A. Hamma, Stabilizer Rényi Entropy, Phys. Rev. Lett. **128**, 050402 (2022).

[11] L. Leone, S. F. Oliviero, Y. Zhou, and A. Hamma, Quantum chaos is quantum, Quantum **5**, 453 (2021).

[12] J. Haferkamp, Random quantum circuits are approximate unitary t-designs in depth $o\left(nt^{5+o(1)}\right)$, arXiv:2203.16571 (2022).

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

[14] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature **549**, 172 (2017).

[15] P. W. Shor, in *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE, Burlington, VT, USA, 1996), p. 56.

[16] J. Preskill, in *Introduction to Quantum Computation and Information* (World Scientific, Singapore, 1998), p. 213.

[17] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, Nature **402**, 390 (1999).

[18] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. (NY) **303**, 2 (2003).

[19] E. T. Campbell, Catalysis and activation of magic states in fault-tolerant architectures, Phys. Rev. A **83**, 032317 (2011).

[20] M. Howard and E. Campbell, Application of a Resource Theory for Magic States to Fault-Tolerant Quantum Computing, Phys. Rev. Lett. **118**, 090501 (2017).

[21] O. Hahn, A. Ferraro, L. Hultquist, G. Ferrini, and L. García-Álvarez, Quantifying Qubit Magic with Gottesman-Kitaev-Preskill Encoding, Phys. Rev. Lett. **128**, 210502 (2022).

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

[23] M. Heinrich and D. Gross, Robustness of magic and symmetries of the stabiliser polytope, Quantum **3**, 132 (2019).

[24] X. Wang, M. M. Wilde, and Y. Su, Quantifying the magic of quantum channels, New J. Phys. **21**, 103002 (2019).

[25] M. Beverland, E. Campbell, M. Howard, and V. Kliuchnikov, Lower bounds on the non-Clifford resources for

