# Quasiprobabilistic approaches for quantum error mitigation and open system dynamics simulation

Workshop “Quantum Technologies for Fundamental Physics”, September 6, 2023, EMFCSC, Erice

**Peter P. Orth (Ames National Laboratory and Saarland University)**

Reference: McDonough et al., 2022 IEEE/ACM Third International Workshop on Quantum Computing Software (QCS), Dallas, TX, USA, 2022, pp. 83-93. See also: arXiv:2210.08611.

# Motivation and Take-Home Messages

- Current quantum hardware experiences significant levels of noise
  - How can we mitigate the effects of noise on quantum computations?
  - Can we exploit the presence of noise when simulating open (ie noisy) quantum systems?
- Quasiprobability methods are systematic approach of removing noise induced bias
  - Probabilistic Error Cancellation (PEC) [1] trades bias reduction for increased variance
  - Incurs exponentially large classical sampling overhead $\propto$ noise strength, circuit depth
  - Probabilistic Error Reduction (PER) [2] alleviates sampling overhead by reducing noise only partially $\rightarrow$ combined with extrapolation can yield results comparable to PEC
- Inherent device noise can act as resource in simulations of open quantum systems
  - Device noise can reduce the sampling overhead if somewhat close to what you want to simulate [3]
  - Only some type of noise acts as a resource (e.g. non-unital noise)

[1] Temme et al. (2017); Endo et al. (2017); [2] Mari et el. (2021); McDonough, PPO et al. (2022); [3] Aftergood, PPO et al. (to be submitted); Guimarães et al. (2023)

# Collaborators

Ben McDonough
(Yale; SQMS)

Andrea Mari
(Unitary Fund)

Misty Wahl
(Unitary Fund)

Nate Stemen
(Unitary Fund)

Nathan Shammah
(Unitary Fund)

Will Zeng
(Unitary Fund)

Joshua Aftergood
(Ames National Lab)

Thomas Iadecola
(Iowa State University & Ames National Lab)

# Quasiprobabilistic approach to quantum error mitigation

Noisy superoperator of quantum circuit $\mathcal{U}(\rho) = U\rho U^\dagger$

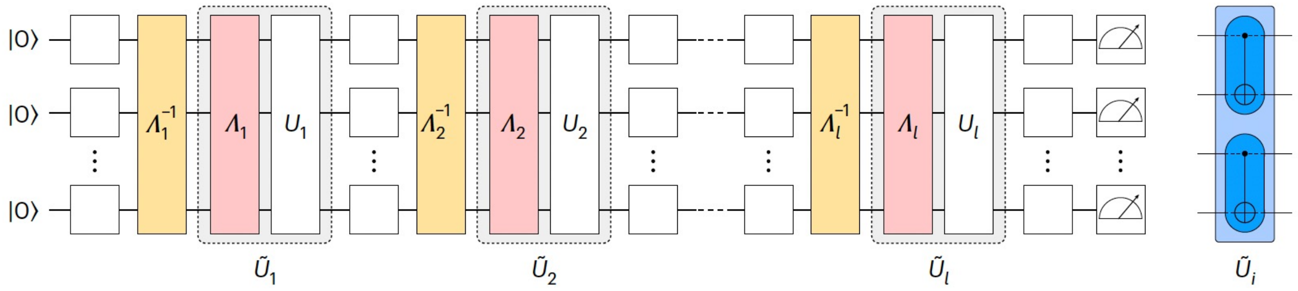
## Probabilistic Error Cancellation (PEC) [1]

$\tilde{\mathcal{U}} = \tilde{\mathcal{U}}_l \circ \ldots \circ \tilde{\mathcal{U}}_1$

$\tilde{\mathcal{U}}_i = \Lambda_i \circ \mathcal{U}_i$

- Removes noise induced bias on expectation value $\langle \mathcal{O} \rangle = \mathrm{Tr}[\mathcal{O}\tilde{\mathcal{U}}(\rho_0)]$
- Learn noise channels $\Lambda_i$ using noise tomography
  - Can be done efficiently for Pauli noise channels [2]: $\Lambda(\rho) = \sum_a c_a P_a \rho P_a^\dagger$
  - Use Gate Set Tomography for general noise models
- Apply channel inverse $\Lambda^{-1}$ (on average) using classical postprocessing
- Cancels physical noise and recovers the ideal noiseless result $\langle \mathcal{O} \rangle_{\mathrm{ideal}} = \mathrm{Tr}[\mathcal{O}\mathcal{U}(\rho_0)]$

**a** From van den Berg et al. (2023). **b**




[1] Temme et al. (2017); Endo et al. (2018); Cai et al. (2023); [2] van den Berg et al. (2023); Flammia, Wallmann (2019)

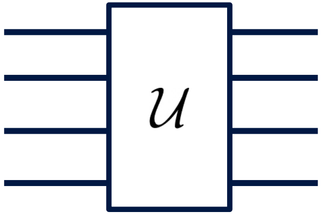
# Ideal circuit evaluation of expectation value

- Applicable to algorithms in which the figure of merit is an expectation value averaged over many shots of a unitary circuit

Ideal circuit

Input state is prepared

$\rho_0$




One or more expectation values are obtained

$$\langle \mathcal{O} \rangle_{\text{ideal}} = \mathrm{Tr}[\mathcal{O}\mathcal{U}(\rho_0)]$$

A unitary circuit is run

$$\mathcal{U}(\rho_0) = (\mathcal{U}_l \circ \ldots \circ \mathcal{U}_1)(\rho_0)$$

[1] Temme et al. (2017); Endo et al. (2018); [2] van den Berg et al. (2023); Cai et al. (2023).

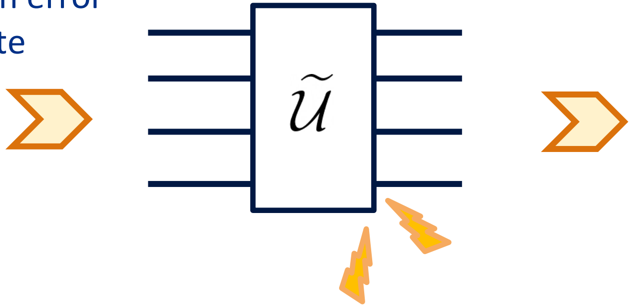
# Noisy expectation value from noisy quantum circuit

- Noise introduces bias into the estimator of this expectation value.

Noisy circuit

State preparation error affects input state

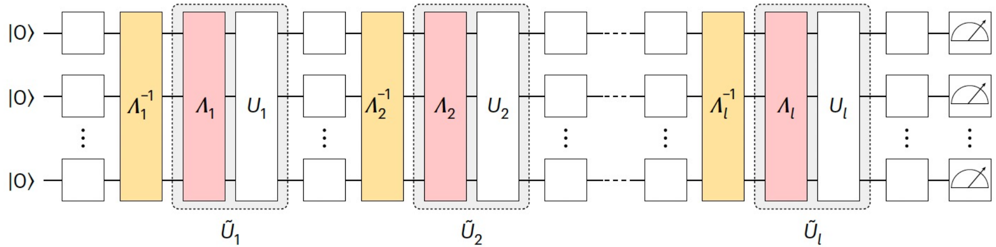
$\tilde{\rho}_0$




Measurement error also adds to noise in final expectation value

$$\langle \mathcal{O} \rangle = \mathrm{Tr}[\tilde{\mathcal{O}}\tilde{\mathcal{U}}(\rho_0)]$$

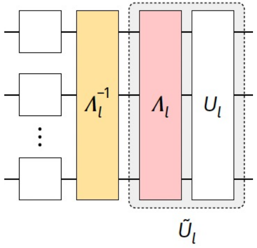
Circuit is non-unitary $\tilde{\mathcal{U}} = \tilde{\mathcal{U}}_l \circ \ldots \circ \tilde{\mathcal{U}}_1$ with $\tilde{\mathcal{U}}_i = \Lambda_i \circ \mathcal{U}_i$




From van den Berg et al. (2023).

