Well-conditioned multi-product formulas for hardware-friendly Hamiltonian simulation

Almudena Carrera Vazquez[1,2], Daniel J. Egger[1], David Ochsner[1,2], and Stefan Woerner[1]

[1]IBM Quantum, IBM Research Europe – Zurich
[2]ETH Zurich

arXiv:2207.11268v3 [quant-ph] 24 Jul 2023

Simulating the time-evolution of a Hamiltonian is one of the most promising applications of quantum computers. Multi-Product Formulas (MPFs) are well suited to replace standard product formulas since they scale better with respect to time and approximation errors. Hamiltonian simulation with MPFs was first proposed in a fully quantum setting using a linear combination of unitaries. Here, we analyze and demonstrate a hybrid quantum-classical approach to MPFs that classically combines expectation values evaluated with a quantum computer. This has the same approximation bounds as the fully quantum MPFs, but, in contrast, requires no additional qubits, no controlled operations, and is not probabilistic. We show how to design MPFs that do not amplify the hardware and sampling errors, and demonstrate their performance. In particular, we illustrate the potential of our work by theoretically analyzing the benefits when applied to a classically intractable spin-boson model, and by computing the dynamics of the transverse field Ising model using a classical simulator as well as quantum hardware. We observe an error reduction of up to an order of magnitude when compared to a product formula approach by suppressing hardware noise with Pauli Twirling, pulse efficient transpilation, and a novel zero-noise extrapolation based on scaled cross-resonance pulses. The MPF methodology reduces the circuit depth and may therefore represent an important step towards quantum advantage for Hamiltonian simulation on noisy hardware.

Stefan Woerner: wor@zurich.ibm.com

1 Introduction

A major goal of quantum mechanics is to accurately describe how a given quantum system evolves with time. This was the original incentive argued by Feynman in 1981 to develop quantum computers [37]. More specifically, Hamiltonian simulation seeks to implement the operator e^{-iHt} as accurately as possible with an efficient quantum algorithm, which allows to study the time dynamics of a corresponding quantum system. Here, H is a Hermitian operator and *efficient* means that the running time should be polynomial in the size of the corresponding system, i.e., polylog(N) for a $N \times N$ Hamiltonian. For example, in the context of the spin-boson model this would enable the study of decoherence models at scale [60, 66].

Hamiltonian simulation can be done with system specific machines in which the hardware natively implements H, as exemplified by cold atoms [43], or with general purpose gate-based quantum computers. System specific machines allow large-scale simulation but only for the subset of Hamiltonians they natively support. Accurate Hamiltonian simulation on gate-based quantum computers is however challenging since e^{-iHt} has to be implemented on hardware with a native gate set and limited connectivity [87].

Many quantum algorithms have been proposed to simulate Hamiltonians since Feynman's talk [66]. The choice of the algorithm depends on the structure of the Hamiltonian [19, 20], e.g., whether it is sparse or dense, time-dependent [5, 8] or time-independent, and whether the time evolution is real or imaginary [35, 54].

Here, we focus on real-time simulations of a time-independent sparse Hamiltonian. Then, the two main approaches are variational quantum algorithms (VQA) [28, 61, 89, 91] and product formulas (PF) [2, 3, 7, 21, 24, 63]. VQAs do

not implement the operator e^{-iHt} but instead create a parameterized variational ansatz $|\psi(\boldsymbol{\theta})\rangle$ which approximates the state $|\Phi(t)\rangle$ governed by the Schrödinger's equation [61]. The quantum computer is then used to determine the coefficients in an equation of motion (EOM) that is classically integrated. The depth and structure of the circuit determine how well the variational approach can approximate the true state evolution. A priori, this is not clear, and we have to rely on error bounds that can be evaluated a posteriori to improve the circuit design [91]. Further, VQAs must compute the coefficients in the EOM to high accuracy. This requires a large number of circuit executions rendering the VQA approach in its current implementation unlikely to lead to a quantum advantage for time-dependent problems [66]. First steps towards overcoming this computational cost could include approximating the Quantum Fisher Information matrix using simultaneous perturbation stochastic approximation techniques [40]. In addition, the cost of estimating complex observables can be reduced with positive operator valued measures [38, 41] or classical shadows [44, 51]. In contrast, PFs are analytical techniques which perform well in practice and implement e^{-iHt} as an operator by breaking the Hamiltonian into sub-Hamiltonians, each simulated as a circuit for a small time-step. The circuit depth grows polynomially with the inverse approximation error and the simulation time. Interestingly, the existing error bounds for PFs can be several orders of magnitude larger than measured errors, even for small systems [25, 26, 46, 59], which suggests that further investigation is needed.

A promising candidate to replace PFs are multi-product formulas (MPF) [10, 23], which can be seen as a form of Richardson extrapolation [13, 71, 74]. Here, we show how to use MPFs to mitigate algorithmic errors in PFs by *classically* combining expectation values instead of implementing a quantum Linear Combination of Unitaries (LCU) [23], which was the original quantum algorithm proposed in 2012. We run a PF several times on a gate-based quantum computer with a different number of Trotter steps to estimate the same expectation value. Then, we linearly combine these estimates on a classical computer into a better estimate of the expectation value, which would otherwise only be obtainable with a large number of Trotter steps, i.e., significantly deeper circuits.

Our work makes both an algorithmic and a hardware contribution to Hamiltonian simulation with MPFs. We now describe how we build on existing references. The idea to classically combine expectation values to mitigate algorithmic errors was first introduced in 2018 in Ref. [34] applied to Hamiltonian simulation. Next, Ref. [83] showed that such a classical combination reduces the resources needed to implement the HHL algorithm [47]. In particular, Ref. [34] noticed that Richardson extrapolation can mitigate both physical [61, 79] and algorithmic errors, and showed promise in numerical experiments performed with two and three extrapolation points. In 2019 Ref. [46] noticed that a non-careful choice of sequences of Trotter steps leads to exponentially small probabilities of success for MPFs implemented as an LCU. Our work generalizes the classical combination which Ref. [83] applied to tridiagonal Toeplitz symmetric matrices to arbitrary Hamiltonians. We further improve upon Ref. [83], which focused on the HHL algorithm and thus a Hamiltonian simulation embedded in the quantum phase estimation algorithm, by showing that the classical combination is applicable to observables arising in a general Hamiltonian simulation. We also solve the vanishing success probability issue raised by Ref. [46]. We also formalize the classical combination of expectation values for Hamiltonian simulation of Ref. [34], our contribution is an algorithm to create well-conditioned MPFs implemented as a classical combination of expectation values. Here, we explain how ill-conditioning, presented in Ref. [46], increases non-algorithmic errors when classically combining expectation values. The result of our work is a hardware-friendly MPF. Its circuit depth is quadratically lower than MPFs implemented as an LCU [46] since the individual product formulas in the MPF are independently run on the hardware. The ill-conditionning is avoided by a careful choice of Trotter exponents. We compare the different algorithms in Table 1.

Crucially, we implement for the first time Hamiltonian simulation by MPF on quantum hardware. We observe an algorithmic error reduction of up to an order of magnitude while suppressing hardware noise [57] with Pauli Twirling,

pulse efficient transpilation [32], and a novel zero-noise extrapolation (ZNE) [61, 79] based on scaled cross-resonance pulses [76]. This last method is similar to digital ZNE [31, 42, 49] but allows for a continuous range of stretch factors without the need for costly pulse-calibration. We illustrate our work by computing the dynamics in simulation and on quantum hardware of the transverse field Ising model; a model ubiquitous in condensed matter and statistical mechanics [72]. The Ising Hamiltonian is

$$H = -J \sum_{\langle i,j \rangle} Z_i Z_j - h \sum_i X_i, \tag{1}$$

where J denotes the interaction strength between nearest neighboring spins $\langle i, j \rangle$, h the transverse magnetic field, and X and Z are Pauli operators. This model is often studied on both gate-based [17, 57] and analogue quantum simulations [6, 12, 90]. Additionally, we extend the scaling predictions made by Ref. [66] of the resources needed to simulate a classically intractable spin-boson model by a first-order PF and VQA. We show that in this setting our method requires exponentially less resources with respect to the system size than a first-order PF.

The main limitation of PF based approaches, including this work, is that the number of gates depends linearly on the number of terms L in the Hamiltonian, which can grow very large for chemistry applications [86]. Using randomization, the qDRIFT protocol [15] addresses this problem by trading accuracy for an L-independent algorithm. Hybrid approaches that interpolate between qDRIFT and PFs improve the asymptotic complexity with respect to both methods [45, 68]. However, they do not improve the scaling with respect to the simulation time nor the precision. Randomizing the implementation of the LCU has also been explored. For example, Ref. [36] shows a drastic reduction in circuit depth by implementing the LCU on average and increasing the number of circuits. However, this method still requires two controlled PFs per circuit. Alternatively, Ref. [85] presents a randomized phase estimation algorithm based partially on the LCU approach and has an L-independent complexity. Although only one ancilla is required and most of the complexity is handled by the sampling, the circuits still require $\mathcal{O}(\epsilon^{-2})$ controlled Pauli rotations, where ϵ is the desired precision. The MPF approach we present here is suitable for simulating general dynamics, in contrast to Ref. [85], and for applications that require higher precision, since we achieve more accurate results by increasing the number of shallow circuits we run.

This paper is structured as follows. In Sec. 2 we review the theory of PFs and then show how to implement well-conditioned MPFs by *classically* combining the outcome of multiple PFs. Further, in Sec. 3 we update the resource estimate for spin-boson models in Ref. [66] with our MPF approach to illustrate that we can achieve the same approximation accuracy as a first-order PF using shallower circuits. In Sec. 4, we discuss error mitigation strategies for quantum hardware and then illustrate our MPF method on fixed-frequency superconducing transmon qubits by simulating an Ising model. We conclude our work in Sec. 5 and provide suggestions for further research.

2 Multi-product formulas

Product formulas are used to build a unitary matrix approximating e^{-iHt} by (i) writing H as $\sum_j H_j$ (e.g. where each H_j is a 1-sparse matrix [3, 24]) such that (ii) each $e^{-iH_j t}$ has an efficient quantum circuit representation, and finally (iii), if the matrices H_j do not commute with each other, use an approximation. The first-order PF—also known as the Lie-Trotter formula—is

$$S_1(t) := \prod_{j=1}^{J} e^{-iH_j t} \tag{2}$$

and has a quadratic error term in t, i.e. $S_1(t) = e^{-iHt} + \mathcal{O}(t^2)$. Higher order PFs converge faster and are defined recursively as follows.

