Letter

Compressed quantum error mitigation

Maurits S. J. Tepaske and David J. Luitz [*]

Physikalisches Institut, Universität Bonn, Nußallee 12, 53115 Bonn, Germany

We introduce a quantum error mitigation technique based on probabilistic error cancellation to eliminate errors which have accumulated during the application of a quantum circuit. Our approach is based on applying an optimal "denoiser" after the action of a noisy circuit and can be performed with an arbitrary number of extra gates. The denoiser is given by an ensemble of circuits distributed with a quasiprobability distribution. For a simple noise model, we show that efficient, local denoisers can be found and we demonstrate their effectiveness for the digital quantum simulation of the time evolution of simple spin chains.

Introduction. Quantum information processing has been theoretically shown to hold great promise and quantum algorithms were developed which can in principle achieve an exponential speedup over their classical counterparts, both for general purpose computing [1–4] and quantum simulation [5–9]. However, present day quantum computing prototypes still suffer from significant noise processes which hinder the execution of many potentially groundbreaking quantum algorithms [10]. Nontrivial quantum algorithms typically require large sequences of quantum gates, each of which introduces dissipation and hence an overall loss of coherence, eventually rendering the results useless.

Until quantum error correction [11,12] becomes practical, *quantum error mitigation* seems to be more feasible to increase the accuracy of expectation values. Here the goal is to induce the (partial) cancellation of errors that stem from noisy quantum gates by extending the circuit corresponding to the desired algorithm with an ensemble of gates [13,14], sampled from a quasiprobability distribution.

The traditional way to accomplish this is with the gate-wise method from [13,14], where noise is mitigated by inverting the noise channel of each gate separately, i.e., the cancellation of errors is performed for each gate on its own. Here the local noise channel is approximated in a way such that it can be easily inverted analytically, e.g., using Pauli twirling [14]. Gates are then sampled from the inverted noise channel by interpreting it as a quasiprobability distribution. Because in this gatewise approach every noisy gate has to be modified separately, the sign problem is exponentially large in the number of gates, limiting the practicality of the mitigation. The success of the gatewise approach resulted in a large body of work concerning these methods [15–23], including extensions for simultaneous mitigation of multiple gates by Pauli-twirling entire layers [24,25] or variationally constructing a mitigating matrix product operator [26].

In principle, errors during the execution of a circuit can propagate and accumulate. These propagated errors can potentially blow up and lead to large errors for the circuit as a whole [27,28]. Here we introduce a mitigation technique that takes into account the propagation of errors, can be performed with a tunable number of extra gates, and works for non-Clifford local noise channels since the inversion of the accumulated global noise channel is implicit. We first execute the targeted noisy circuit completely, letting the noise propagate and accumulate, and only afterwards we apply an extra random circuit sampled from a quasiprobability distribution. We call the corresponding ensemble of random circuits a *denoiser* and we construct it such that upon averaging the accumulated errors cancel. Essentially, the denoiser inverts a global noise channel. Since we will construct it as a local brickwall circuit, following the classical preprocessing approach from [29], we call this *compressed* quantum error mitigation.

Method. Due to the inevitable coupling of a quantum processor to its environment, every qubit operation is affected by noise. Therefore, the simplest technique to minimize the impact of the resulting noise is to minimize the number of operations when performing a quantum algorithm. In [29] we showed that many-body time evolution operators can be efficiently compressed into brickwall circuits with high fidelity per gate.

In this Letter, we consider the noise explicitly by treating quantum operations as (generally nonunitary) quantum channels, corresponding to completely positive and trace preserving (CPTP) maps [30]. For example, instead of a noiseless two-qubit gate G, which acts on a quantum state $|\rho\rangle\rangle$ in superoperator form as $\mathcal{G}|\rho\rangle\rangle = G \otimes G^*|\rho\rangle\rangle$, we get the noisy channel $\tilde{\mathcal{G}} = \mathcal{N}\mathcal{G}$, where the noise channel $\mathcal{N}$ implements the two-qubit noise [31]. These channels are used to construct a "supercircuit" $\tilde{\mathcal{C}} = \prod_{i=1}^{N_G} \tilde{\mathcal{G}}_i$, consisting of N_G channels, which is affected by multiqubit accumulated noise. This supercircuit encodes an ensemble of circuits [31]. For simplicity, we assume that the noisy channels $\tilde{\mathcal{G}}_i$ in each half brickwall layer are lattice inversion and translation invariant, such that we can construct a denoiser with these properties, limiting the number of variational parameters.