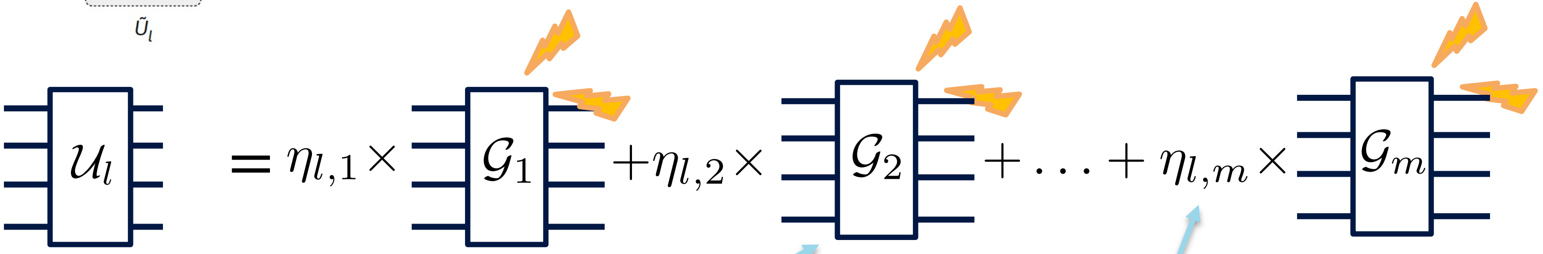
# Ideal expectation value as linear combination of noisy circuits

- After characterizing the noisy operations $\{G_\alpha\}$, the ideal circuit is decomposed into $\mathcal{U}(\rho) = \sum_\alpha \eta_\alpha G_\alpha(\rho)$ with real coefficients $\eta_\alpha$.
- Linearity of the expectation value allows writing the ideal value as $\langle O \rangle_{ideal} = \sum_\alpha \eta_\alpha \langle O \rangle_\alpha$




$$\boxed{\mathcal{U}_l = \Lambda_l^{-1} \circ \tilde{\mathcal{U}}_l = \sum_k \eta_{l,k} \mathcal{G}_k}$$

Inverse channel is not physical.
Thus implemented on average by classical postprocessing.




Ideal (noiseless) circuit layer

Noisy circuits

Real, possibly negative coefficients



SQMS SUPERCONDUCTING QUANTUM MATERIALS & SYSTEMS CENTER

# Exponential sampling overhead due to negativity

- Number of terms grows exponentially with circuit depth $l$

$$\mathrm{Tr}[\mathcal{O}(\mathcal{U}_l \circ \ldots \circ \mathcal{U}_1)(\rho_0)] = \sum_{k_1,\ldots,k_l} \eta_{1,k_1} \cdots \eta_{l,k_l} \mathrm{Tr}[\mathcal{O}(G_{l,k_l} \circ \ldots \circ G_{1,k_1})(\rho_0)]$$

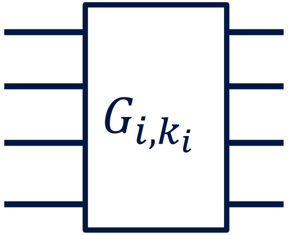
- The linear combination is converted to a quasiprobability distribution (QPD) and sampled

$$\mathrm{Tr}[\mathcal{O}(\mathcal{U}_l \circ \ldots \circ \mathcal{U}_1)(\rho_0)] = \gamma_1 \cdots \gamma_l \sum_{k_1,\ldots,k_l} \mathrm{sgn}(\eta_{1,k_1}) \cdots \mathrm{sgn}(\eta_{l,k_l}) \frac{|\eta_{1,k_1}|}{\gamma_1} \cdots \frac{|\eta_{l,k_l}|}{\gamma_l} \mathrm{Tr}[\mathcal{O}(G_{l,k_l} \circ \ldots \circ G_{1,k_1})(\rho_0)]$$

Increased range ☞ increased variance $\Delta\mathcal{O}^2 \propto \gamma^2$

Obtain via Monte Carlo sampling

Sample




with probability $\frac{|\eta_{i,k_i}|}{\gamma_i}$

Then, scale expectation value by

$\gamma_i \,\mathrm{sgn}(\eta_{i,k_i})$ with $\gamma_i = \sum_{k_i} |\eta_{i,k_i}| \geq 1$

Sampling overhead due to negative volume of QPD ∝ noise strength

# Example: Pauli noise

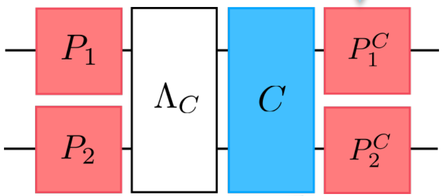
- Transform hardware noise to Pauli noise by Pauli twirling $\Lambda_{\mathcal{P}} = \frac{1}{|\mathcal{P}|}\sum_a \mathcal{P}_a \Lambda \mathcal{P}_a$
  - Works only for layers composed of Clifford gates
  - Assumes dominant noise source are CNOT gates

$$\sum_a (\mathcal{P}_a^C \mathcal{C} \Lambda \mathcal{P}_a)(\rho) = \sum_{a,\nu} \underbrace{P_a^C C}_{= C P_a} M_\nu P_a \rho P_a^\dagger M_\nu^\dagger C^\dagger (P_a^C)^\dagger = (\Lambda_{\mathcal{P}} \mathcal{C})(\rho)$$

Randomly sampled Pauli gates

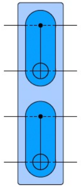
Conjugate Pauli operators

$$C P_a C^\dagger = P_a^C$$




Noise associated with layer

Clifford layer



Dressing noise by random Paulis diagonalizes $\Lambda$ in the Pauli basis

[1] van den Berg et al. (2023); Flammia, Wallmann (2019)