Definition 2.1 (Lie-Trotter-Suzuki, [78])**.** Let $H \in \mathbb{C}^{N \times N}$ be a Hermitian matrix with $H = \sum_{j=1}^{J} H_j$, and $\chi \in \mathbb{N}$. Then, the 2χ-order symmetric formula approximating e^{-iHt} is recursively defined by

$$\begin{aligned} S_2(t) &:= \prod_{j=1}^{J} e^{-iH_j t/2} \prod_{j=J}^{1} e^{-iH_j t/2}, \\ S_{2\chi}(t) &:= S_{2\chi-2}\left(s_\chi t\right)^2 S_{2\chi-2}\left[(1-4s_\chi)\, t\right] S_{2\chi-2}\left(s_\chi t\right)^2, \end{aligned} \tag{3}$$

where $s_p = \left(4 - 4^{1/(2p-1)}\right)^{-1}$ for $p \in \mathbb{N}$.

Table 1: Second-order PF compared to different MPFs. The last row corresponds to the algorithm presented in this paper. l denotes the number of extrapolation points, $k_1 < \ldots < k_l$ to the Trotter exponents, and ε' to the non-algorithmic errors such as physical or sampling noise. We assume that the MPFs are built from S_2, that $l \leq 15$, and that the simulation time satisfies $t \leq 1$. The harmonic Trotter sequence corresponds to $[k_1, k_1+1, \ldots, k_1+l-1]$ and exemplifies a generic ill-conditioned sequence that may arise from a non-careful choice of Trotter exponents, such as proposed in Refs. [23, 34], where no particular sequences were specified. The constant factor of the circuit depth scaling inside the $\mathcal{O}(\cdot)$-notation for LCU implementations can be very large once the circuit is transpiled to the hardware due to the limited connectivity and the many controlled gates.

Algo.	Combination	Trotter seq.	#circuits	Max. circ. depth	Success prob.	Approx. error	$\|\vec{a}\|_1$
S_2	N/A	$[k]$	1	$\mathcal{O}(k)$	1	$\mathcal{O}\left(\frac{1}{k^2}+\varepsilon'\right)$	N/A
MPF	quantum (LCU)	harmonic [23]	1	$\mathcal{O}(l^2)$	$\mathcal{O}(\|\vec{a}\|_1^{-2})$	$\mathcal{O}\left(\frac{1}{k_1^{2l+1}}+\varepsilon'\right)$	$e^{\Omega(l)}$
		well-conditioned [46]		$\mathcal{O}(l^2\log(l))$			≤ 2
	classical	harmonic [34]	l	$\mathcal{O}(l)$	1	$\mathcal{O}\left(\frac{1}{k_1^{2l+1}}+\|\vec{a}\|_1\varepsilon'\right)$	$e^{\Omega(l)}$
		well-conditioned (our work)		$\mathcal{O}(l\log(l))$			≤ 2

When the duration t of the time evolution is long we approximate e^{-iHt} by splitting t into k equal *Trotter steps* of length t/k and apply a product formula S_χ in each interval such that

$$S_\chi^k\left(\tfrac{t}{k}\right) = \left[S_\chi\left(\tfrac{t}{k}\right)\right]^k = e^{-iHt} + \mathcal{O}\Big(t\left(\tfrac{t}{k}\right)^\chi\Big). \quad (4)$$

The last equality follows from the triangle inequality. As the order χ of the PF increases the number of Trotter steps k required to reach a target error decreases. However, the depth of the quantum circuit scales exponentially with χ [23, 78]. Therefore, choosing the optimal formula for a given problem is a non-trivial task further complicated by the lack of tight analytical bounds [25]. Moreover, even if one could find the PF that minimizes the resources needed for a given Hamiltonian, it would still require deep circuits, rendering the PF approach very challenging to implement on noisy quantum hardware.

Multi-product formulas, which can be seen as a form of Richardson extrapolation, are a promising candidate to replace product formulas [10, 23]. MPFs only require a sum of polynomially many unitary operations to achieve the same accuracy as PFs with exponentially many unitaries would. They achieve this by linearly combining PFs of the same order χ but varying Trotter exponents k to cancel lower order error terms.

Definition 2.2 (Multi-product formula)**.** For $l \in \mathbb{N}$, we define the Trotter exponents as $\vec{k} = (k_1, \ldots, k_l)$ where the k_i's are distinct natural numbers, and the extrapolation weights as $\vec{a} = (a_1, \ldots, a_l)$ with $a_i \in \mathbb{R}$. Then, the multi-product formula $M_{l,\chi}(t)$ is

$$M_{l,\chi}(t) = \sum_{j=1}^{l} a_j S_\chi^{k_j}\left(\tfrac{t}{k_j}\right), \quad (5)$$

where χ, S_χ are as in Def. 2.1. To cancel the first $l-1$ error terms, the extrapolation weights must satisfy

$$\sum_{j=1}^{l} a_j = 1, \quad \text{and} \quad \sum_{j=1}^{l-1} a_j/k_j^\eta = 0, \quad (6)$$

where $\eta = \chi+2n$ for symmetric product formulas and $\eta = \chi+n$ otherwise, for $n = 0, ..., l-2$. For more details, see Appendix A.

2.1 Circuit implementation

Ref. [23] introduced MPFs as a quantum algorithm to approximate Hamiltonian simulation. The linear combination in Eq. (5) is translated to a quantum circuit by the Linear Combination of Unitaries (LCU) method, illustrated in Fig. 1. Asymptotically, this method with $k_j = j$, i.e. the minimal sequence with l different natural numbers, implements $\mathcal{O}(l^2)$ Trotter steps and approximates the Hamiltonian better than PFs with as many Trotter steps. More precisely, MPFs require exponentially fewer Trotter steps [27] to

achieve the same accuracy. However, the LCU implementation is probabilistic with a success probability that decreases exponentially as $e^{-\Omega(l)}$. Furthermore, even the shallowest circuits are too deep to implement on noisy hardware since the LCU method sums Trotter steps of varying order as controlled unitaries, requiring additional qubits and controlled operations. For example, simulating a five linearly connected spin Ising model with an MPF with $\vec{k} = [k_1, k_2, k_3]$ implemented by an LCU requires $108(k_1+k_2)+40k_3+4$ CNOT gates, see Appendix B.

While Eq. (5) is a linear combination of operators, in practice we measure operator expectation values after a quantum circuit. Crucially, as we show here and illustrate in Fig. 1, the linear combination in MPFs can thus be done classically. This allows to parallelize the circuits implementing the PF for different Trotter steps, resulting in shallow circuits with $\mathcal{O}(l)$ Trotter steps, which is better for noisy hardware. We therefore implement each $S_\chi^{k_j}$ in Eq. (5) as an independent quantum circuit. As shown in Appendix C we can substitute the S_χ's in Eq. (5) by their expectation values. More precisely, computing an expectation value E after approximately simulating e^{-iHt} with the PF $S_{2\chi}^{k_j}(t/k_j)$ yields an approximation E_j of the expectation value E for each Trotter exponent k_j with an approximation error related to χ and k_j by Eq. (4). Finally, combining the E_j's as

$$\sum_{j=1}^{l} a_j E_j = E + \mathcal{O}\left(\frac{t^{2(\chi+l)+1}}{k_1^{2(\chi+l)}}\right), \tag{7}$$

gives an estimation of E where the remaining asymptotic error depends on the base PF, see Appendix A.

2.2 Choosing Trotter exponents

MPFs can achieve exponentially lower algorithmic errors than PFs while requiring similar circuit depths. However, a poor choice of the Trotter exponents and extrapolation weights amplifies other error sources such as physical or sampling noise so that the error in the final quantity is larger than the largest algorithmic error in the MPF. This choice-of-sequences issue was raised in Ref. [46] where the authors noted that choosing the arithmetic progression for the Trotter exponents $k_j = j$, which gives the shortest circuits,

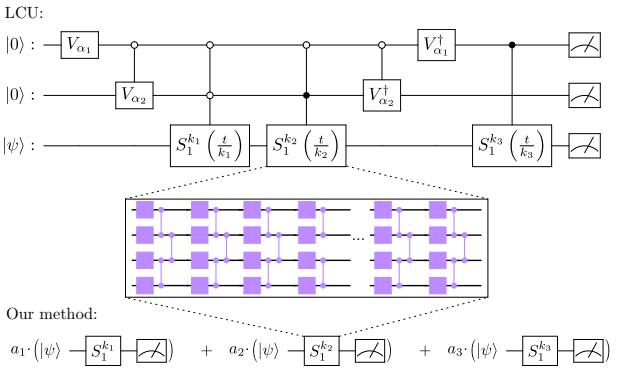


Figure 1: Implementation of MPFs. The LCU circuit probabilistically implements the linear combination $\sum_j S_1^{k_j}(t/k_j)$ and is successful if the measurement outcome of the top two qubits is 0. Our method implements each of the PFs as a separate quantum circuit and calculates the linear combination classically. Running the PFs separately instead of in sequence circumvents the need of auxiliary qubits and significantly reduces the circuit depth, especially as the number of terms in the linear combination increases.

leads to an ill-conditioned MPF, i.e. $\|\vec{a}\|_1 = e^{\Omega(l)}$ where $\|\vec{a}\|_1 := \sum_j^l |a_j|$. When implemented as a quantum circuit as an LCU this ill-conditioning translates into an exponentially small success probability scaling as $\mathcal{O}(\|\vec{a}\|_1^{-2})$. Since our approach is not probabilistic, i.e., we classically compute the linear combination in Eq. (7) rather than implementing $M_{l,\chi}$ as an LCU, the condition number $\|\vec{a}\|_1$ amplifies any error that is not the Trotter error by a factor of $\|\vec{a}\|_1$. More precisely, suppose that in Eq. (7) we calculated the expectation values E_j with a quantum computer and we have several error sources. If $\varepsilon(k_j)$ denotes the algorithmic error of the Trotter exponent k_j and other sources arising from, e.g., sampling or physical noise are labelled ε', then

$$E_j = E + \varepsilon(k_j) + \varepsilon'. \tag{8}$$

Now, the result of the MPF is

$$\sum_{j=1}^{l} a_j \left(E + \varepsilon(k_j) + \varepsilon'\right) \tag{9}$$

$$= E + \mathcal{O}\left(\frac{t^{2(l+\chi)+1}}{k_1^{2(l+\chi)}}\right) + \mathcal{O}\left(\|\vec{a}\|_1 \varepsilon'\right). \tag{10}$$

Eq. (10) shows that if $\|\vec{a}\|_1$ is large, then ε' is amplified and as a result, the sum in Eq. (9) can be a worse approximation to E than the E_j's. Appendix D shows a toy example illustrating the

asymptotic behaviour, where ε' is the sampling noise of a Bernoulli variable. Here, the increase in non-algorithmic errors results in an increase in the sampling overhead.