[*]david.luitz@uni-bonn.de
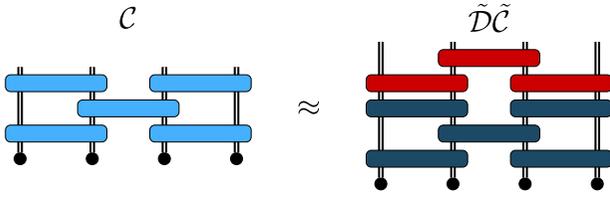
FIG. 1. Example of the quantum error mitigation procedure used in this work for the time evolution of the wave function of a spin chain. The ideal second-order Trotter supercircuit $\mathcal{C}$ of depth $M_{\text{trot}} = 1$ (light blue) is approximated by applying a denoiser $\tilde{\mathcal{D}}$ of depth $M = 1$ (red) to the noisy Trotter supercircuit $\tilde{\mathcal{C}}$ (dark blue). Because the denoiser is applied after fully executing the noisy Trotter supercircuit, it represents an approximate inverse of the global noise channel with a precision tunable by the depth of the denoiser.

The purpose of quantum error mitigation is to modify the ensemble of circuits described by $\tilde{\mathcal{C}}$ in a way that we can use it to obtain the noiseless expectation values. In superoperator language, we do this by following the supercircuit $\tilde{\mathcal{C}}$ with a denoiser supercircuit $\tilde{\mathcal{D}}$, such that $\tilde{\mathcal{D}}\tilde{\mathcal{C}}$ is as close to the noiseless supercircuit $\mathcal{C} = C \otimes C^*$ as possible. Here C is the target unitary circuit. Because the noise channel $\mathcal{N}$ is nonunitary, hence making the supercircuit $\tilde{\mathcal{C}}$ nonunitary, we need to use a nonunitary denoiser to retrieve the unitary $\mathcal{C}$.

We illustrate the mitigation procedure in Fig. 1, where a denoiser with one layer is used to mitigate errors for a second-order Trotter supercircuit with one layer. This circuit architecture is commonly used to simulate the time evolution of a quantum many-body system, until some time t, with controllable precision [29,32–42], and we will use it to benchmark the denoiser. In practice, we cannot directly implement a supercircuit, and so we have to utilize its interpretation as an ensemble of circuits. Essentially, after executing a shot of the noisy circuit we sample the denoiser and apply it. The goal is to construct the denoiser in a way that averaging over many of its samples cancels the accumulated errors and gives us a good approximation of the noiseless expectation values.

It should be noted that our approach requires more gate applications on the quantum processor than with the gate-wise scheme, since there each sample from the mitigation quasiprobability distribution can be absorbed into the original circuit, whereas our approach increases the circuit depth. We take this into account by imposing the same noise on the denoiser. Furthermore, within our scheme, the dimensionality of the quasiprobabilistic mitigating ensemble can be controlled, in contrast to the gatewise approach where it is equal to the gate count.

To facilitate the stochastic interpretation we parametrize each two-qubit denoiser channel $\mathcal{G}_i$ as a sum of CPTP maps, such that we can sample the terms in this sum and execute the sampled gate on the quantum processor. Concretely, we use a trace preserving sum of a unitary and a nonunitary channel. For the unitary part we take a two-qubit unitary channel $\mathcal{U}(\vec{\phi}_i) = U(\vec{\phi}_i) \otimes U^*(\vec{\phi}_i)$, with $U(\vec{\phi}_i)$ a two-qubit unitary gate parametrized by $\vec{\phi}_i$. For this we take the two-qubit ZZ rotation $\exp[-i\alpha(\sigma_z \otimes \sigma_z)]$ with angle α, which can be obtained from native gates on current hardware [43], and dress it with four general one-qubit unitaries, only two of which

are independent if we want a circuit that is space inversion symmetric around every bond. The resulting gate has seven real parameters $\vec{\phi}_i$.