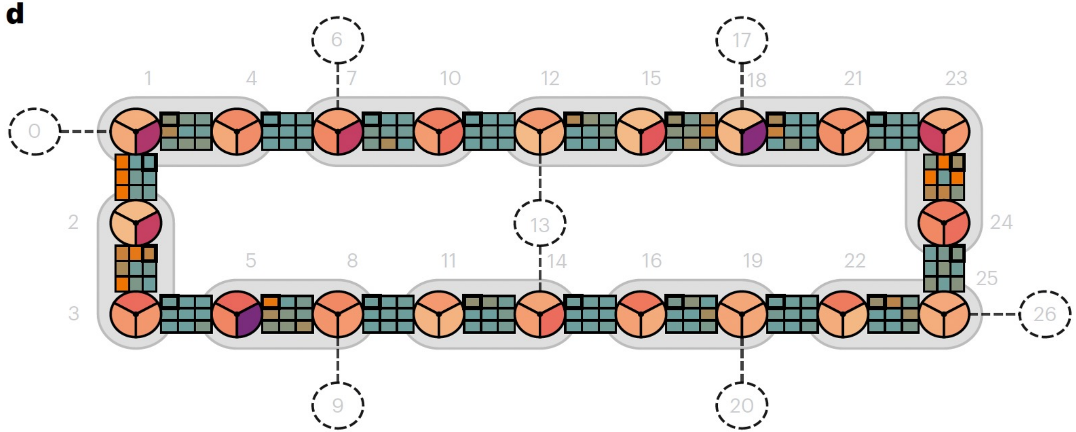
# Sparse Pauli-Lindblad noise model

- Sparse Pauli noise model parameters $\lambda_k$ of a given layer can be efficiently learned [1,2]

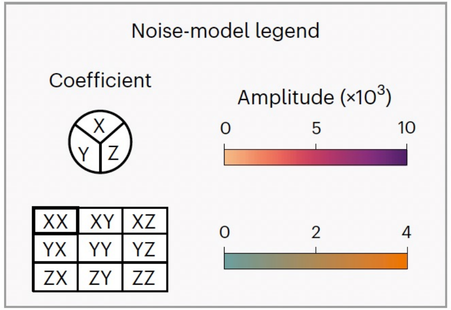
Lindbladian $$\mathcal{L}(\rho) = \sum_k \lambda_k (P_k \rho P_k - \rho)$$ [1] van den Berg et al. (2023); Flammia, Wallmann (2019)

Pauli noise channel $$\Lambda(\rho) = e^{\mathcal{L}}(\rho) = \prod_k [w_k(\cdot) + (1 - w_k) P_k(\cdot) P_k](\rho)$$

Noise coefficients $$w_k = \frac{1}{2}(1 + e^{-2\lambda_k}) \approx 1 - 2\lambda_k$$




From van den Berg et al. (2023)



# Sparse Pauli-Lindblad noise model

- Sparse Pauli noise model parameters $\lambda_k$ of a given layer $i$ can be efficiently learned [1,2]

Lindbladian $$\mathcal{L}(\rho) = \sum_k \lambda_k (P_k \rho P_k - \rho)$$

Pauli noise channel $$\Lambda(\rho) = e^{\mathcal{L}}(\rho) = \prod_k [w_k(\cdot) + (1 - w_k) P_k (\cdot) P_k](\rho)$$

$$\gamma_i = e^{2\sum_k \lambda_k} \geq 1$$

Inverse map $$\Lambda^{-1}(\rho) = e^{-\mathcal{L}}(\rho) = \gamma_i \prod_k [w_k(\cdot) - (1 - w_k) P_k (\cdot) P_k](\rho)$$

[1] van den Berg et al. (2023); Flammia, Wallmann (2019)

# PEC at work: Trotterized time evolution

McDonough, PPO et al, IEEE (2022).

**Algorithm 1** Description of PER routine

**Input:** Circuit with layers $l \in \{1,..,l_{tot}\}$, each with noise model parameters $\{w_{l1}^{(\xi)},...,w_{ln}^{(\xi)}\}$
**Output:** A sample of the PER expectation value (before readout error mitigation)

```
1:  Let α ≡ 1
2:  for l ∈ {1, ..., l_tot} do
3:      Compose layer l into circuit
4:      for k ∈ {1, ..., n} do
5:          Sample I with probability w_lk^(ξ) and P_k otherwise
6:          Multiply α by γ_l^(ξ)
7:          if P_k was sampled then
8:              Multiply α by −1
9:          end if
10:         Compose sampled operator into circuit
11:     end for
12: end for
13: Run the circuit and get the expectation value
14: if ξ < 1 then
15:     Scale result by α
16: end if
```

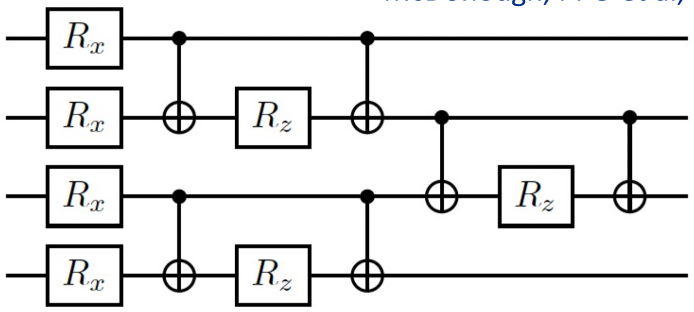
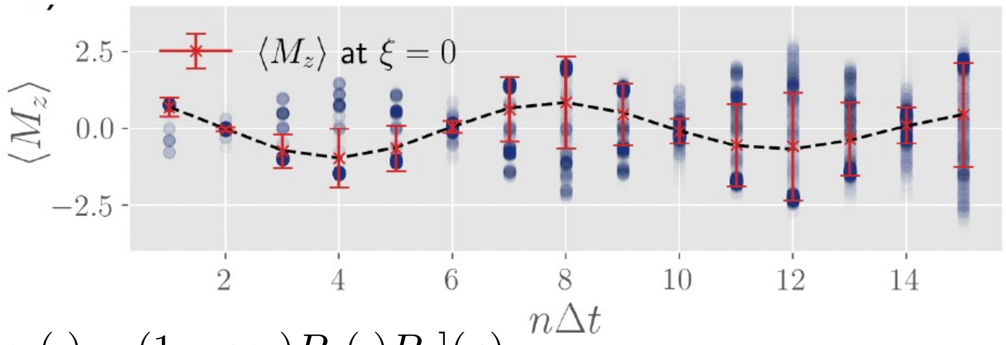



Fig. 7: The realization of a single Trotter step as a quantum circuit. Here we have defined $R_x \equiv RX(-2h\delta t)$ and $R_z \equiv RZ(2J\delta t)$




Inverse map $\Lambda^{-1}(\rho) = e^{-\mathcal{L}}(\rho) = \gamma_i \prod_k [w_k(\cdot) - (1-w_k)P_k(\cdot)P_k](\rho)$



SUPERCONDUCTING QUANTUM MATERIALS & SYSTEMS CENTER

# Probabilistic Error Reduction (PER)

Mari et al. (2021); McDonough, PPO et al, IEEE (2022).