Ref. [46] provides choices of $\vec{k}$ for which $\|\vec{a}\|_1$ remains "small" to avoid exponentially low success probabilities, i.e., ill-conditioned MPFs. The authors minimize either $\|\vec{a}\|_1\|\vec{k}\|_1$ or $\|\vec{a}\|_1$ and provide the corresponding values for $\chi = 2$ and 4. Their method requires $\mathcal{O}(l^2 \log l)$ Trotter steps to implement the LCU instead of $\mathcal{O}(l^2)$ for $k_j = j$. By contrast, our classical combination of expectation values needs only $\mathcal{O}(l \log l)$ Trotter steps for the longest circuit. Importantly, the condition number decreases exponentially from $\|\vec{a}\|_1 = e^{\Omega(l)}$ to $\|\vec{a}\|_1 = \mathcal{O}(\log l)$. Furthermore, Tab. I in Ref. [46] shows that for an MPF based on S_2 and a practical value of $l \leq 15$, the condition number $\|\vec{a}\|_1 < 1.7$ and is therefore negligible. To avoid increasing non-algorithmic errors, see Eq. (10), we chose the Trotter exponents by minimizing $\|\vec{a}\|_1$ under the constraints in Eq. (6) for a given possible range of the k_j's. The algorithm is described in Appendix E.

We illustrate how to choose the Trotter exponents with the Ising model for five linearly connected spins, see Eq. (1), with $H_1 = -J\sum_{i=0}^{3} Z_i Z_{i+1}$ and $H_2 = -h\sum_{i=0}^{4} X_i$. Therefore, $e^{-iH_1 t}$ and $e^{-iH_2 t}$ are implemented with layers of two-qubit R_{ZZ} and single-qubit R_X rotations, respectively. As benchmark we compute the exact local magnetization $\langle Z_0\rangle^*$ for $h = 1$, $J = 0.5$ and $t = 0.5$ by numerically exponentiating $H_1 + H_2$. Next, we compute the local magnetization $\langle Z_0(k_j)\rangle$ with a first-order PF and k_j Trotter steps. We calculate the relative error $|\langle Z_0(k_j)\rangle - \langle Z_0\rangle^*|/|\langle Z_0\rangle^*|$ which we fit to k_j^{-1}, see the blue dots and dashed line in Fig. 2. The condition number of an MPF based on S_1 grows faster than for S_2, see Appendix D. Therefore, we empirically set the threshold defining a well-conditioned MPF to $\|\vec{a}\|_1 \leq 3$ instead of 1.7. Thus, we compute well- and ill-conditioned MPFs, with $\|\vec{a}\|_1 \leq 3$ and $\|\vec{a}\|_1 > 3$, respectively, with different Trotter exponents by classically combining the expectation values $\langle Z_0(k_j)\rangle$. The values of $\vec{a}$ are given in Appendix E. In the absence of non-algorithmic errors, i.e., $\varepsilon' = 0$, all MPFs reduce the error compared to the benchmark, see triangles in Fig. 2. Next, we artificially add a fixed perturbation $\varepsilon' = 10^{-3}$ following $\langle Z_0(k_j)\rangle \to \langle Z_0(k_j)\rangle + \text{sign}(a_j)\varepsilon'$ before performing the linear combination. The well-conditioned MPFs still reduce the error compared to ill-conditioned sequences which are worse than the benchmark, see stars in Fig. 2. For example, the noisy S_1-based MPF $[2, 4]$ whose deepest circuit requires $k_2 = 4$ Trotter steps has a lower error than the PF with $k = 8$ Totter steps. This is a factor of two reduction in circuit depth.

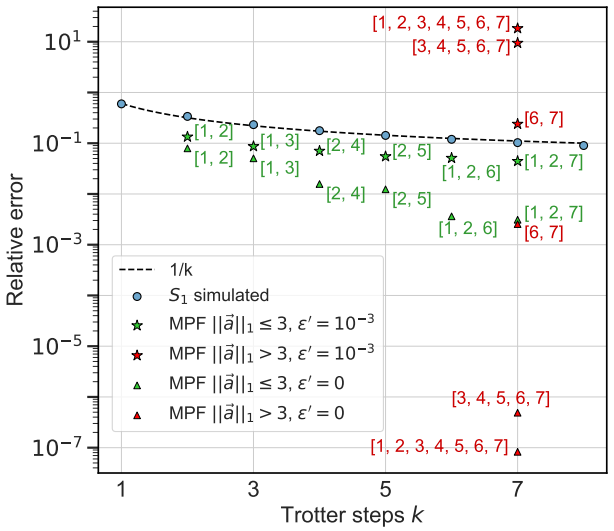


Figure 2: Local magnetization of the five-spin Ising model. The Ising simulation is done with a first-order Trotter formula. The y-axis shows the relative error $|\langle Z_0\rangle - \langle Z_0\rangle^*|/|\langle Z_0\rangle^*|$ between the simulated and exact values. The extrapolated values are computed using well-conditioned (green markers) and ill-conditioned (red markers) MPFs with different Trotter exponents (shown in the brackets $[k_1, k_2, ...]$) by combining the expectation values classically. The calculation is done without non-algorithmic errors ($\varepsilon' = 0$, triangular markers) and with non-algorithmic errors ($\varepsilon' = 10^{-3}$, star markers).

3 Spin-boson model

The spin-boson model [60] is a model of quantum dissipation. It gives insights into the transition between coherent and incoherent behaviour as well as a phase transition suppressing quantum tunneling. Its dynamics can be simulated with analogue simulators [33, 52] or gate-based quantum computers [66]. In Ref. [66] the authors compare the performance of VQAs based on McLachlan's variational principle [65] against a first-order PF to simulate the dynamics of the spin-boson model. The authors made scaling predictions for a classically intractable system for

both, VQA and first-order PFs. They conclude that VQA in its current form is unlikely to lead to a quantum advantage due to the large number of circuit executions, although it is assumed to have shallower circuits than first-order PFs. Further, they show that first-order PFs that achieve the desired accuracy lead to very deep circuits, which might render them infeasible as well. We now extend their results by adding second-order PFs and MPFs to the comparison, showing a significant circuit depth reduction.

More precisely, the spin-boson Hamiltonian

$$H = \sum_{k=1}^{M} \omega_k a_k^\dagger a_k + \frac{\omega_s}{2} Z + \Delta X + \sum_{k=1}^{M} g_k X \left(a_k^\dagger + a_k\right) \tag{11}$$

describes a two-level system, such as an atom, coupled to a bath of M bosons [30, 39]. Here, the bosonic operators $a_k^\dagger$ (a_k) create (annihilate) harmonic basis states with eigenfrequencies ω_k. The Pauli matrices X, Z act on the state of the spin with eigenfrequency ω_s and tunneling rate Δ. The coupling strength between the spin and the k^{th} bosonic mode is g_k. As in Ref. [66] we consider the resonant case $\omega_k \equiv \omega$, $g_k \equiv g$ and ultrastrong coupling, i.e. $g/\omega = 0.5$ with $\omega_s = -1$ and $\Delta = 0$. We consider different spin-boson models in which the M bosonic modes are truncated at different occupation numbers $n^{\max}$.

As in Ref. [66], we are interested in $t = 10$. Since we suspected that the algorithm would also work for sequences with $t/k_1 > 1$, we first computed several sequences of Trotter exponents with $l = 2, 3$ and 4. These sequences and their corresponding extrapolation weights satisfied $\|\vec{a}\|_1 \leq 2$ to ensure that errors arising from noise or sampling are amplified by at most a factor of 2. Next, we computed the approximation error ε for each MPF and selected the one which required less resources to achieve the target accuracy ε_t. We observed that this method works well in practice even when $t/k_1 > 1$, confirming that the analytical error bounds on PFs are not tight. Furthermore, we found that the second-order PF always requires shallower circuits than the PFs S_4 and S_6 when used in an MPF. We also found that the PF S_2 alone always requires shallower circuits than S_1. For small spin-boson systems we see a reduction of up to a factor of eight in circuit depth, see Appendix F.

We derive the scaling of the MPF approach taking into account the dependence of the error on the number of qubits N_q and the number of different Trotter exponents l. First, the circuit depth of the spin-boson model grows linearly with N_q [66]. Next, as a worst case estimate we set $\alpha = t$ such that $t/k_j \leq 1$ for $j = 1, \cdots, l$ so that we can bound the error in the MPF by $\mathcal{O}\left(t^{2l+1}/(2l+1)!\right)$ [46, 83]. A short calculation detailed in Appendix F shows that our hybrid scheme requires the deepest circuit to have

$$\mathcal{O}\left(\frac{t \log^2(N_q/\varepsilon_t)}{\log^2 \log(N_q/\varepsilon_t)}\right) \tag{12}$$

Trotter steps instead of the $\mathcal{O}(\sqrt{t^3 N_q^2/\varepsilon_t})$ required by S_2. A more careful choice of Trotter exponents and α may allow better error bounds. Although the sampling overhead of ε_t^{-2}, not shown in Eq. (12), cannot be avoided in either approach, the depth of the MPF circuits grows logarithmically instead of polynomially in ε_t^{-1}. Our method is advantageous for large systems as it reduces the linear dependency of errors on N_q to a logarithmic one. Ref. [66] concluded that first-order Trotter formulas are a better approach than VQA to simulate spin-boson models of intractable system sizes. However, S_1 exhibits a scaling linear in N_q, making such a simulation challenging to implement on noisy hardware. MPFs implemented as a linear combination of expectation values are therefore promising to simulate classically intractable quantum systems.

4 Ising model computation on hardware

Quantum computing hardware has increased in scale and quality [53]. However, gate errors on cross-resonance based hardware are non-negligible. For example, a typical CNOT gate on IBM Quantum hardware has an average error ranging from 0.5% to 1%. This makes it challenging to implement MPFs on noisy quantum hardware as hardware-related errors ε' must be kept comparable to or smaller than the total Trotter error which is close to 10%, see Eq. (10) and Fig. 2. We can however suppress hardware errors in the measured observables with noise mitigation methods which we describe in Sec. 4.1 – 4.3 in the same order as we apply them as Qiskit transpiler passes [70]. Hardware results are in Sec. 4.4. They show a factor of six reduction in

circuit depth of an MPF implementation over a PF.