For the nonunitary part, which is essential because $\tilde{\mathcal{D}}$ has to cancel the nonunitary accumulated noise to obtain the noiseless unitary circuit, we use a general one-qubit measurement followed by conditional preparation channel $\mathcal{M}(\vec{\zeta}_i)|\rho\rangle\rangle = \sum_l K_l \otimes K_l^* |\rho\rangle\rangle$. It has Kraus operators $K_1 = |\psi_1\rangle\langle\psi|$ and $K_2 = |\psi_2\rangle\langle\bar{\psi}|$ if we measure in the orthonormal basis $\{|\psi\rangle, |\bar{\psi}\rangle\}$, where $|\bar{\psi}\rangle$ is uniquely defined by $|\psi\rangle$ as they are antipodal points on the Bloch sphere. If the measurement yields $|\psi\rangle$ we prepare $|\psi_1\rangle$ and if we measure $|\bar{\psi}\rangle$ we prepare $|\psi_2\rangle$. These states can be arbitrary points on the Bloch sphere, i.e., $|\psi_1\rangle = V(\vec{\kappa}_1)|0\rangle$, $|\psi_2\rangle = V(\vec{\kappa}_2)|0\rangle$, and $|\psi\rangle = V(\vec{\kappa}_3)|0\rangle$, with V a general one-qubit unitary and each $\vec{\kappa}_i$ a three-dimensional vector, resulting in a real nine-dimensional $\vec{\zeta}_i$. This yields the two-qubit correlated measurement $\mathcal{M}(\vec{\zeta}_i) \otimes \mathcal{M}(\vec{\zeta}_i)$.

With these parts we construct the parametrization

$$\mathcal{G}_i = \eta_0 \mathcal{U}(\vec{\phi}_i) + \eta_1 \mathcal{M}(\vec{\zeta}_i) \otimes \mathcal{M}(\vec{\zeta}_i), \tag{1}$$

with coefficients $\eta_i \in \mathbb{R}$ that satisfy $\eta_0 + \eta_1 = 1$ because $\mathcal{G}_i$ is trace preserving. Note that here the tensor product symbol corresponds to combining two one-qubit channels to make a two-qubit channel, whereas in most of the paper it is used to link the column and row indices of a density matrix. We construct the denoiser from the noisy channels $\tilde{\mathcal{G}}_i = \mathcal{N}\mathcal{G}_i$. With this parametrization one denoiser channel has 17 independent real parameters, such that a denoiser of depth M, i.e., consisting of M brickwall layers, has $34M$ real parameters (we use one unique channel per half brickwall layer). For reference, a general channel has $544M$ parameters.

To determine the mitigated expectation values we use the full expression

$$\langle\hat{O}\rangle_{p=0} = \langle\langle\mathbb{1}|(\hat{O}\otimes\mathbb{1})\mathcal{C}|\rho_0\rangle\rangle \approx \langle\langle\mathbb{1}|(\hat{O}\otimes\mathbb{1})\tilde{\mathcal{D}}\tilde{\mathcal{C}}|\rho_0\rangle\rangle, \tag{2}$$

where $|\rho_0\rangle\rangle$ is the initial state and $|\mathbb{1}\rangle\rangle$ is the vectorized identity operator on the full Hilbert space. To evaluate this on a quantum processor, we use the stochastic interpretation of (1) to resample (2). In particular, from each channel (1) we get a unitary with probability $p_0 = |\eta_0|/\gamma$ and a measurement followed by conditional preparation with probability $p_1 = |\eta_1|/\gamma$. Here $\gamma = |\eta_0| + |\eta_1|$ is the sampling overhead, which characterizes the magnitude of the sign problem from negative η_i [13,14,18,20,44,45]. For quasiprobability distributions, i.e., with $\gamma > 1$, every denoiser sample has an extra sign $\text{sgn}(\eta) = \prod_{g=1}^{N_G} \text{sgn}(\eta_g)$, where $\text{sgn}(\eta_g)$ is the sign of the sampled coefficient of the gth channel. $\gamma = 1$ means that all signs are positive. Observables $\langle\hat{O}\rangle_{p=0}$ for the noiseless circuit are then approximated by resampling the observables from the denoiser ensemble [13]

$$\langle\hat{O}\rangle_{p=0} \approx \gamma\langle\text{sgn}(\eta)\hat{O}\rangle_p, \tag{3}$$

where $\gamma = \prod_{g=1}^{N_G} \gamma_g$ is the overall sampling overhead, with γ_g the overhead of the gth gate. Clearly, a large γ implies a large variance of $\langle\hat{O}\rangle_{p=0}$ for a given number of samples, with accurate estimation requiring the cancellation of large signed terms.

The number of samples required to resolve this cancellation of signs is bounded by Hoeffding's inequality, which states that a sufficient number of samples to estimate $\langle \hat{O} \rangle_{p=0}$ with error δ at probability $1 - \omega$ is bounded by $(2\gamma^2/\delta^2) \ln(2/\omega)$ [44]. Since γ scales exponentially in γ_g, it is clear that a denoiser with large M and $\gamma \gg 1$ will require many samples.