- PEC sampling overhead increases exponentially with noise strength x depth
- Remove noise only partially to reduce overhead

$$\Lambda^{(\xi)} = \gamma^{(\xi)} \prod_k \left( w_k^{(\xi)} \mathcal{I} + \mathrm{sgn}(\xi - 1)(1 - w_k^{(\xi)}) \mathcal{P}_k \right). \quad (11)$$

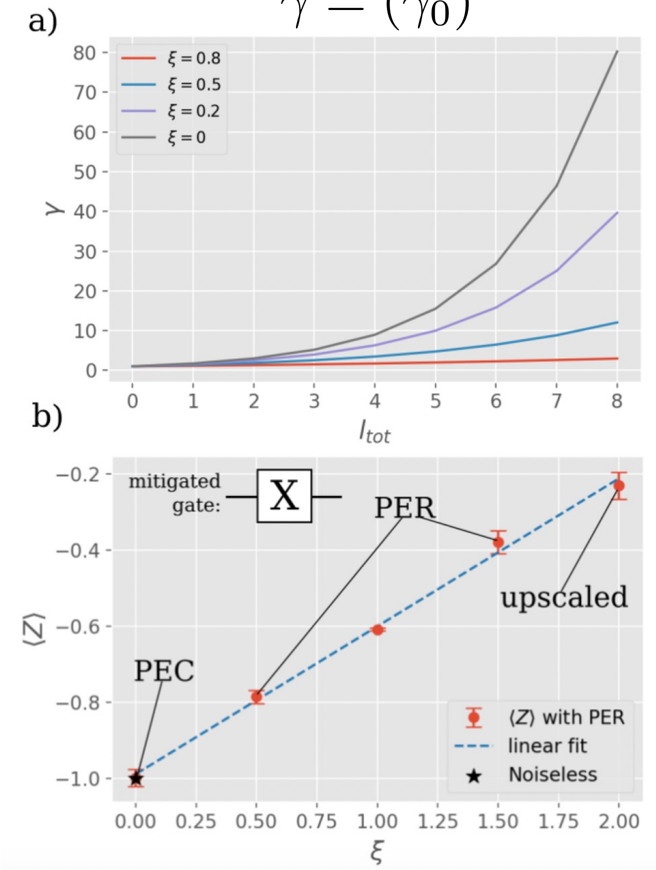
Here, $w^{(\xi)} \equiv \frac{1}{2}(1 + e^{-2|1-\xi|\lambda_k})$ and the sampling overhead

$$\gamma^{(\xi)} = \begin{cases} \exp[2(1-\xi)\sum_k \lambda_k] & \xi < 1 \\ 1 & \xi \geq 1 \end{cases} \quad (12)$$

- Combine with virtual zero-noise extrapolation (vZNE) to recover the ideal result

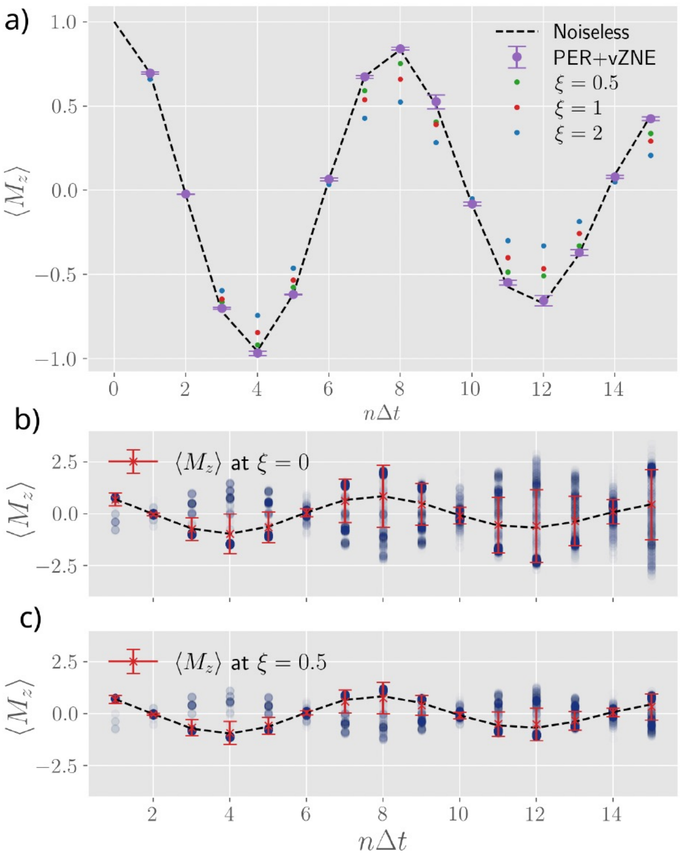
Example: error mitigation on single X gate on Rigetti hardware. Obtained noise model using Gate Set Tomography (GST): $\gamma_0 = 1.73$

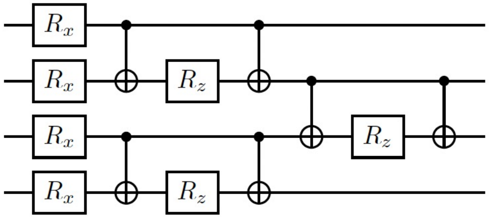
$$\gamma = (\gamma_0)^l$$

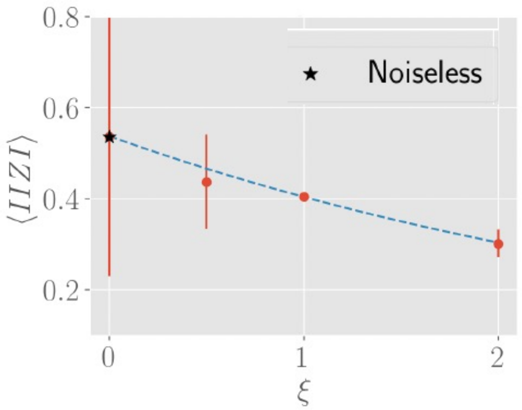


# PER + vZNE reduces sampling overhead

McDonough, PPO et al, IEEE (2022).




Scaling up noise using this method has been used in recent IBM “quantum utility” work: Kim et al. Nature (2023).

# Open source software for automated QEM

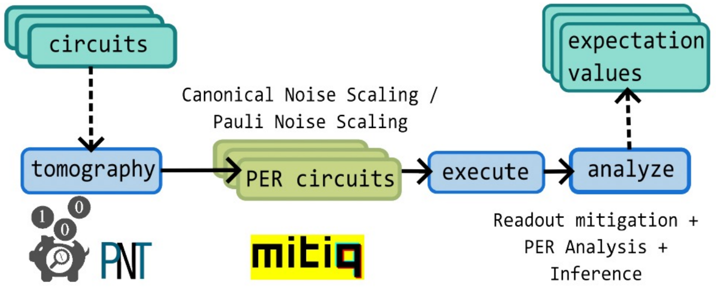
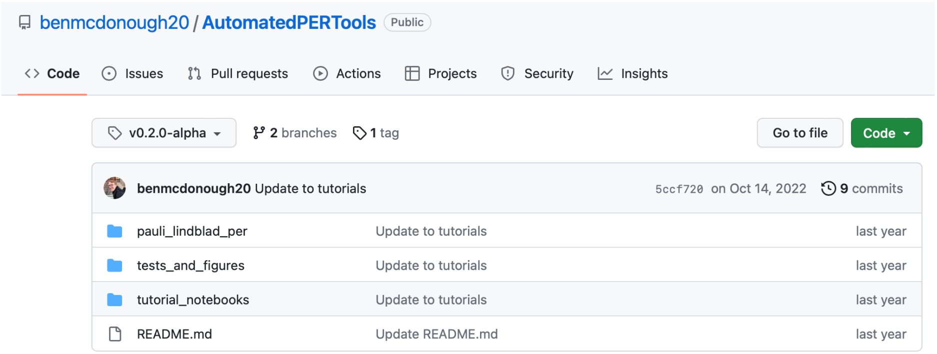



Fig. 1: Illustration of automated error mitigation protocol starting from user defined circuits and returning noise-mitigated expectation values. It includes a noise tomography step involving PNT or GST, whose results are used to generate sampled PER circuits via canonical or Pauli noise scaling.