4.1 Pauli Twirling

Pauli twirling (PT) converts an arbitrary noise channel into a stochastic Pauli error channel, suppressing off-diagonal coherent error contributions [57, 84]. We first apply PT to the R_{ZZ} gates, the dominant error source, to convert their noise channels into a stochastic Pauli error channel. PT is implemented by sandwiching each R_{ZZ} gate between randomly sampled Pauli gates chosen so that the net operation of the circuit is unaffected in the absence of noise. The gate noise is reduced to a stochastic form by averaging the results of $N_{\mathrm{PT}} \in \mathbb{N}$ logically equivalent random circuits [14, 17, 48, 57, 84]. As in Ref. [57], we randomly sample twirling gates from the set $\mathbb{G}_{ZZ} = \{I \otimes I, X \otimes X, Y \otimes Y, Z \otimes Z, X \otimes Y, Y \otimes X, Z \otimes I, I \otimes Z\}$ that commutes with $R_{ZZ}(\theta)$. We apply the same Pauli gates before and after the $R_{ZZ}(\theta)$ gate, see Fig. 3(a) and (b). Each quantum circuit is executed with N/N_{PT} shots on that hardware where N is the total shot budget.

4.2 Pulse efficient transpilation and calibration

Second, we use a pulse-efficient circuit transpilation which replaces the $R_{ZZ}(\theta)$ gates by R_{ZX} gates with echos exposed in the circuit as presented in Ref. [32, 76], see Fig. 3(c). This avoids the onerous double CNOT implementation of $R_{ZZ}(\theta)$. The R_{ZX} gates are implemented by scaling the hardware native cross-resonance (CR) pulses [73, 77] found in the calibrated schedules of the CNOT gates. Since we run a Trotter simulation where the rotation angles θ are small the flat-top of the square pulses with Gaussian flanks vanish. We must therefore implement $R_{ZX}(\theta)$ with Gaussian CR pulses of varying amplitude. To overcome non-linear effects we calibrate the amplitude of the CR pulses, see Appendix H.

4.3 Zero-noise extrapolation with scaled cross-resonance gates

The noise in quantum computers can be mitigated with zero-noise extrapolation (ZNE) which runs logically equivalent circuits with different noise levels [61, 79]. Ideally, such circuits are implemented by reducing the amplitude and increasing the duration of all the underlying pulses to increase the noise [56]. This requires extensive gate calibration, which is impractical and reduces the availability of a cloud-based quantum computing system. To overcome this limitation digital ZNE replaces each CNOT gate with $2n+1$ CNOT gates [16, 31, 75]. This, however, limits the possible stretch factors to $c = 2n+1$. Similarly, the Mitiq package [58] inserts pairs of CNOT gates but allows for fractional stretch factors by repeating selected gate layers instead of repeating all CNOT gates.

Here, we implement ZNE by adding the gate sequence $R_{ZX}(-\theta_c)R_{ZX}(\theta_c)$ after the first layer of $R_{ZZ}(\theta)$ gates, see Fig. 3(d). Crucially, the rotation angle $\theta_c \neq \theta$ which allows us to create a *continuous* range of stretch factors by linearly scaling the width w of the echoed CR pulses in the $R_{ZX}(\pm\theta_c)$ gates as discussed in Ref. [32, 76], see Fig. 3(e). Finally, we calculate the effective stretch factor c of each circuit as the fraction between the duration of the stretched circuit and the duration of the original circuit, both without measurements. The smallest possible stretch factor c_{min} we can implement, indicated in Fig. 4(b), occurs when the flat-top w of the CR pulses vanishes, see Fig. 3(e). By scaling only the R_{ZX} gates rather than each gate in the circuit, it is possible to use a wider range of stretch factors for the same T_1 times of the hardware. Each observable we measure is then extrapolated to the zero-noise limit with an exponential fit $ae^{-bc} + d$ where c is the stretch factor and a, b, and d are fit parameters, see Fig. 4(b). We compare our ZNE method to the Mitiq package in Appendix I.

4.4 Results

We compute the dynamics of the Ising model with five linearly connected spins and $J = 0.5$, $h = 1$, and an evolution time of $t = 0.5$. To select five good qubits on *ibmq_montreal* we perform quantum process tomography [9, 67] of scaled-CR $R_{ZX}(\theta_c)R_{ZX}(-\theta_c)$ gates and retain the five linearly connected qubits 13, 14, 16, 19, and 20 as their gate fidelity decay visually resembles their T_1 decay, see Appendix J. We calculate the global magnetization $\langle Z \rangle$ by measuring the local magnetization $\langle Z_i \rangle$ of each qubit and take the average $\langle Z \rangle = \frac{1}{5}\sum_{i=1}^{5}\langle Z_i \rangle$, see Fig. 4(a). As benchmark we compute the first-order Trotter formula S_1 with exponents $k_j \in \{1, ..., 10\}$ on a noiseless state-vector simulator and compare it to the

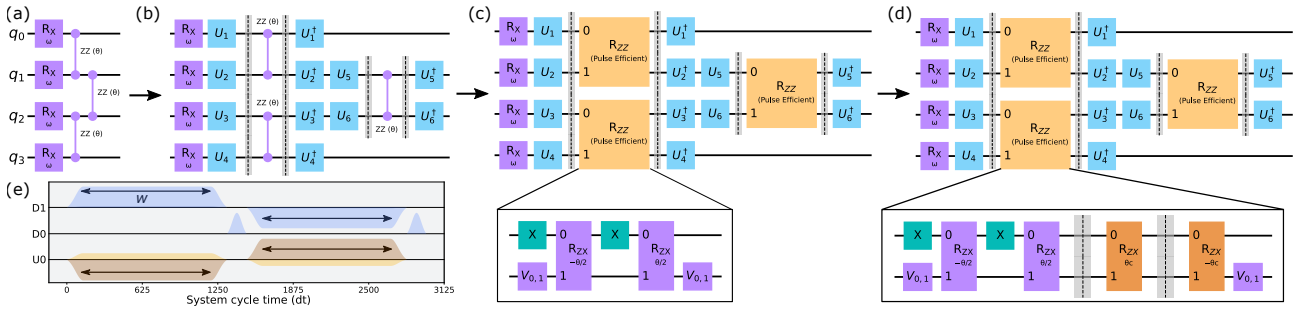


Figure 3: Transpiler steps that implement error mitigation. (a) A single Trotter step on four qubits with H_1 and H_2 implemented as single- and two-qubit gates, respectively. (b) Pauli Twirling applied to the R_{ZZ} gates. The Pauli gates U_j applied to the qubits before the $R_{ZZ}(\theta)$ gates are chosen uniformly from the set $\mathbb{G}_{ZZ}$, which commutes with the twirled gate. The same Pauli gates are applied after the $R_{ZZ}(\theta)$ gate so that the net operation is unchanged. (c) A pulse-efficient transpiler pass scales the cross-resonance pulses of the backend-provided calibrated CNOT gates to the intended rotation angle θ. The R_{ZZ} gates are implemented as echoed cross-resonance gates shown in the inset. (d) We identify all $R_{ZX}(-\theta/2)XR_{ZX}(\theta/2)$ gate sequences in the first Trotter layer and append the sequence $R_{ZX}(-\theta_c)R_{ZX}(\theta_c)$ to perform ZNE. (e) Pulse sequence implementing $R_{ZX}(\theta_c)$ with $\theta_c = \pi$. The width w of the flat-top of the CR pulses and the corresponding rotation angle θ_c are chosen to produce a target stretch factor c.

exact analytical value of e^{-iHt}, see Eq. (2) and the blue dots in Fig. 4(c). Next, we compute S_1 on *ibmq_montreal* with the error mitigation strategies described in Sec. 4.1-4.3. The properties of *ibmq_montreal* are shown in Appendix J. We also ran the corresponding circuits with and without dynamical decoupling but did not see a meaningful effect. We elaborate on this in Appendix G. For each Trotter step we create eight random PT circuits by dressing the R_{ZZ} gates with gates uniformly drawn from $\mathbb{G}_{ZZ}$. Each of these eight circuits is executed 20 times with different stretch-factors θ_c linearly spaced in the interval $[\theta_c^{\min}, \theta_c^{\max}]$. Here, $\theta_c^{\max}$ is chosen such that the pulse-schedule of ten Trotter steps does not exceed the shortest T_1 time. For every θ_c each of the eight circuits is evaluated with $10^5/8$ shots to keep the physical and sampling noise below $\varepsilon' = 10^{-2}$. This process is repeated ten times to gather statistics. Therefore, a zero-noise extrapolated global magnetization $\langle Z(k_j)\rangle$, indicated as green stars in Fig. 4(b) and (c), is computed with a total of $10 \times 20 \times 10^5 = 2 \cdot 10^7$ shots. Before computing each $\langle Z(k_j)\rangle$ we calibrate the R_{ZX} gates using the procedure in Appendix H. All the measured expectation values are corrected for readout errors with a tensor product of individual qubit calibration matrices [11, 57]. The noise mitigation significantly increases the accuracy in the measured $\langle Z(k_j)\rangle$, see Fig. 4(b). For up to four Trotter steps the measured $\langle Z(k_j)\rangle$ agree well with the noiseless benchmark. At Trotter steps $k_j > 4$ the $\langle Z(k_j)\rangle$ have a lower error than predicted by the noiseless benchmark which is a coincidence.

We build the MPF estimation of the zero-noise extrapolated values of $\langle Z\rangle$ using Eq. (7). We select the Trotter exponents k_j to combine and the corresponding weights a_j following the optimization algorithm described in Appendix E, which also contains the values of $\vec{k}$ and $\vec{a}$, with the threshold $\|\vec{a}\|_1 \leq 3$. Crucially, we observe that the MPF significantly reduces the error. For example, the combinations [1, 2], [1, 3], and [2, 4] all have the same accuracy as a first-order formula with 24 Trotter steps, see green crosses in Fig. 4(c). This is a reduction in circuit depth by a factor of twelve to six over the product formula.