We observed that decompositions with $\gamma > 1$ are crucial for an accurate denoiser. Restricting to $\gamma = 1$ leads to large infidelity and no improvement upon increasing the number of terms in (1) or the depth M of the denoiser. Simply put, probabilistic error cancellation of gate noise introduces a sign problem and it is crucial to find optimal parametrizations (1) which minimize γ to make the approach scalable. This issue arises in all high performance error mitigation schemes [13,20,24,44], because the inverse of a physical noise channel is unphysical and cannot be represented as a positive sum over CPTP maps. This is clearly visible in the spectra of the denoiser, which lies outside the unit circle (cf. Fig. 4). This makes the tunability of the number of gates in each denoiser sample a crucial ingredient, which allows control over the sign problem, because we can freely choose the η_i in (1).

For the parametrization (1) of denoiser channels, we try to find a set of parameters for error mitigation by minimizing the normalized Frobenius distance between the noiseless and denoised supercircuits [29]

$$\epsilon = ||\mathcal{C} - \tilde{\mathcal{D}}\tilde{\mathcal{C}}||_F^2 / 4^L, \tag{4}$$

which bounds the distance of output density matrices and becomes zero for perfect denoising.

We carry out the minimization of ϵ on a classical processor, using gradient descent with the differential programming algorithm from [29]. Instead of explicitly calculating the accumulated global noise channel and subsequently inverting it, we approximate the noiseless supercircuit $\mathcal{C}$ with the denoised supercircuit $\tilde{\mathcal{D}}\tilde{\mathcal{C}}$, effectively yielding a circuit representation $\mathcal{D}$ of the inverse noise channel.

Results. To benchmark the denoiser we apply it to the second-order Trotter circuits of the spin-1/2 Heisenberg chain with periodic boundary conditions (PBC)

$$H = \sum_{i=1}^{L} \left(\sigma_1^i \sigma_1^{i+1} + \sigma_2^i \sigma_2^{i+1} + \sigma_3^i \sigma_3^{i+1} \right), \tag{5}$$

where $\sigma_\alpha^i = (\mathbb{1}^i, \sigma_x^i, \sigma_y^i, \sigma_z^i)$ is the Pauli algebra acting on the local Hilbert space of site i. A second-order Trotter circuit for evolution time t with depth M_{trot} consists of $M_{\text{trot}} - 1$ half brickwall layers with time step t/M_{trot} and two layers with half time step [29,34]. We consider circuits that are affected by uniform depolarizing noise with probability p for simplicity, but our approach can be used for any non-Clifford noise. The two-qubit noise channel is

$$\mathcal{N} = \left(1 - \frac{16p}{15} \right) \mathbb{1} + \frac{p}{15} \bigotimes_{j=i}^{i+1} \left(\sum_{\alpha=0}^{3} \sigma_\alpha^j \otimes \sigma_\alpha^{j*} \right), \tag{6}$$

which acts on neighboring qubits i and $i + 1$ and is applied to each Trotter and denoiser gate and $p = 0.01$ unless stated otherwise. We study circuits with depths $M_{\text{trot}} = 16, 32, 64$
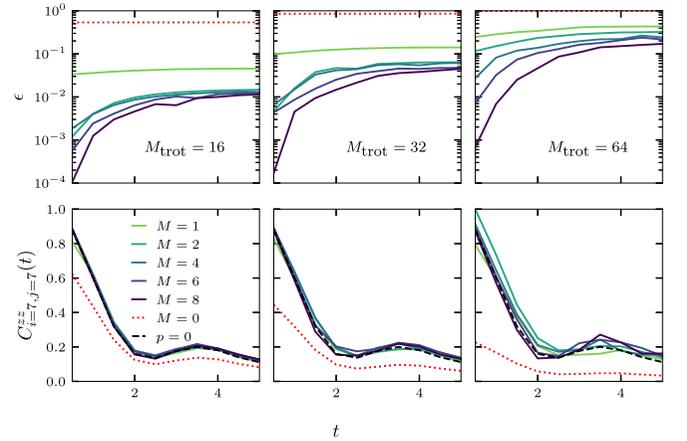
FIG. 2. Normalized distance ϵ between the denoised Trotter supercircuit $\tilde{\mathcal{D}}\tilde{\mathcal{C}}$ and the noiseless Trotter supercircuit $\mathcal{C}$ (top panels), at evolution times $t = 0.5, 1, \ldots, 5$, and the two-point z-spin correlator $C_{i=L/2, j=L/2}^{zz}(t)$ of a spin on the middle site at times 0 and t (bottom panels), for the infinite temperature initial state. We consider denoisers with depths $M = 1, 2, 4, 6, 8$ and second-order Trotter circuits with depths $M_{\text{trot}} = 16, 32, 64$. In the top panels we use a Heisenberg chain with $L = 8$ and in the bottom panels with $L = 14$, both with periodic boundary conditions. All gates are affected by two-qubit depolarizing noise with $p = 0.01$. The nondenoised results are labeled with $M = 0$ and the noiseless values with $p = 0$.