- Open source software package that implements Pauli noise tomography and PER + vZNE quantum error mitigation
  - https://github.com/benmcdonough20/AutomatedPERTools
- Similar method has been implemented by IBM in Qiskit, but source code is not openly accessible

# QPD noise scaling for open system dynamics simulations

- Can arbitrarily scale different Pauli noise parameters $\lambda_k$
- Go from a given (device) Pauli noise model $\{\lambda_k\}$ to another one $\{\phi_k\}$:

$$\boxed{\Lambda_\phi = \Lambda^{-1}_{\lambda\to\phi}\Lambda_\lambda}$$

"Partial" inverse $\Lambda^{-1}_{\lambda\to\phi} = \gamma \prod_{k\in\mathcal{K}} [\omega_k \mathcal{I} - \mathrm{sgn}(\lambda_k - \phi_k)(1-\omega_k)\mathcal{P}_k]$

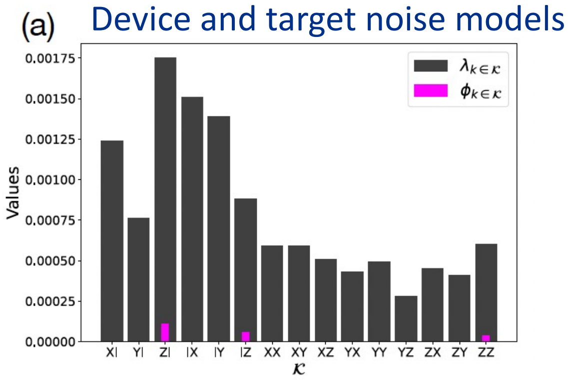
with coefficients $\omega_k = \frac{1}{2}(1 + e^{-2|\lambda_k - \phi_k|})$

Sampling overhead $\gamma = \exp(\sum_{k|\lambda_k>\phi_k} 2|\lambda_k - \phi_k|)$

Scaling up Pauli noise does not cost overhead!

# Example: noisy transverse-field Ising model dynamics

- Simulate dynamics of transverse-field Ising model (TFIM) with Pauli noise




Hamiltonian $H = J_z \sum_{i=1}^{L-1} Z_i Z_{i+1} + h_x \sum_{i=1}^{L} X_i$

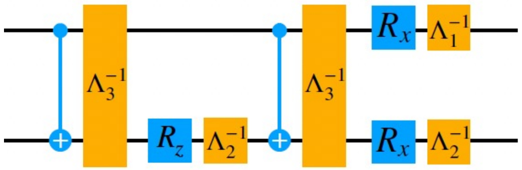
Dissipative dynamics described by Lindblad equation of motion

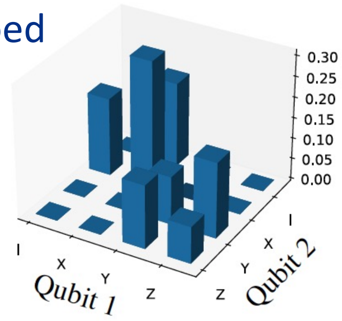
$$\dot{\rho}(t) = -i[H, \rho(t)] + \hat{\mathcal{L}}[\rho(t)]$$

Desired Pauli noise described by model parameters $\{\phi_k\}$

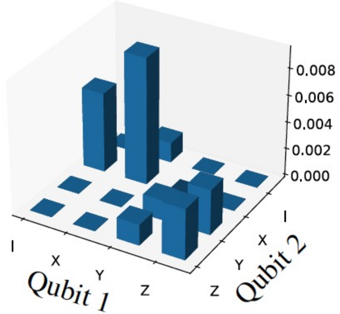
Assume hardware experiences Pauli noise described by parameters $\{\lambda_k\}$ (after Pauli twirling)

$$\boxed{\Lambda^{-1}_{\lambda_k \to \phi_k}}$$




Difference of Pauli expectation values after 10 Trotter steps




After QPD sampling, the difference is much reduced



SQMS SUPERCONDUCTING QUANTUM MATERIALS & SYSTEMS CENTER

# Simulating dynamics with general 1-qubit noise models

- Most general single qubit noise model described by Kraus operators [1]

$$K_0 = U \begin{pmatrix} s_0 & 0 \\ 0 & s_1 \end{pmatrix} V^\dagger$$

$$K_1 = U \begin{pmatrix} 0 & \sqrt{1-s_1^2} \\ \sqrt{1-s_0^2} & 0 \end{pmatrix} V^\dagger$$

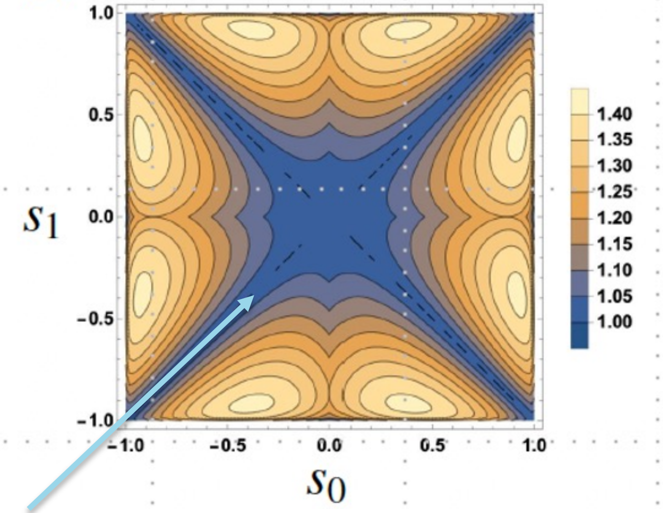
- Cannot express nonunital noise by unitary gates only [2]
- Include RESET gate $|\psi\rangle \xrightarrow{\text{RESET}} |0\rangle$ into noisy gate set $\{G_k\} \equiv \{B_k\}$
  - We use the IBM native gateset $\{I, X, \sqrt{X}, R_z(\phi), \text{RESET}\}$.
- Construct noise channel from IBM basis gates using QPD approach $\Lambda = \sum_i \eta_i \mathcal{B}_i$

[1] Verstraete (2000; [2] Temme et al. (2017); Endo et al. (2018)

# Construct general 1-qubit channel using QPD approach

- We compare the sampling overhead for constructing general 1-qubit channel using noiseless and noisy basis sets
- Exact solution for noiseless basis set: $\Lambda = \sum_i \eta_i \mathcal{B}_i$

$$
\begin{aligned}
\eta_I &= \eta_0 \\
\eta_X &= \eta_0 + \frac{1}{2}\left(1 - s_0^2 - s_0 s_1 - s_1^2 \sqrt{\left(s_0^2 - 1\right)\left(s_1^2 - 1\right)}\right) \\
\eta_Z &= \eta_0 - s_0 s_1 \\
\eta_R &= s_0(s_0 + s_1) - 2\eta_0 \\
\eta_{XZ} &= \eta_0 - \frac{1}{2}\left(1 - s_0^2 - s_0 s_1 - s_1^2 - \sqrt{\left(s_0^2 - 1\right)\left(s1^2 - 1\right)}\right) \\
\eta_{XR} &= s_1(s_0 + s_1) - 2\eta_0\,,
\end{aligned}
$$