5 Conclusion

We present a hybrid quantum-classical approach to MPFs that requires neither additional qubits nor controlled operations as do previous approaches based on an LCU [23, 46]. Our method approximates a Hamiltonian evolution with the same accuracy as PFs or MPFs implemented with an LCU but with significantly shallower circuits by implementing each unitary in the LCU as a standalone circuit. Therefore, even though the total number of matrix exponentials in both approaches is the same, the reduction in gate count for the deepest circuit can be substantial as illustrated by the $l = 3$ example for which our approach has 22 times less two qubit gates than an LCU implementation when $\vec{k} = [1, 2, 7]$. Re-

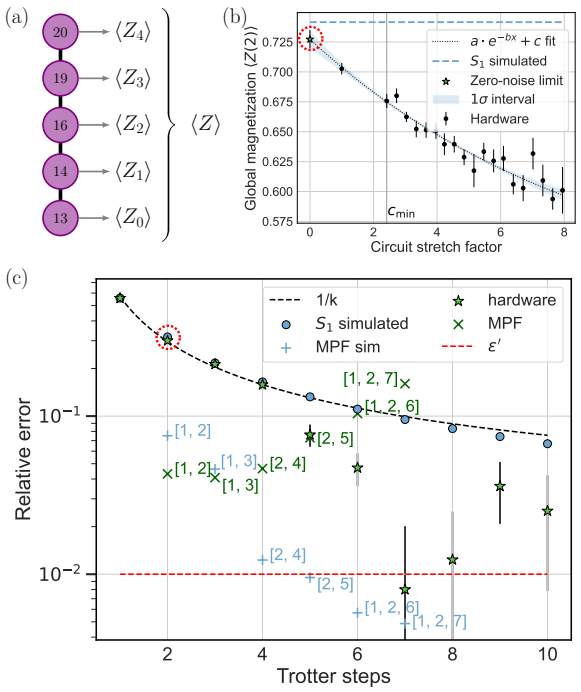


Figure 4: Global magnetization of a five-spin Ising model. (a) The global magnetization is the average local magnetization of qubits 13, 14, 16, 19, and 20. (b) Zero-noise extrapolation of $\langle Z(k_j)\rangle$ for $k_j = 2$ Trotter steps. The black dots represent $\langle Z(2)\rangle$ at different circuit stretch factors c. (c) Error on the simulated global magnetization relative to the exact value $\langle Z\rangle^*$ obtained numerically by exponentiating H, i.e., $|\langle Z\rangle - \langle Z\rangle^*|/|\langle Z\rangle^*|$. The blue dots show a noiseless state-vector simulation. The corresponding hardware measured ZNE expectation values are depicted as green stars. The error on the global magnetization obtained from MPFs based on the state-vector simulation and hardware measurements are shown as blue pluses and green crosses, respectively. The Trotter exponents underlying the MPF are shown in square brackets. The results were obtained between May 19th and June 2nd 2022.

ducing the gate and qubit count makes it possible to execute MPFs on noisy hardware that achieve accurate results, such as those shown in Fig. 4, that would not be possible if implemented as an LCU.

The fast convergence of MPFs to the exact operator e^{-iHt} also means that the approximation error soon becomes smaller than the noise levels on the hardware. Error mitigation is thus crucial to suppress hardware noise. As the hardware and error mitigation techniques improve, the advantages of our MPF method will become increasingly noticeable. To mitigate errors we used PT and pulse-efficient transpilation and introduced an implementation of ZNE that provides a continuous range of stretch factors while simultaneously avoiding onerous pulse calibration. Crucially, this method overcomes the $c \in \{1, 3, 5, ...\}$ limitation of the digital ZNE which rapidly becomes infeasible for deep circuits. With these error mitigation methods we demonstrated on hardware that MPFs can achieve the same accuracy as a PF but with up to a factor of 12 reduction in circuit depth for a five spin Ising Hamiltonian.

Our approach has shallower circuits than MPFs based on an LCU for $t \leq 1$. Future work may investigate whether there is a regime for large values of t in which the LCU circuits are shallower than the classical combination of expectation values we propose. For example, since an MPF implemented as an LCU is one quantum circuit, it may be possible to repeat $t/\Delta t$ times the circuit for a time Δt. This may use less Trotter steps than a classical combination of expectation values. In addition, tighter bounds could be derived, since as mentioned in Sec. 3, our experiments showed that the scheme still works even when $t/k_1 > 1$. Regarding ZNE as presented in Sec. 4.3: a deeper study of the optimal number of stretch factors is warranted. Furthermore, we empirically found that inserting the gate sequence $R_{ZX}(-\theta_c)R_{ZX}(\theta_c)$ after the gates in the first layer of R_{ZZ} gates allowed us to mitigate errors. However, the choices of the sequence and where to insert them are not unique. While we tried several possibilities, a more rigorous study is left as further work. We obtained our ZNE expectation values with an exponential fit using the effective stretch factors as the independent variable. These choices could be either studied empirically in more detail or backed up with theory.

We conjecture that our scheme is optimal when used with S_2, where optimal means requiring the shallowest circuits to achieve a given approximation error. In summary, the linear combinations of expectation values makes it possible to use the MPF approach on noisy hardware. This allows us to reduce circuit depth by orders of magnitude which we believe is crucial to reach a quantum advantage for Hamiltonian simulation on noisy quantum hardware.

6 Acknowledgements

The authors thank Nate Earnest, Andreas Fuhrer, Youngseok Kim, David Layden and Alexander Miessen for useful discussions. This work was supported as a part of NCCR QSIT, a National Centre of Competence (or Excellence) in Research, funded by the Swiss National Science Foundation (grant number 51NF40-185902).

References

[1] IBM Quantum. https://quantum-computing.ibm.com/.

[2] D. Aharonov and A. Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge. In *Proc. 35th Annu. ACM Symp. Theory Comput.*, page 20–29, 2003. DOI: 10.1145/780542.780546.

[3] G. Ahokas. Improved algorithms for approximate quantum Fourier transforms and sparse Hamiltonian simulations. 2004. DOI: 10.11575/PRISM/22839.

[4] T. Alexander, N. Kanazawa, D. J. Egger, L. Capelluto, C. J. Wood, A. Javadi-Abhari, and D. C. McKay. Qiskit pulse: programming quantum computers through the cloud with pulses. *Quantum Sci. Technol.*, 5(4):044006, aug 2020. DOI: 10.1088/2058-9565/aba404.

[5] D. An, D. Fang, and L. Lin. Time-dependent Hamiltonian simulation of highly oscillatory dynamics and superconvergence for Schrödinger equation. *Quantum*, 6:690, 2022. DOI: 10.22331/q-2022-04-15-690.

[6] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin. Probing many-body dynamics on a 51-atom quantum simulator. *Nature*, 551(7682):579–584, Nov 2017. ISSN 1476-4687. DOI: 10.1038/nature24622.

[7] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders. Efficient quantum algorithms for simulating sparse Hamiltonians. *Commun. Math. Phys.*, 270(2):359–371, Mar 2007. ISSN 1432-0916. DOI: 10.1007/s00220-006-0150-x.

[8] D. W. Berry, A. M. Childs, Y. Su, X. Wang, and N. Wiebe. Time-dependent Hamiltonian simulation with L^1-norm scaling. *Quantum*, 4:254, 2020. DOI: 10.22331/q-2020-04-20-254.

[9] R. C. Bialczak, M. Ansmann, M. Hofheinz, E. Lucero, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, J. Wenner, M. Steffen, A. N. Cleland, and J. M. Martinis. Quantum process tomography of a universal entangling gate implemented with Josephson phase qubits. *Nat. Phys.*, 6(6):409–413, Jun 2010. ISSN 1745-2481. DOI: 10.1038/nphys1639.

[10] S. Blanes, F. Casas, and J. Ros. Extrapolation of symplectic Integrators. *Celestial Mech. Dyn. Astron.*, 75:149–161, 1999. DOI: 10.1023/A:1008364504014.

[11] S. Bravyi, D. Gosset, R. König, and M. Tomamichel. Quantum advantage with noisy shallow circuits. *Nat. Phys.*, 16(10): 1040–1045, Oct 2020. ISSN 1745-2481. DOI: 10.1038/s41567-020-0948-z.

[12] A. Browaeys and T. Lahaye. Many-body physics with individually controlled Rydberg atoms. *Nat. Phys.*, 16(2):132–142, Feb 2020. ISSN 1745-2481. DOI: 10.1038/s41567-019-0733-z.

[13] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, 3 edition, 7 2016. ISBN 9781119121534. DOI: 10.1002/9781119121534. Chapter 33.

[14] Z. Cai and S. C. Benjamin. Constructing smaller Pauli twirling sets for arbitrary error channels. *Sci. Rep.*, 9(1):11281, Aug 2019. ISSN 2045-2322. DOI: 10.1038/s41598-019-46722-7.

[15] E. Campbell. Random compiler for fast hamiltonian simulation. *Phys. Rev. Lett.*, 123:070503, Aug 2019. DOI: 10.1103/PhysRevLett.123.070503.

[16] A. Carrera Vazquez and S. Woerner. Efficient state preparation for quantum amplitude estimation. *Phys. Rev. Applied*, 15: 034027, Mar 2021. DOI: 10.1103/PhysRevApplied.15.034027.

[17] I. 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, Oct 2022. DOI: 10.1103/PhysRevResearch.4.043027.

[18] A. Chiesa, F. Tacchino, M. Grossi, P. Santini, I. Tavernelli, D. Gerace, and S. Car-

retta. Quantum hardware simulating four-dimensional inelastic neutron scattering. *Nat. Phys.*, 15(5):455–459, May 2019. ISSN 1745-2481. DOI: 10.1038/s41567-019-0437-4.

[19] A. M. Childs. On the relationship between continuous- and discrete-time quantum walk. *Commun. Math. Phys.*, 294(2): 581–603, Mar 2010. ISSN 1432-0916. DOI: 10.1007/s00220-009-0930-1.

[20] A. M. Childs and R. Kothari. Limitations on the simulation of non-sparse Hamiltonians. *Quantum Info. Comput.*, 10(7):669–684, July 2010. ISSN 1533-7146. DOI: 10.5555/2011373.2011380.

[21] A. M. Childs and R. Kothari. Simulating sparse Hamiltonians with star decompositions. In *Proc. 5th Conf. Theory Quantum Comput. Commun. Cryptogr.*, pages 94–103, 2011. ISBN 3-642-18072-8, 978-3-642-18072-9. DOI: 10.5555/1946127.19461355.

[22] A. M. Childs and Y. Su. Nearly optimal lattice simulation by product formulas. *Phys. Rev. Lett.*, 123:050503, Aug 2019. DOI: 10.1103/PhysRevLett.123.050503.

[23] A. M. Childs and N. Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Info. Comput.*, 12(11–12):901–924, November 2012. ISSN 1533-7146. DOI: 10.5555/2481569.2481570.