for evolution times $t = 0.5, 1, \ldots, 5$ and denoisers $\tilde{\mathcal{D}}$ with depths $M = 1, 2, 4, 6, 8$.

In the top panels of Fig. 2 we show ϵ (4) for a chain of size $L = 8$ as a function of time t. Here it can be seen that even for $M_{\text{trot}} = 32$ a denoiser with $M = 1$ already improves ϵ by roughly an order of magnitude at all considered t. Depending on M_{trot} and t, further increasing M lowers ϵ, with the biggest improvements occurring for high precision Trotter circuits with large depth $M_{\text{trot}} = 64$ and short time $t = 0.5$, where the Trotter gates are closer to the identity than in the other cases. At the other extreme, for $M_{\text{trot}} = 16$ the improvements are relatively small upon increasing $M > 2$. In all cases the denoiser works better at early times than at late times, again indicating that it is easier to denoise Trotter gates that are relatively close to the identity.

To probe the accuracy of the denoiser on quantities that do not enter the optimization, as a first test we consider the two-point correlator between spins at different times [46]

$$C_{ij}^{zz}(t) = \langle\langle \mathbb{1} | (\sigma_i^z \otimes \mathbb{1}) \tilde{\mathcal{D}}\tilde{\mathcal{C}}(t) (\sigma_j^z \otimes \mathbb{1}) | \mathbb{1} \rangle\rangle / 2^L, \tag{7}$$

where we have chosen the infinite temperature initial state and $\tilde{\mathcal{C}}(t)$ is the Trotter supercircuit for time t. In the bottom panels of Fig. 2 we show $C_{i=L/2, j=L/2}^{zz}(t)$ for the supercircuits from the upper panels, now for a $L = 14$ chain. Here we see that at $M_{\text{trot}} = 16$ we can retrieve the noiseless values already with $M = 1$, but that increasing M_{trot} makes this more difficult. At $M_{\text{trot}} = 64$ we see larger deviations and improvement upon increasing M is less stable, but nonetheless we are able to mitigate errors to a large extent.
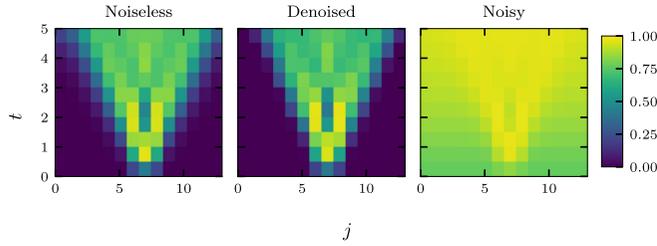
FIG. 3. Out-of-time-ordered correlator $C_{i=L/2,j}^{\text{otoc}}(t)$ as a function of the operator position j and time t, for the infinite temperature initial state, for a denoised second-order Trotter supercircuit with Trotter depth $M_{\text{trot}} = 32$ and denoiser depth $M = 2$. We consider evolution times $t = 0.5, 1, \ldots, 5$, for the periodic $L = 14$ Heisenberg chain that is affected by two-qubit depolarizing noise with $p = 0.01$.

As a further test, we compute the out-of-time-ordered correlator (OTOC) [29,39,47–50]

$$C_{ij}^{\text{otoc}}(t) = \text{Re}\langle\langle \mathbb{1} | (\sigma_j^{z\dagger} \otimes \mathbb{1}) \tilde{\mathcal{D}} \tilde{\mathcal{C}}(-t)$$
$$\times (\sigma_i^z \otimes \sigma_i^{z*}) \tilde{\mathcal{D}} \tilde{\mathcal{C}}(t) (\sigma_j^z \otimes \mathbb{1}) | \mathbb{1} \rangle\rangle / 2^L. \quad (8)$$

In Fig. 3 we show the results for $i = L/2$, for a Trotter circuit with depth $M_{\text{trot}} = 32$ and a denoiser with depth $M = 2$. Here we see that a denoiser with $M \ll M_{\text{trot}}$ is able to recover the light cone of correlations, which are otherwise buried by the noise. In the Supplemental Material [51] we consider how the denoiser performs at different noise levels p and how the denoised supercircuits perform under stacking. There we also calculate domain wall magnetization dynamics and show the distribution of the optimized denoiser parameters and the sampling overhead associated to the denoiser as a whole.