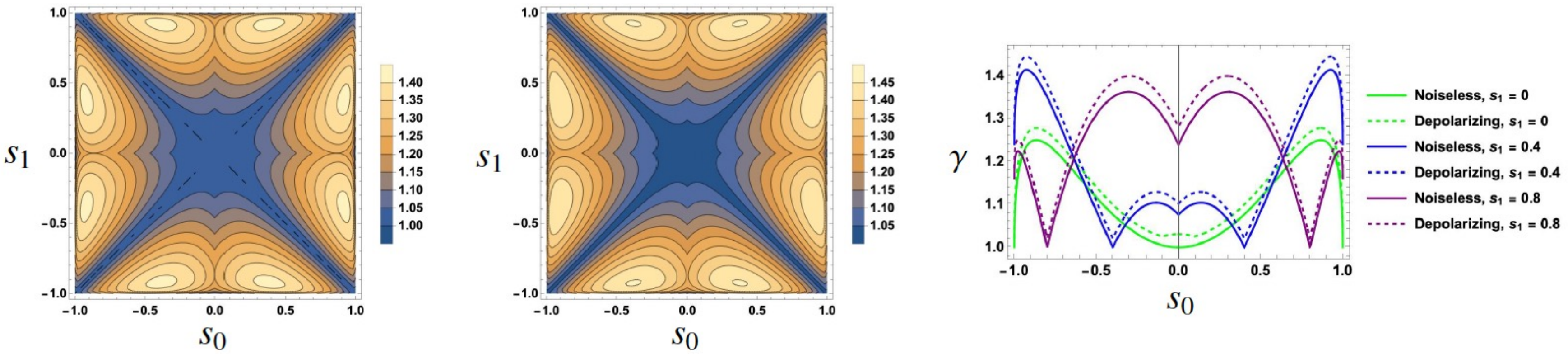



Pauli noise: no overhead



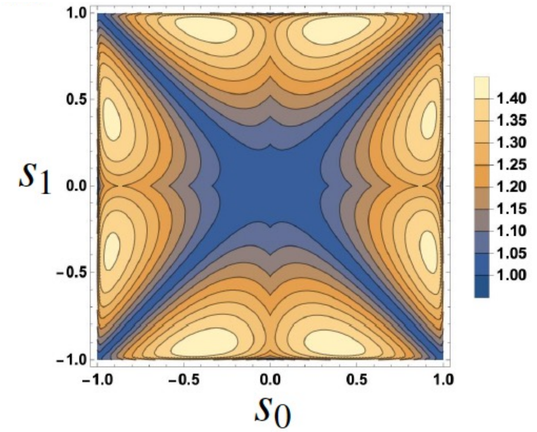
SUPERCONDUCTING QUANTUM MATERIALS & SYSTEMS CENTER

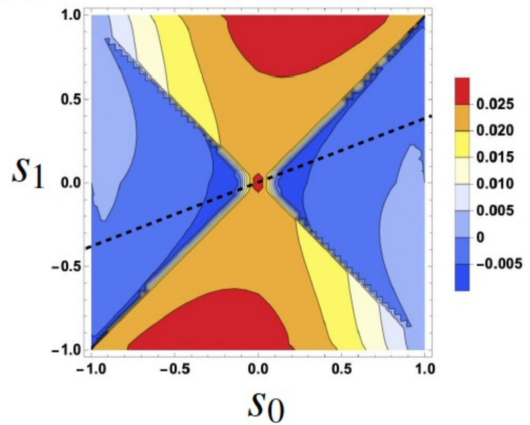
# Comparing noiseless and depolarizing noisy gate sets

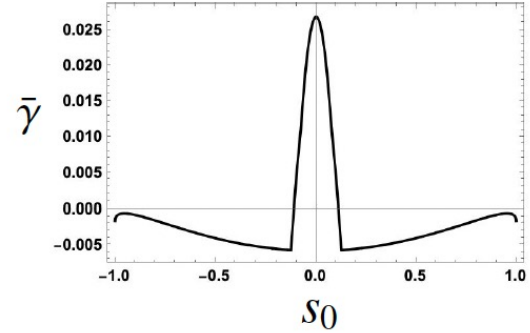



- Gate set experiencing depolarizing noise (symmetric Pauli noise) has always larger overhead
  - Depolarizing channel $\Lambda_p = (1-3p)\mathcal{B}_I + p\mathcal{B}_X + p\mathcal{B}_Y + p\mathcal{B}_Z$
  - We use noise strength $p = 10^{-3}$ here
- Pauli noise does not act as a resource
- Increasing Pauli noise is “free”, removing it costs nonzero overhead

# Nonunital device noise is a resource



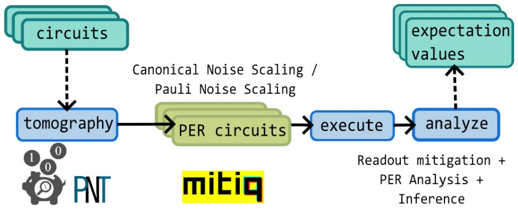



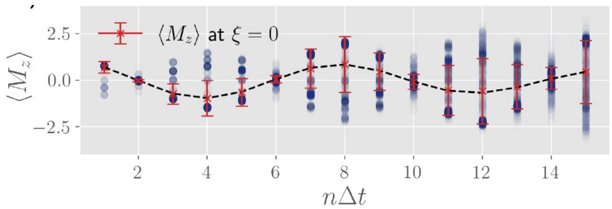



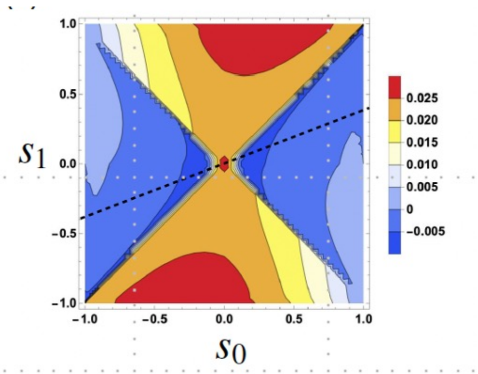

- Gate set with amplitude + phase damping, taken from FakeArmonkV2 backend
- Slightly boost noise strength to be comparable to $p = 10^{-3}$ depolarizing channel
- Nonunital noise reduces overhead compared to noiseless case
  - ☞ Noise acts as a resource
  - May seem small quantitatively for a single application of channel, but it exponentially increases with gate depth

# Summary

- Quasiprobability methods can scale effective noise affecting expectation values
  - Systematic scaling of noise that requires noise tomography
  - Sampling overhead increases exponentially with noise strength and circuit depth
  - Probabilistic Error Reduction combined with zero-noise extrapolation reduces overhead
- QPD scaling of noise can be used for open system dynamics simulations
  - Nonunital noise acts as a resource by reducing overhead compared to noiseless gate set