[24] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman. Exponential algorithmic speedup by a quantum walk. In *Proc. 35th Annu. ACM Symp. Theory Comput.*, pages 59–68, 2003. ISBN 1-58113-674-9. DOI: 10.1145/780542.780552.

[25] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su. Toward the first quantum simulation with quantum speedup. *PNAS*, 115(38):9456–9461, 2018. ISSN 0027-8424. DOI: 10.1073/pnas.1801723115.

[26] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu. Theory of Trotter error with commutator scaling. *Phys. Rev. X*, 11(1): 011020, February 2021. DOI: 10.1103/PhysRevX.11.011020.

[27] S. A. Chin. Multi-product splitting and Runge-Kutta-Nyström integrators. *Celestial Mech. Dyn. Astron.*, 106(4):391–406, Apr 2010. ISSN 1572-9478. DOI: 10.1007/s10569-010-9255-9.

[28] C. Cîrstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger. Variational fast forwarding for quantum simulation beyond the coherence time. *npj Quantum Inf.*, 6(1):82, Sep 2020. ISSN 2056-6387. DOI: 10.1038/s41534-020-00302-0.

[29] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Adv. Comput. Math.*, 5(4): 329–359, 1996. DOI: 10.1007/BF02124750.

[30] A. Di Paolo, P. Kl. Barkoutsos, I. Tavernelli, and A. Blais. Variational quantum simulation of ultrastrong light-matter coupling. *Phys. Rev. Research*, 2:033364, Sep 2020. DOI: 10.1103/PhysRevResearch.2.033364.

[31] E. F. Dumitrescu, A. J. McCaskey, G. Hagen, G. R. Jansen, T. D. Morris, T. Papenbrock, R. C. Pooser, D. J. Dean, and P. Lougovski. Cloud quantum computing of an atomic nucleus. *Phys. Rev. Lett.*, 120:210501, May 2018. DOI: 10.1103/PhysRevLett.120.210501.

[32] N. Earnest, C. Tornow, and D. J. Egger. Pulse-efficient circuit transpilation for quantum applications on cross-resonance-based hardware. *Phys. Rev. Research*, 3: 043088, Oct 2021. DOI: 10.1103/PhysRevResearch.3.043088.

[33] D. J. Egger and F. K. Wilhelm. Multimode circuit quantum electrodynamics with hybrid metamaterial transmission lines. *Phys. Rev. Lett.*, 111:163601, Oct 2013. DOI: 10.1103/PhysRevLett.111.163601.

[34] S. Endo, Q. Zhao, Y. Li, S. Benjamin, and X. Yuan. Mitigating algorithmic errors in a Hamiltonian simulation. *Phys. Rev. A*, 99:012334, Jan 2019. DOI: 10.1103/PhysRevA.99.012334.

[35] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan. Variational quantum simulation of general processes. *Phys. Rev. Lett.*, 125:010501, Jun 2020. DOI: 10.1103/PhysRevLett.125.010501.

[36] P. K. Faehrmann, M. Steudtner, R. Kueng, M. Kieferova, and J. Eisert. Randomizing multi-product formulas for Hamiltonian simulation. *Quantum*, 6:806, September 2022. ISSN 2521-327X. DOI: 10.22331/q-2022-09-19-806.

[37] R. P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21(6):467–

488, Jun 1982. ISSN 1572-9575. DOI: 10.1007/BF02650179.

[38] L. E. Fischer, D. Miller, F. Tacchino, P. Kl. Barkoutsos, D. J. Egger, and I. Tavernelli. Ancilla-free implementation of generalized measurements for qubits embedded in a qudit space. *Phys. Rev. Research*, 4: 033027, Jul 2022. DOI: 10.1103/PhysRevResearch.4.033027.

[39] A. Frisk Kockum, A. Miranowicz, S. De Liberato, S. Savasta, and F. Nori. Ultrastrong coupling between light and matter. *Nat. Rev. Phys.*, 1(1):19–40, Jan 2019. ISSN 2522-5820. DOI: 10.1038/s42254-018-0006-2.

[40] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner. Simultaneous Perturbation Stochastic Approximation of the Quantum Fisher Information. *Quantum*, 5:567, October 2021. ISSN 2521-327X. DOI: 10.22331/q-2021-10-20-567.

[41] G. García-Pérez, M. A. C. Rossi, B. Sokolov, F. Tacchino, P. Kl. Barkoutsos, G. Mazzola, I. Tavernelli, and S. Maniscalco. Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms. *PRX Quantum*, 2:040342, Nov 2021. DOI: 10.1103/PRXQuantum.2.040342.

[42] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng. Digital zero noise extrapolation for quantum error mitigation. In *IEEE Int. Conf. on QCE*, pages 306–316, 2020. DOI: 10.1109/QCE49297.2020.00045.

[43] C. Gross and I. Bloch. Quantum simulations with ultracold atoms in optical lattices. *Science*, 357(6355):995–1001, 2017. DOI: 10.1126/science.aal3837.

[44] C. Hadfield, S. Bravyi, R. Raymond, and A. Mezzacapo. Measurements of quantum Hamiltonians with locally-biased classical shadows. *Commun. Math. Phys.*, 391(3): 951–967, May 2022. ISSN 1432-0916. DOI: 10.1007/s00220-022-04343-8.

[45] M. Hagan and N. Wiebe. Composite quantum simulations. 2022. URL `https://arxiv.org/abs/2206.06409`.

[46] G. Hao Low, V. Kliuchnikov, and N. Wiebe. Well-conditioned multiproduct Hamiltonian simulation. 2019. URL `https://arxiv.org/abs/1907.11679`.

[47] A. W. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103: 150502, Oct 2009. DOI: 10.1103/PhysRevLett.103.150502.

[48] A. Hashim, R. K. Naik, A. Morvan, J. L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor. *Phys. Rev. X*, 11:041039, Nov 2021. DOI: 10.1103/PhysRevX.11.041039.

[49] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer. Zero-noise extrapolation for quantum-gate error mitigation with identity insertions. *Phys. Rev. A*, 102:012426, Jul 2020. DOI: 10.1103/PhysRevA.102.012426.

[50] A. Hoorfar and M. Hassani. Inequalities on the Lambert W function and hyperpower function. *JIPAM*, 9, 01 2008. URL `http://eudml.org/doc/130024`.

[51] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.*, 16(10):1050–1057, Oct 2020. ISSN 1745-2481. DOI: 10.1038/s41567-020-0932-7.

[52] S. Indrajeet, H. Wang, M.D. Hutchings, B.G. Taketani, F. K. Wilhelm, M.D. LaHaye, and B.L.T. Plourde. Coupling a superconducting qubit to a left-handed metamaterial resonator. *Phys. Rev. Applied*, 14: 064033, Dec 2020. DOI: 10.1103/PhysRevApplied.14.064033.

[53] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, and *et al.* Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Sci. Technol.*, 6(2):025020, mar 2021. DOI: 10.1088/2058-9565/abe519.

[54] H. Kamakari, S. N. Sun, M. Motta, and Austin J. Minnich. Digital quantum simulation of open quantum systems using quantum imaginary–time evolution. *PRX Quantum*, 3:010320, Feb 2022. DOI: 10.1103/PRXQuantum.3.010320.

[55] Naoki Kanazawa, Daniel J. Egger, Yael Ben-Haim, Helena Zhang, William E. Shanks, Gadi Aleksandrowicz, and Christopher J. Wood. Qiskit experiments: A python pack-

age to characterize and calibrate quantum computers. *J. Open Source Softw.*, 8(84): 5329, 2023. DOI: 10.21105/joss.05329.

[56] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, Mar 2019. ISSN 1476-4687. DOI: 10.1038/s41586-019-1040-7.

[57] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala. Scalable error mitigation for noisy quantum circuits produces competitive expectation values. *Nat. Phys.*, 19(5):752–759, May 2023. DOI: 10.1038/s41567-022-01914-3.

[58] R. LaRose, A. M., S. Kaiser, P. J. Karalekas, A. A. Alves, P. Czarnik, M. El Mandouh, M. H. Gordon, Y. Hindy, A. Robertson, P. T., N. Shammah, and W. J. Zeng. Mitiq: A software package for error mitigation on noisy quantum computers. *Quantum*, 6:774, August 2022. ISSN 2521-327X. DOI: 10.22331/q-2022-08-11-774.

[59] D. Layden. First-order Trotter error from a second-order perspective. *Phys. Rev. Lett.*, 128:210501, May 2022. DOI: 10.1103/PhysRevLett.128.210501.

[60] A. J. Leggett, S. Chakravarty, A. T. Dorsey, Matthew P. A. Fisher, A. Garg, and W. Zwerger. Dynamics of the dissipative two-state system. *Rev. Mod. Phys.*, 59:1–85, Jan 1987. DOI: 10.1103/RevModPhys.59.1.

[61] Y. Li and S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X*, 7: 021050, Jun 2017. DOI: 10.1103/PhysRevX.7.021050.

[62] D. A. Lidar. *Review of Decoherence-Free Subspaces, Noiseless Subsystems, and Dynamical Decoupling*, pages 295–354. John Wiley & Sons, Ltd, 2014. ISBN 9781118742631. DOI: https://doi.org/10.1002/9781118742631.ch11.

[63] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996. DOI: 10.1126/science.273.5278.1073.

[64] E. Magesan and J. M. Gambetta. Effective Hamiltonian models of the cross-resonance gate. *Phys. Rev. A*, 101:052308, May 2020. DOI: 10.1103/PhysRevA.101.052308.

[65] A. D. McLachlan. A variational solution of the time-dependent Schrodinger equation. *Mol. Phys.*, 8(1):39–44, 1964. DOI: 10.1080/00268976400100041.

[66] A. Miessen, P. J. Ollitrault, and I. Tavernelli. Quantum algorithms for quantum dynamics: A performance study on the spin-boson model. *Phys. Rev. Research*, 3:043212, Dec 2021. DOI: 10.1103/PhysRevResearch.3.043212.

[67] M. Mohseni, A. T. Rezakhani, and D. A. Lidar. Quantum-process tomography: Resource analysis of different strategies. *Phys. Rev. A*, 77:032322, Mar 2008. DOI: 10.1103/PhysRevA.77.032322.

[68] Y. Ouyang, D. R. White, and E. T. Campbell. Compilation by stochastic Hamiltonian sparsification. *Quantum*, 4:235, February 2020. ISSN 2521-327X. DOI: 10.22331/q-2020-02-27-235.