In Fig. 4 we show the eigenvalues of the noisy supercircuits for a noisy second-order Trotter supercircuit with $M_{\text{trot}} = 16$ at $t = 1$ (left), the corresponding optimized denoiser with $M = 4$ (center), and the denoised supercircuit (right). The eigenvalues λ of a unitary supercircuit lie on the unit circle and in the presence of dissipation they are pushed to the center. We see that the spectrum of the denoiser lies outside the unit circle, making it an unphysical channel which

cures the effect of the noise on the circuit, such that the spectrum of the denoised circuit is pushed back to the unit circle. The noiseless eigenvalues are shown as blue bars, making it clear that the denoiser is able to recover the noiseless eigenvalues from the noisy circuit. In the Supplemental Material [51] we show the spectra for a $p = 0.036$ denoiser, where we observe a clustering of eigenvalues reminiscent of Refs. [52–54]. There we also investigate the channel entropy of the various supercircuits [55,56].

Conclusion. We have introduced a probabilistic error cancellation scheme, where a classically determined denoiser mitigates the accumulated noise of a (generally non-Clifford) local noise channel. The required number of mitigation gates, i.e., the dimensionality of the corresponding quasiprobability distribution, is tunable and the parametrization of the corresponding channels provides control over the sign problem that is inherent to probabilistic error cancellation. We have shown that a denoiser with one layer can already significantly mitigate errors for second-order Trotter circuits with up to 64 layers.

This effectiveness of low-depth compressed circuits for denoising, in contrast with the noiseless time evolution operator compression from [29], can be understood from the nonunitarity of the denoiser channels. In particular, measurements can have nonlocal effects, since the measurement of a single qubit can reduce some highly entangled state (e.g., a GHZ state) to a product state, whereas in unitary circuits the spreading of correlations forms a light cone.

To optimize a denoiser with convenience at $L > 8$, the optimization can be formulated in terms of matrix product operators [26,29] or channels [16], which is convenient because the circuit calculations leading to the normalized distance ϵ and its gradient are easily formulated in terms of tensor contractions and singular value decompositions [29,57]. This provides one route to a practical denoiser, which is relevant because the targeted noiseless circuit and the accompanying noisy variant in (4) need to be simulated classically, confining the optimization procedure to limited system sizes with an exact treatment or limited entanglement with tensor networks. Nonetheless, we can use, e.g., matrix product operators to calculate (4) for some relatively small t, such that the noiseless and denoised supercircuits in (4) have relatively small entanglement, and then stack the final denoised supercircuit on a quantum processor to generate classically intractable states. Analogously, we can optimize the channels exactly at some classically tractable size and then execute them on a quantum processor with larger size. Both approaches are limited by the light cone of many-body correlations, as visualized in Fig. 3, because finite-size effects appear when the light-cone width becomes comparable with system size.

FIG. 4. Complex eigenvalues λ of the noisy second-order Trotter supercircuit with $M_{\text{trot}} = 16$ at time $t = 1$ (left), the corresponding optimized denoiser with $M = 4$ (center), and the denoised Trotter supercircuit (right). The Trotter circuit is for a $L = 6$ Heisenberg model with PBC and all two-qubit channels are affected by depolarizing noise with $p = 0.0046$. The unit circle, on which unitary eigenvalues must lie, is shown in black and the noiseless eigenvalues are shown as blue bars. It is evident that the denoiser recovers all the noiseless eigenvalues from the noisy circuit.

[1] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, Electron. Colloquium Comput. Complex. **TR96-003** (1996).

[2] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. **26**, 1484 (1997).

[3] S. Ebadi *et al.*, Quantum optimization of maximum independent set using Rydberg atom arrays, Science **376**, 1209 (2022).

[4] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[5] R. P. Feynman, Simulating physics with computers, Int. J. Theor. Phys. **21**, 467 (1982).

[6] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[7] B. Stefano, V. Filippo, and C. Giuseppe, An efficient quantum algorithm for the time evolution of parameterized circuits, Quantum **5**, 512 (2021).