quantum computations, Quantum Sci. Technol. **5**, 035009 (2020).
[26] G. Saxena and G. Gour, Quantifying dynamical magic with completely stabilizer preserving operations as free, arXiv:2202.07867 (2022).
[27] H. Dai, S. Fu, and S. Luo, Detecting magic states via characteristic functions, Int. J. Theor. Phys. **61**, 1 (2022).
[28] N. Delfosse, P. A. Guerin, J. Bian, and R. Raussendorf, Wigner Function Negativity and Contextuality in Quantum Computation on Rebits, Phys. Rev. X **5**, 021003 (2015).
[29] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, Science **364**, 260 (2019).
[30] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).
[31] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).
[32] C. Ryan-Anderson, J. Bohnet, K. Lee, D. Gresh, A. Hankin, J. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, Realization of Real-Time Fault-Tolerant Quantum Error Correction, Phys. Rev. X **11**, 041058 (2021).
[33] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, and C. Monroe, Fault-tolerant control of an error-corrected qubit, Nature **598**, 281 (2021).
[34] Y. Zhao *et al.*, Realization of an Error-Correcting Surface Code with Superconducting Qubits, Phys. Rev. Lett. **129**, 030501 (2022).
[35] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, and J. Herrmann *et al.*, Realizing repeated quantum error correction in a distance-three surface code, arXiv:2112.03708 (2021).
[36] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, Quantum certification and benchmarking, Nat. Rev. Phys. **2**, 382 (2020).
[37] J. Carrasco, A. Elben, C. Kokail, B. Kraus, and P. Zoller, Theoretical and Experimental Perspectives of Quantum Verification, PRX Quantum **2**, 010102 (2021).
[38] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature **574**, 505 (2019).
[39] Z.-W. Liu and A. Winter, Many-Body Quantum Magic, PRX Quantum **3**, 020333 (2022).
[40] C. D. White, C. Cao, and B. Swingle, Conformal field theories are magical, Phys. Rev. B **103**, 075145 (2021).
[41] M. A. Nielsen and I. Chuang, Quantum computation and quantum information (2002).
[42] A. K. Ekert, C. M. Alves, D. K. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Direct Estimations of Linear and Nonlinear Functionals of a Quantum State, Phys. Rev. Lett. **88**, 217901 (2002).
[43] D. Aharonov, J. Cotler, and X.-L. Qi, Quantum algorithmic measurement, Nat. Commun. **13**, 1 (2022).
[44] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Denver, CO, USA, 2022), p. 574.
[45] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, Science **376**, 1182 (2022).
[46] J. C. Garcia-Escartin and P. Chamorro-Posada, Swap test and Hong-Ou-Mandel effect are equivalent, Phys. Rev. A **87**, 052330 (2013).
[47] A. W. Harrow and A. Montanaro, Testing product states, quantum Merlin-Arthur games and tensor optimization, J. ACM (JACM) **60**, 1 (2013).
[48] R. Islam, R. Ma, P. M. Preiss, M. E. Tai, A. Lukin, M. Rispoli, and M. Greiner, Measuring entanglement entropy in a quantum many-body system, Nature **528**, 77 (2015).
[49] A. Montanaro, Learning stabilizer states by Bell sampling, arXiv:1707.04012 (2017).
[50] 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).
[51] J. Vovrosh, K. E. Khosla, S. Greenaway, C. Self, M. S. Kim, and J. Knolle, Simple mitigation of global depolarizing errors in quantum simulations, Phys. Rev. E **104**, 035309 (2021).
[52] T. Haug, C. N. Self, and M. Kim, Quantum machine learning of large datasets using randomized measurements, arXiv:2108.01039 (2021).
[53] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).
[54] M. Urbanek, B. Nachman, V. R. Pascuzzi, A. He, C. W. Bauer, and W. A. de Jong, Mitigating Depolarizing Noise on Quantum Computers with Noise-Estimation Circuits, Phys. Rev. Lett. **127**, 270502 (2021).
[55] T. Haug, K. Bharti, and M. Kim, Capacity and Quantum Geometry of Parametrized Quantum Circuits, PRX Quantum **2**, 040309 (2021).
[56] M. Fishman, S. White, and E. Stoudenmire, The ITensor software library for tensor network calculations, SciPost Physics Codebases, 004 (2022).
[57] G. Torlai and M. Fishman, PastaQ: A package for simulation, tomography and analysis of quantum computers (2020).
[58] K. Wright *et al.*, Benchmarking an 11-qubit quantum computer, Nat. Commun. **10**, 1 (2019).
[59] D. A. Meyer and N. R. Wallach, Global entanglement in multiparticle systems, J. Math. Phys. **43**, 4273 (2002).
[60] F. G. Brandao, A. W. Harrow, and M. Horodecki, Local random quantum circuits are approximate polynomial-designs, Commun. Math. Phys. **346**, 397 (2016).
[61] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth, Efficient unitary designs with a system-size independent number of non-Clifford gates, Commun. Math. Phys., 1 (2022).
[62] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. Obrien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[63] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).
[64] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).
[65] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).
[66] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).
[67] B. van Straaten and B. Koczor, Measurement Cost of Metric-Aware Variational Quantum Algorithms, PRX Quantum **2**, 030324 (2021).
[68] T. Haug and M. S. Kim, Optimal training of variational quantum algorithms without barren plateaus, arXiv:2104.14543 (2021).
[69] T. Haug and M. S. Kim, Natural parametrized quantum circuit, Phys. Rev. A **106**, 052611 (2022).
[70] A. Mari, T. R. Bromley, and N. Killoran, Estimating the gradient and higher-order derivatives on quantum hardware, Phys. Rev. A **103**, 012405 (2021).
[71] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 (2014).
[72] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, Adv. Quantum Technol. **2**, 1900070 (2019).
[73] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).
[74] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus, PRX Quantum **3**, 010313 (2022).
[75] Z. Webb, The Clifford group forms a unitary 3-design, arXiv preprint arXiv:1510.02769 (2015).
[76] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, Science **353**, 794 (2016).
[77] D. Gross, S. Nezami, and M. Walter, Schur-Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations, Commun. Math. Phys. **385**, 1325 (2021).
[78] N. Gisin and R. Thew, Quantum communication, Nat. Photonics **1**, 165 (2007).
[79] D. Gottesman, Theory of fault-tolerant quantum computation, Phys. Rev. A **57**, 127 (1998).
[80] R. A. Low, Learning and testing algorithms for the Clifford group, Phys. Rev. A **80**, 052314 (2009).
[81] K. P. Seshadreesan, M. Berta, and M. M. Wilde, Rényi squashed entanglement, discord, and relative entropy differences, J. Phys. A: Math. Theor. **48**, 395303 (2015).
[82] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1 (2021).
[83] C. O. Marrero, M. Kieferová, and N. Wiebe, Entanglement-Induced Barren Plateaus, PRX Quantum **2**, 040316 (2021).
[84] M. Kieferova, O. M. Carlos, and N. Wiebe, Quantum generative training using Renyi divergences, arXiv:2106.09567 (2021).
[85] S. Zhou, Z. Yang, A. Hamma, and C. Chamon, Single T gate in a Clifford circuit drives transition to universal entanglement spectrum statistics, SciPost Physics **9**, 087 (2020).
[86] S. True and A. Hamma, Transitions in entanglement complexity in random circuits, Quantum **6**, 818 (2022).
[87] T. Haug, Code for Bell magic, https://github.com/txhaug/bell-magic.
[88] S. F. E. Oliviero, L. Leone, A. Hamma, and S. Lloyd, Measuring magic on a quantum processor, arXiv: arXiv:2204.00015 (2022).
[89] J. Dehaene and B. De Moor, Clifford group, stabilizer states, and linear and quadratic operations over GF(2), Phys. Rev. A **68**, 042318 (2003).
[90] G. K. Brennen, An observable measure of entanglement for pure states of multi-qubit systems, arXiv preprint arXiv:quant-ph/0305094 (2003).
[91] I. Nechita, in *Annales Henri Poincaré* (Springer, Basel/Switzerland, 2007), Vol. 8, p. 1521.