[69] B. Pokharel, N. Anand, B. Fortman, and D. A. Lidar. Demonstration of fidelity improvement using dynamical decoupling with superconducting qubits. *Phys. Rev. Lett.*, 121:220502, Nov 2018. DOI: 10.1103/PhysRevLett.121.220502.

[70] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023. URL `https://doi.org/10.5281/zenodo.2562111`.

[71] L. F. Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Philos. Trans. R. Soc. London, Ser. A*, 210:307–357, 1911. ISSN 02643952. URL `http://www.jstor.org/stable/90994`.

[72] S. Sachdev. *Quantum Phase Transitions*. Cambridge University Press, 2 edition, 2011. DOI: 10.1017/CBO9780511973765.

[73] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta. Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Phys. Rev. A*, 93: 060302, Jun 2016. DOI: 10.1103/PhysRevA.93.060302.

[74] A. Sidi. *Practical Extrapolation Methods: Theory and Applications*. Cambridge Uni-

versity Press, 2003. ISBN 9780511546815. DOI: 10.1017/CBO9780511546815.

[75] N. Stamatopoulos, D. J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner. Option Pricing using Quantum Computers. *Quantum*, 4:291, July 2020. ISSN 2521-327X. DOI: 10.22331/q-2020-07-06-291.

[76] J. P. T. Stenger, N. T. Bronn, D. J. Egger, and D. Pekker. Simulating the dynamics of braiding of Majorana zero modes using an IBM quantum computer. *Phys. Rev. Research*, 3:033171, Aug 2021. DOI: 10.1103/PhysRevResearch.3.033171.

[77] N. Sundaresan, I. Lauer, E. Pritchett, E. Magesan, P. Jurcevic, and J. M. Gambetta. Reducing unitary and spectator errors in cross resonance with optimized rotary echoes. *PRX Quantum*, 1:020318, Dec 2020. DOI: 10.1103/PRXQuantum.1.020318.

[78] M. Suzuki. General theory of fractal path integrals with applications to many-body theories and statistical physics. *J. Math. Phys.*, 32(2):400–407, 1991. DOI: 10.1063/1.529425.

[79] K. Temme, S. Bravyi, and J. M. Gambetta. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.*, 119: 180509, Nov 2017. DOI: 10.1103/PhysRevLett.119.180509.

[80] The IBM Decision Optimization on Cloud team. DOCPLEX, 2022. URL `https://pypi.org/project/docplex/`.

[81] C. Tornow, N. Kanazawa, W. E. Shanks, and D. J. Egger. Minimum quantum runtime characterization and calibration via restless measurements with dynamic repetition rates. *Phys. Rev. Applied*, 17: 064061, Jun 2022. DOI: 10.1103/PhysRevApplied.17.064061.

[82] V. Tripathi, H. Chen, M. Khezri, K. W. Yip, E. M. Levenson-Falk, and D. A. Lidar. Suppression of crosstalk in superconducting qubits using dynamical decoupling. *Phys. Rev. Appl.*, 18:024068, Aug 2022. DOI: 10.1103/PhysRevApplied.18.024068.

[83] A. Vazquez Carrera, R. Hiptmair, and S. Woerner. Enhancing the quantum linear systems algorithm using richardson extrapolation. *ACM Trans. Quantum Comput.*, 3(1), jan 2022. ISSN 2643-6809. DOI: 10.1145/3490631.

[84] J. J. Wallman and J. Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A*, 94: 052325, Nov 2016. DOI: 10.1103/PhysRevA.94.052325.

[85] K. Wan, M. Berta, and E. T. Campbell. Randomized quantum algorithm for statistical phase estimation. *Phys. Rev. Lett.*, 129:030503, Jul 2022. DOI: 10.1103/PhysRevLett.129.030503.

[86] D. Wecker, B. Bauer, B. K. Clark, M. B. Hastings, and M. Troyer. Gate-count estimates for performing quantum chemistry on small quantum computers. *Phys. Rev. A*, 90:022305, Aug 2014. DOI: 10.1103/PhysRevA.90.022305.

[87] J. Weidenfeller, L. C. Valor, J. Gacon, C. Tornow, L. Bello, S. Woerner, and D. J. Egger. Scaling of the quantum approximate optimization algorithm on superconducting qubit based hardware. *Quantum*, 6: 870, December 2022. ISSN 2521-327X. DOI: 10.22331/q-2022-12-07-870.

[88] Y. Yongxin. Quantum many-body scars on QC, 2022. URL `https://gitlab.com/QANED/rydberg_mfim`.

[89] X. Yuan, S. Endo, Q. Zhao, Ying Li, and S. Benjamin. Theory of variational quantum simulation. *Quantum*, 3:191, 2019. DOI: 10.22331/q-2019-10-07-191.

[90] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature*, 551(7682):601–604, Nov 2017. ISSN 1476-4687. DOI: 10.1038/nature24654.

[91] C. Zoufal, D. Sutter, and S. Woerner. Error bounds for variational quantum time evolution. 2021. URL `https://arxiv.org/abs/2108.00022`.

A Error cancellations in multi-product formulas

We illustrate how errors in multi-product formulas cancel with the first-order formula. The example is easily adapted to the 2χ-order. We write $S_1^{k_j}(t/k_j)$ as a power series in terms of the er-

ror [27]

$$S_1^{k_j}\left(\tfrac{t}{k_j}\right) = e^{-iHt} + \sum_{n=1}^{\infty} A_n t^{n+1}/k_j^n. \tag{13}$$

Here, the A_n are matrices that depend only on commutators of the H_j. In particular, they are independent of the time t and the Trotter exponents k_j. Therefore, the lower-order error terms can be cancelled out by a careful choice of extrapolation weights a_j that are independent of the Hamiltonian. The A_n are thus irrelevant in subsequent calculations. For the MPF $M_{l,1}(t)$ to approximate e^{-iHt} without bias, the extrapolation weights must satisfy $\sum_{j=1}^{l} a_j = 1$. Furthermore, to cancel the first $l-1$ error terms in S_1 we impose the constraint $\sum_{j=1}^{l-1} a_j/k_j^{1+n} = 0$ for $n = 0, ..., l-2$. Therefore, the resulting error is given by

$$\begin{aligned} M_{l,1}(t) &= \sum_{j=1}^{l} a_j S_1^{k_j}\left(\tfrac{t}{k_j}\right) \\ &= \sum_{j=1}^{l} a_j \left(e^{-iHt} + \sum_{n=1}^{l-1} A_n \frac{t^{n+1}}{k_j^n} + \sum_{n=l}^{\infty} A_n \frac{t^{n+1}}{k_j^n} \right) \\ &= e^{-iHt} + \mathcal{O}\left(\frac{t^{l+1}}{k_1^l}\right). \end{aligned} \tag{14}$$

B Comparison with LCU

Here, we compare the circuit depth of MPFs implemented by classically combining expectation values to MPFs implemented as LCUs with $l = 3$ extrapolation points as example. The LCU method implements an operation proportional to the linear combination $\kappa U + V$ for some unitary gates U, V and $\kappa \geq 0$. A linear combination of three unitaries is implemented recursively as $\alpha_1(\alpha_2 U_1 + U_2) + U_3$. Similarly, we can implement an operation proportional to $\sum_{j=1}^{3} a_j S_1^{k_j}(t/k_j)$ by choosing the appropriate coefficients $\alpha_1, \alpha_2 \geq 0$, shown in Fig. 5. We count the number of CNOT gates for the Ising model described in Sec. 4.4 required by each approach. The first-order Trotter formula implementing a five qubit Ising model requires four R_{ZZ} gates and five R_X gates. Thus, the deepest circuit of a classical combination will contain $4k_3$ R_{ZZ} gates. The number of different gates required by an LCU implementation is given in Table 2. The table also shows the number of CNOT gates required to implement each gate, as transpiled by Qiskit into the native gate set. The LCU approach would therefore require $4(k_1 + k_2) \cdot (20 + 7) + 4k_3 \cdot (8 + 2) + 2 \cdot 2 = 108(k_1 + k_2) + 40k_3 + 4$ CNOTs and two additional qubits. By contrast, our approach only requires $4k_3$ CNOTs. For $l = 3$, this is thus a reduction of the number of CNOT gates by a factor of $[108(k_1 + k_2) + 40k_3 + 4]/4k_3$. With, e.g., $\vec{k} = [1, 2, 7]$ this corresponds to 22× less CNOT gates than an LCU implementation.

Table 2: Number of times each gate is required by an MPF with $l = 3$ implemented as an LCU or as a classical linear combination. The second column shows the number of CNOT gates required to implement each gate. Here we do not include pulse-efficient transpilation. Here, we do not explicitly include pulse-efficient transpilation but treat the cost of the $R_Z Z$ gate as a single CNOT gate.

Gate	CNOTs	LCU	Class.
R_{ZZ}	1	0	$4k_3$
1-controlled R_{ZZ}	8	$4k_3$	0
1-controlled single-qubit U	2	$4k_3 + 2$	0
2-controlled R_{ZZ}	20	$4(k_1 + k_2)$	0
2-controlled R_X	7	$4(k_1 + k_2)$	0

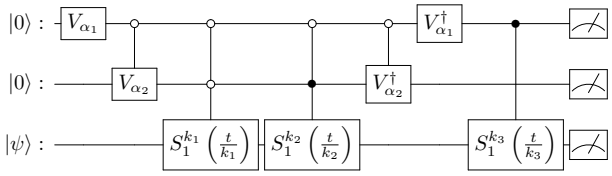


Figure 5: Circuit implementing an MPF with $l = 3$ as an LCU. For a five spin Ising model the LCU method therefore requires seven qubits for $l = 3$. Here, each first-order PF $S_1^{k_j}$ requires $4k_j$ R_{ZZ} gates.

C Extrapolation properties

A crucial step in our scheme is to compute expectation values with the S_χ^k's in Eq. (5) instead of $M_{l,\chi}$, since it allows us to directly compute the extrapolated observable from the measured expectation values without measuring cross-terms in the MPF. Here, cross-terms correspond to $\langle\psi| S_\chi^{k_l} O S_\chi^{k_j} |\psi\rangle$ for $k_j \neq k_l$ when computing the expectation of an observable O with Eq. (5). The proof included below for completeness follows similar arguments as those from Refs. [34, 83].