[8] M. C. Bañuls *et al.*, Simulating lattice gauge theories within quantum technologies, Eur. Phys. J. D **74**, 165 (2020).

[9] P. Scholl *et al.*, Quantum simulation of 2d antiferromagnets with hundreds of Rydberg atoms, Nature (London) **595**, 233 (2021).

[10] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[11] A. R. Calderbank and P. W. Shor, Good quantum error-correcting codes exist, Phys. Rev. A **54**, 1098 (1996).

[12] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, Phys. Rev. A **52**, R2493 (1995).

[13] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[14] S. Endo, S. C. Benjamin, and Y. Li, Practical Quantum Error Mitigation for Near-Future Applications, Phys. Rev. X **8**, 031027 (2018).

[15] 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).

[16] S. Filippov, B. Sokolov, M. A. C. Rossi, J. Malmi, E.-M. Borrelli, D. Cavalcanti, S. Maniscalco, and G. García-Pérez, Matrix product channel: Variationally optimized quantum tensor network to mitigate noise and reduce errors for the variational quantum eigensolver, arXiv:2212.10225.

[17] N. Cao, J. Lin, D. Kribs, Y.-T. Poon, B. Zeng, and R. Laflamme, Nisq: Error correction, mitigation, and noise simulation, arXiv:2111.02345.

[18] C. Piveteau, D. Sutter, and S. Woerner, Quasiprobability decompositions with reduced sampling overhead, npj Quantum Inf. **8**, 12 (2022).

[19] M. Gutiérrez, C. Smith, L. Lulushi, S. Janardan, and K. R. Brown, Errors and pseudothresholds for incoherent and coherent noise, Phys. Rev. A **94**, 042338 (2016).

[20] J. Jiaqing, W. Kun, and W. Xin, Physical implementability of linear maps and its application in error mitigation, Quantum **5**, 600 (2021).

[21] E. Magesan, D. Puzzuoli, C. E. Granade, and D. G. Cory, Modeling quantum noise for efficient testing of fault-tolerant circuits, Phys. Rev. A **87**, 012324 (2013).

[22] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, arXiv:2210.00921.

[23] S. Ferracin, A. Hashim, J.-L. Ville, R. Naik, A. Carignan-Dugas, H. Qassim, A. Morvan, D. I. Santiago, I. Siddiqi, and J. J. Wallman, Efficiently improving the performance of noisy quantum computers, arXiv:2201.10672.

[24] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors, Nat. Phys. (2023), doi:10.1038/s41567-023-02042-2.

[25] B. McDonough, A. Mari, N. Shammah, N. T. Stemen, M. Wahl, W. J. Zeng, and P. P. Orth, Automated quantum error mitigation based on probabilistic error reduction, in *2022 IEEE/ACM Third International Workshop on Quantum Computing Software (QCS)* (IEEE, New York, 2022).

[26] Y. Guo and S. Yang, Quantum error mitigation via matrix product operators, PRX Quantum **3**, 040313 (2022).

[27] S. Flannigan, N. Pearson, G. H. Low, A. Buyskikh, I. Bloch, P. Zoller, M. Troyer, and A. J. Daley, Propagation of errors and quantitative quantum simulation with quantum advantage, Quantum Sci. Technol. **7**, 045025 (2022).

[28] P. M. Poggi, N. K. Lysne, K. W. Kuper, I. H. Deutsch, and P. S. Jessen, Quantifying the sensitivity to errors in analog quantum simulation, PRX Quantum **1**, 020308 (2020).

[29] M. S. J. Tepaske, D. Hahn, and D. J. Luitz, Optimal compression of quantum many-body time evolution operators into brickwall circuits, SciPost Phys. **14**, 073 (2023).

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

[31] D. Aharonov, A. Kitaev, and N. Nisan, Quantum Circuits with Mixed States, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98* (Association for Computing Machinery, New York, NY, 1998), pp. 20–30.

[32] H. F. Trotter, On the product of semi-groups of operators, Proc. Am. Math. Soc. **10**, 545 (1959).

[33] M. Suzuki, Generalized Trotter's formula and systematic approximants of exponential operators and inner derivations with applications to many-body problems, Commun. Math. Phys. **51**, 183 (1976).

[34] S. Paeckel, T. Köhler, A. Swoboda, S. R. Manmana, U. Schollwöck, and C. Hubig, Time-evolution methods for matrix-product states, Ann. Phys. (NY) **411**, 167998 (2019).