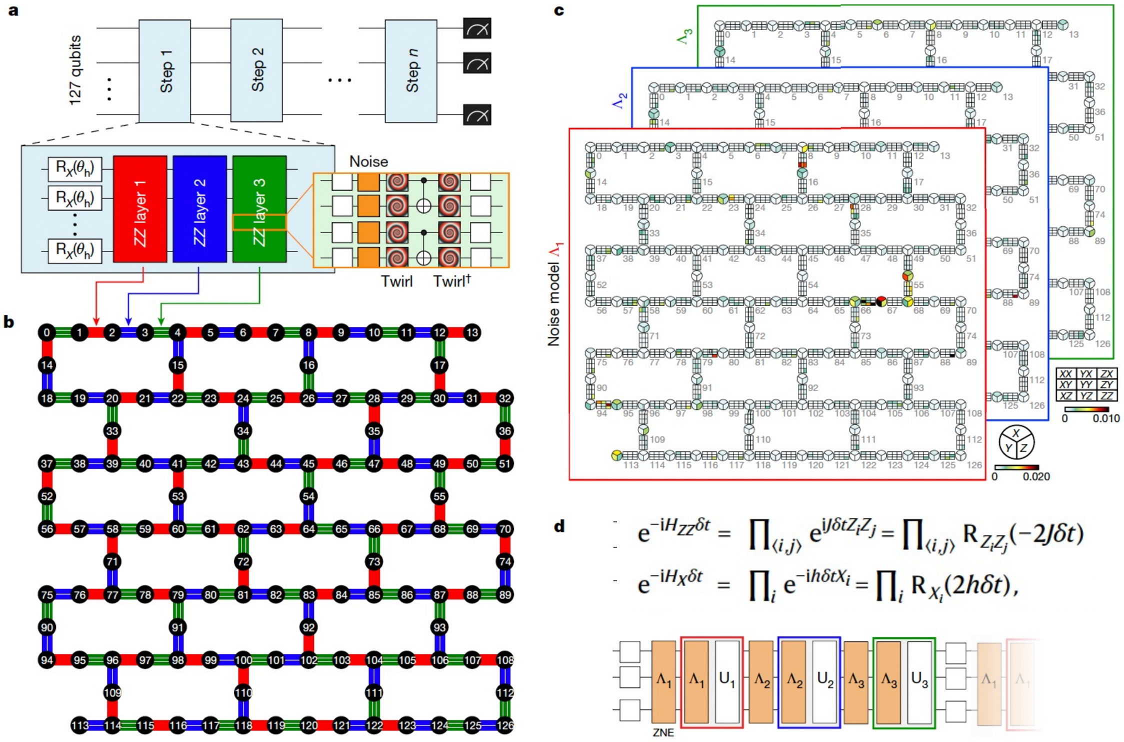

References:


- McDonough et al., 2022 IEEE/ACM Third International Workshop on Quantum Computing Software (QCS), Dallas, TX, USA, 2022, pp. 83-93. See also: arXiv:2210.08611 (2022).
- Aftergood et al. (in preparation).


**Thanks for your attention!**

# Backup Slides

# Trotter dynamics of 127 qubit transverse-field Ising model



- Trotter circuit contains three layers
- Pauli twirling transforms the noise to Pauli noise
- Efficient noise tomography using a sparse Pauli noise model ansatz
- Can precisely tune the noise for ZNE since noise is well characterized (probabilistic noise amplification)

# Trotter dynamics of 127 qubit transverse-field Ising model

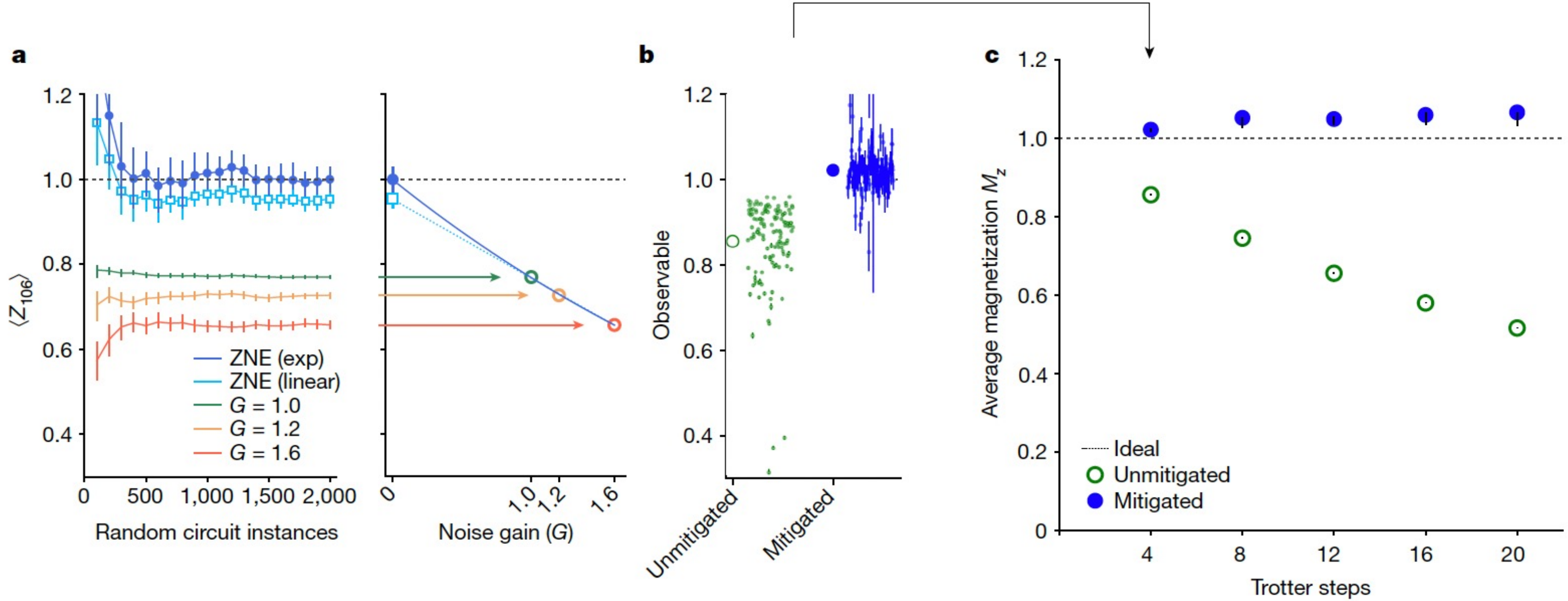



**Fig. 2 | Zero-noise extrapolation with probabilistic error amplification.** Mitigated expectation values from Trotter circuits at the Clifford condition $\theta_h = 0$. **a**, Convergence of unmitigated ($G = 1$), noise-amplified ($G > 1$) and noise-mitigated (ZNE) estimates of $\langle Z_{106} \rangle$ after four Trotter steps. In all panels, error bars indicate 68% confidence intervals obtained by means of percentile bootstrap. Exponential extrapolation (exp, dark blue) tends to outperform linear extrapolation (linear, light blue) when differences between the converged estimates of $\langle Z_{106} \rangle_{G \neq 0}$ are well resolved. **b**, Magnetization (large markers) is computed as the mean of the individual estimates of $\langle Z_q \rangle$ for all qubits (small markers). **c**, As circuit depth is increased, unmitigated estimates of $M_z$ decay monotonically from the ideal value of 1. ZNE greatly improves the estimates even after 20 Trotter steps (see Supplementary Information II for ZNE details).