[35] J. Ostmeyer, Optimised Trotter decompositions for classical and quantum computing, arXiv:2211.02691.

[36] C. Kargi, J. P. Dehollain, F. Henriques, L. M. Sieberer, T. Olsacher, P. Hauke, M. Heyl, P. Zoller, and N. K. Langford, Quantum chaos and universal Trotterisation behaviours in digital quantum simulations, *Quantum Information and Measurement VI 2021* (Optica Publishing Group, 2021).

[37] M. Heyl, P. Hauke, and P. Zoller, Quantum localization bounds trotter errors in digital quantum simulation, Sci. Adv. **5**, eaau8342 (2019).

[38] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of Trotter Error with Commutator Scaling, Phys. Rev. X **11**, 011020 (2021).

[39] K. Hémery, F. Pollmann, and D. J. Luitz, Matrix product states approaches to operator spreading in ergodic quantum systems, Phys. Rev. B **100**, 104303 (2019).

[40] R. Mansuroglu, T. Eckstein, L. Nützel, S. A. Wilkinson, and M. J. Hartmann, Variational Hamiltonian simulation for trans-

lational invariant systems via classical pre-processing, Quantum Sci. Technol. **8**, 025006 (2023).

[41] H. Zhao, M. Bukov, M. Heyl, and R. Moessner, Making Trotterization adaptive for NISQ devices and beyond, arXiv:2209.12653.

[42] N. F. Berthusen, T. V. Trevisan, T. Iadecola, and P. P. Orth, Quantum dynamics simulations beyond the coherence time on NISQ hardware by variational Trotter compression, Phys. Rev. Res. **4**, 023097 (2022).

[43] I-C. Chen, B. Burdick, Y. Yao, P. P. Orth, and T. Iadecola, Error-mitigated simulation of quantum many-body scars on quantum computers with pulse-level control, Phys. Rev. Res. **4**, 043027 (2022).

[44] R. Takagi, Optimal resource cost for error mitigation, Phys. Rev. Res. **3**, 033178 (2021).

[45] Y. Guo and S. Yang, Noise effects on purity and quantum entanglement in terms of physical implementability, npj Quantum Inf. **9**, 11 (2023).

[46] M. Dupont and J. E. Moore, Universal spin dynamics in infinite-temperature one-dimensional quantum magnets, Phys. Rev. B **101**, 121106(R) (2020).

[47] Y.-L. Zhang, Y. Huang, and X. Chen, Information scrambling in chaotic systems with dissipation, Phys. Rev. B **99**, 014303 (2019).

[48] D. J. Luitz and Y. Bar Lev, Information propagation in isolated quantum systems, Phys. Rev. B **96**, 020406(R) (2017).

[49] J. Maldacena, S. H. Shenker, and D. Stanford, A bound on chaos, J. High Energy Phys. 08 (2016) 106.

[50] A. I. Larkin and Yu. N. Ovchinnikov, Quasiclassical method in the theory of superconductivity, Sov. JETP **28**, 1200 (1969).

[51] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.107.L201114 for the denoiser performance at various noise strengths; the denoiser spectra at high noise strength; the entropy of the various supercircuits that appeared in the main text; the performance of the denoised supercircuits upon stacking; histograms of the optimized denoiser parameters; the sampling overhead of the optimized denoisers; and simulations of domain wall dynamics as another test of the denoiser performance.

[52] K. Wang, F. Piazza, and D. J. Luitz, Hierarchy of Relaxation Timescales in Local Random Liouvillians, Phys. Rev. Lett. **124**, 100604 (2020).

[53] J. L. Li, D. C. Rose, J. P. Garrahan, and D. J. Luitz, Random matrix theory for quantum and classical metastability in local Liouvillians, Phys. Rev. B **105**, L180201 (2022).

[54] O. E. Sommer, F. Piazza, and D. J. Luitz, Many-body hierarchy of dissipative timescales in a quantum computer, Phys. Rev. Res. **3**, 023190 (2021).

[55] T. Zhou and D. J. Luitz, Operator entanglement entropy of the time evolution operator in chaotic systems, Phys. Rev. B **95**, 094206 (2017).

[56] W. Roga, Z. Puchała, Ł. Rudnicki, and K. Życzkowski, Entropic trade-off relations for quantum operations, Phys. Rev. A **87**, 032308 (2013).

[57] K. Noh, L. Jiang, and B. Fefferman, Efficient classical simulation of noisy random quantum circuits in one dimension, Quantum **4**, 318 (2020).