# Classically verifiable regime

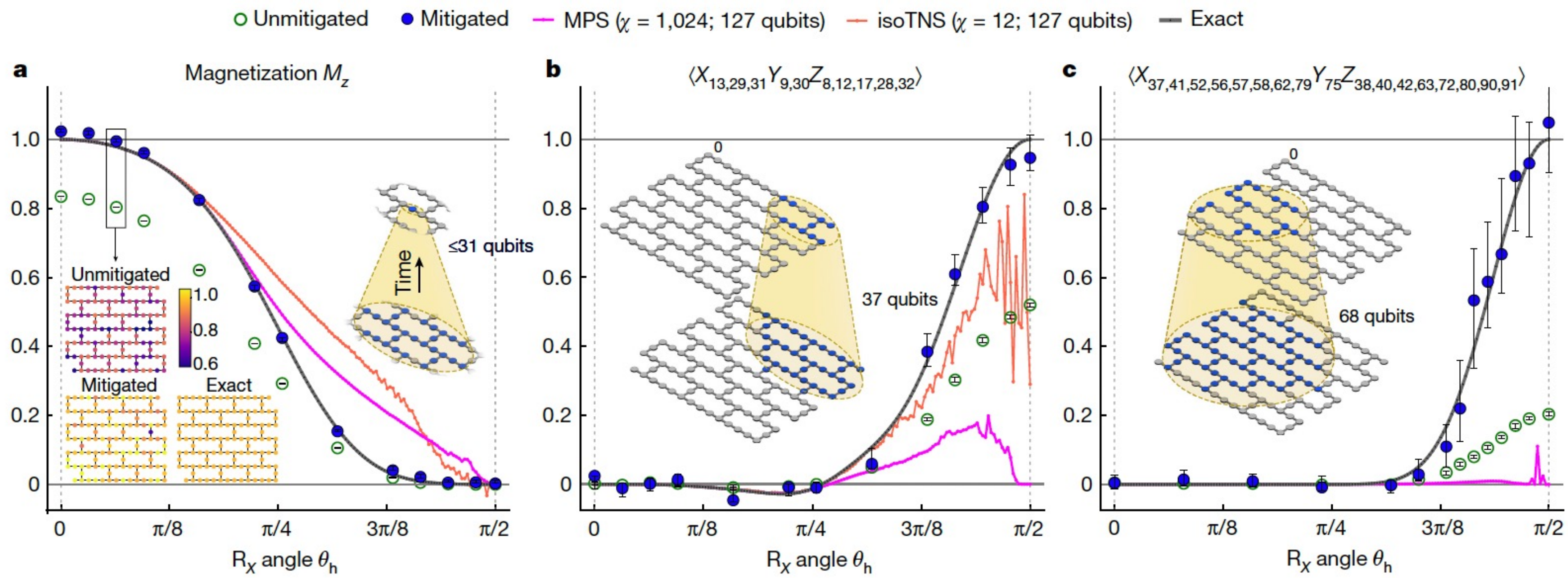



**Fig. 3 | Classically verifiable expectation values from 127-qubit, depth-15 Clifford and non-Clifford circuits.** Expectation value estimates for $\theta_h$ sweeps

Upper insets in all panels illustrate causal light cones, indicating in blue the final qubits measured (top) and the nominal set of initial qubits that can

# Classically “challenging” regime

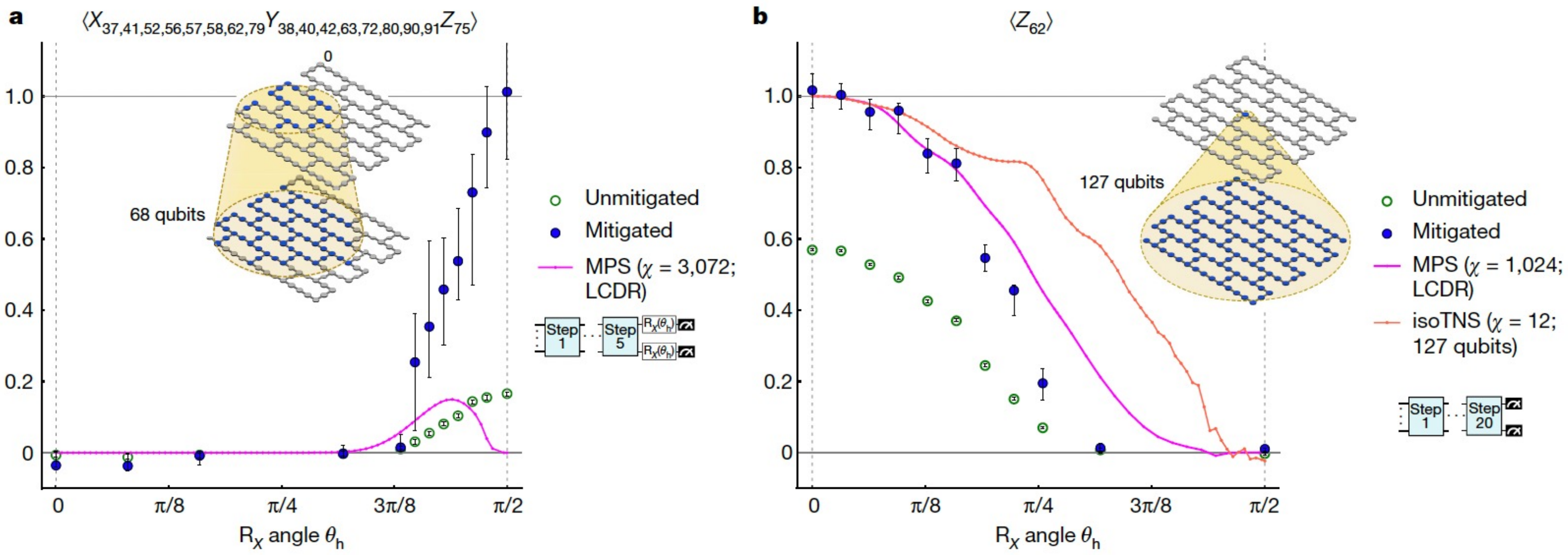
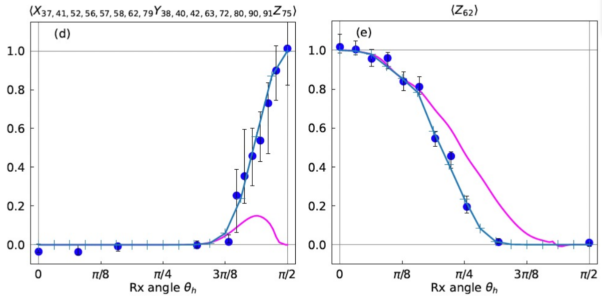



**Fig. 4 | Estimating expectation values beyond exact verification.** Plot

Even if this work is not yet beyond classical capabilities, it is clear that Trotter dynamics is a leading candidate for quantum advantage




Classical simulations using sparse Pauli dynamics method (Begusic, Chan, arXiv (2023).