Let $S_\chi^{k_j} = U + E_j$, where $U = e^{-iHt}$ is the matrix exponential we want to approximate via

a PF and $E_j = \sum_i \alpha_{\eta_i} \frac{t^{\eta_i+1}}{k_j^{\eta_i}}$ is the error incurred in the approximation, where α_k denotes matrix commutators. Let O denote the observable of interest and $a_1, ..., a_l$ the extrapolation weights that satisfy Eq. (6). We will prove that if

$$\sum_{j=1}^{l} a_j S_\chi^{k_j} = \sum_{j=1}^{l} a_j U + \sum_{j=1}^{l} a_j \sum_i \alpha_{\eta_i} \frac{t^{\eta_i+1}}{k_j^{\eta_i}} = U + \mathcal{O}\left(k_1^{-n}\right), \tag{15}$$

where $\eta_{l-1} < n \in \mathbb{Z}$, then

$$\sum_{j=1}^{l} a_j \langle\psi| \left(S_\chi^{k_j}\right)^\dagger O S_\chi^{k_j} |\psi\rangle = \langle\psi| U^\dagger O U |\psi\rangle + \mathcal{O}\left(k_1^{-n}\right). \tag{16}$$

Eq. (16) shows that by ignoring cross-terms we still obtain an approximation with an $\mathcal{O}(k_1^{-n})$ error.

With $S_\chi^{k_j} = U + E_j$ the expectation value $\langle O\rangle = \langle\psi| \left(S_\chi^{k_j}\right)^\dagger O S_\chi^{k_j} |\psi\rangle$ is

$$\langle O\rangle = \langle\psi| U^\dagger O U |\psi\rangle + \langle\psi| U^\dagger O E_j |\psi\rangle + \langle\psi| E_j^\dagger O U |\psi\rangle + \langle\psi| E_j^\dagger O E_j |\psi\rangle. \tag{17}$$

Next, the first order in E_j error terms is

$$\begin{aligned}\epsilon &= \sum_{j=1}^{l} a_j \langle\psi| U^\dagger O E_j |\psi\rangle \\ &= \sum_{j=1}^{l} a_j \sum_i \langle\psi| U^\dagger O \alpha_{\eta_i} \frac{t^{\eta_i+1}}{k_j^{\eta_i}} |\psi\rangle \\ &= \sum_i \langle\psi| U^\dagger O \alpha_{\eta_i} t^{\eta_i+1} |\psi\rangle \sum_{j=1}^{l} \frac{a_j}{k_j^{\eta_i}} \\ &= \mathcal{O}\left(k_1^{-n}\right),\end{aligned} \tag{18}$$

where the last equality comes from combining Eq. (6), i.e. the first $l-1$ error terms cancel out, and Eq. (15), i.e. $\alpha_n \frac{t^{n+1}}{k_j^n} = \mathcal{O}(k_1^{-n})$. A similar result can be shown for $\langle\psi| E_j^\dagger O U |\psi\rangle$ and $\langle\psi| E_j^\dagger O E_j |\psi\rangle$. Thus, Eq. (16) holds and extrapolation can be applied directly to the measured observables.

D Sampling noise of a Bernoulli variable

We illustrate the asymptotic behaviour of the error term $\mathcal{O}\left(\|\vec{a}\|_1 \varepsilon'\right)$ from Eq. (9) restated below

$$\sum_{j=1}^{l} a_j \left(E_j + \varepsilon(k_j) + \varepsilon'\right) = E + \mathcal{O}\left(\frac{t^{2(l+\chi)+1}}{k_1^{2(l+\chi)}}\right) + \mathcal{O}\left(\|\vec{a}\|_1 \varepsilon'\right).$$

We consider as example the task of sampling a Bernoulli discrete random variable $\mathcal{X}$ and computing the probability p that $\mathcal{X} = 1$ by taking M samples. We thus have $E_j = E = \mathbb{E}[\mathcal{X}]$ and $\varepsilon(k_j) = 0$ for all j. The number of samples allows us to control the sampling noise as $\varepsilon'(M) = 1/\sqrt{M}$ on average. We use the extrapolation weights $\vec{a}$ of an ill-conditioned MPF built from the PF S_2 and by taking $k_j = j$ as the sequence of Trotter exponents. Indeed, this setting uniquely defines the sequences $\vec{a}$ by Eq. (6) whose 1-norm scales exponentially with the number of extrapolation points l, i.e., $\|\vec{a}\|_1 \in e^{\Omega(l)}$ [46], see Fig. 6(a). We expect the sampling error to behave as

$$\sum_{j=1}^{l} a_j \left(E + \frac{1}{\sqrt{M}}\right) = E + \mathcal{O}\left(\frac{\|\vec{a}\|_1}{\sqrt{M}}\right), \tag{19}$$

i.e. ill-conditioned sequences have an error that grows exponentially with l. Adding extrapolation points therefore increases the error, see Fig. 6(b). Here, each line corresponds to a different l and we estimate $E_j(M) \coloneqq E_j + \varepsilon'(M)$ for $j = 1, \cdots, l$ by sampling M times from $\mathcal{X}$. Thus, each point in the plot corresponds to the value

$$E^l(M) \coloneqq \sum_{j=1}^{l} a_j E_j(M). \tag{20}$$

This example illustrates how a non-careful choice of Trotter exponents exponentially increases the sampling error of a quantum algorithm. Note that the example above can be seen as a weighted average of l times M samples with weights $\vec{a}$. A simple average is recovered with $a_j = l^{-1}$.

E Optimal Trotter exponents

We obtain sequences of Trotter exponents by solving an optimization problem in Python that

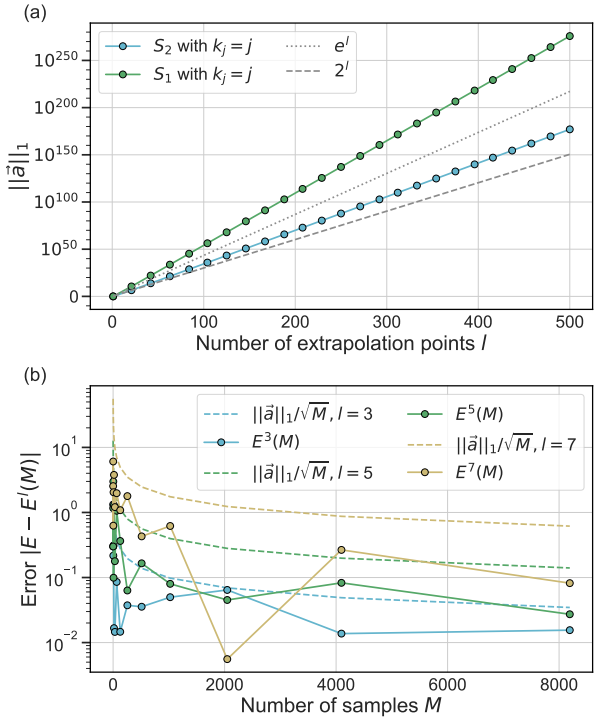


Figure 6: Illustration of an ill-conditioned MPF. (a) Condition number of ill-conditioned MPFs with base PFs S_1 and S_2 when taking Trotter exponents $k_j = j$. (b) Error $\left|E - E^l(M)\right|$ obtained from sampling a Bernoulli variable with $p = 0.3$ using Eq. (20). Different values of l are plotted as different lines.

minimizes $\|\vec{a}\|_1$. We set Eq. (6) as the constraints for the optimization problem. Then, from all the possible sequences of Trotter exponents we accept those for which the condition number returned by the optimization algorithm is below a set threshold. All the results presented here satisfy $\|\vec{a}\|_1 \leq 3$. The Trotter exponents and extrapolation weights used in the Ising simulation experiments are given in Table 3. They were obtained with DOCPLEX [80] using the code shown below.

```
from docplex.mp.model import Model

# npts := number of extrapolation
    points
num_rows = npts - 1
h = np.ones((num_rows, npts))
# 2nd order
h[0, :] = [1/f/f for f in factors]
# 1st order
h[0, :] = [1/f for f in factors]
for i in range(1, num_rows):
    h[i, :] = h[0, :]**(i+1)

# create empty model
mdl = Model()

# create decision variables
a = mdl.continuous_var_list(
        [f'a{i}' for i in range(npts
            )],
        lb=-np.inf,
    )
x = mdl.continuous_var_list(
        [f'x{i}' for i in range(npts
            )],
    )

# minimize sum_i x[i]
mdl.minimize(mdl.sum(x))

# x[i] >= a[i] and x[i] >= -a[i]
for i in range(num_points):
    mdl.add_constraint(x[i] >= a[i])
    mdl.add_constraint(x[i] >= -a[i
        ])

# sum(a) == 1
mdl.add_constraint(mdl.sum(a) == 1)

# for each i: sum_j a_i*h_ij = 0
for i in range(num_rows):
    row = [a[j] * h[i, j]
           for j in range(npts)]
    row_sum = mdl.sum(row)
    mdl.add_constraint(row_sum == 0)
```

F Quantum circuit scaling of the spin-boson model

Mapping the spin-boson Hamiltonian to a qubit-lattice shows that the approximation error of PFs grows linearly with the number of qubits N_q as $\mathcal{O}\left(N_q t^2/k_j\right)$ [18, 22, 66]. Now, we calculate for the spin-boson model the scaling of the deepest circuit needed by an MPF that classically combines expectation values. We start from the Taylor expansion of a symmetric PF

$$S_{2\chi}^{k_j}\left(\tfrac{t}{k_j}\right) = e^{-iHt} + \sum_{n=\chi}^{\infty} A_{2n} t^{2n+1}/k_j^{2n}, \quad (21)$$

where the A_n are matrices that depend on the commutators of the H_j [27, 46]. Using this notation we bound the MPF error by $\mathcal{O}\left(\frac{\|A_{2l+1}\| t^{2l+1}}{(2l+1)!}\right)$ [46, 83]. Ref. [22] shows that $\|A_n\| \leq N_q$ by the triangle inequality and by grouping the Hamiltonian terms H_j in an even-odd pattern. We therefore update the MPF error

Table 3: Trotter exponents and extrapolation weights used to build the S_1-based MPFs to approximate the Ising Hamiltonian. We define well-conditioned MPFs those that satisfy $\|\vec{a}\|_1 \leq 3$.

Trotter exponents $\vec{k}$	Extrapolation weights $\vec{a}$	$\|\vec{a}\|_1$
	Well-conditioned MPF	
$[1,2]$	$[-1.0, 2.0]$	3
$[1,3]$	$[-0.5, 1.5]$	2
$[2,4]$	$[-1.0, 2.0]$	3
$[2,5]$	$[2/3, 5/3]$	2.33
$[1,2,6]$	$[0.2, -1, 1.8]$	3
$[1,2,7]$	$[1/6, -4/5, 49/30]$	2.6
	Ill-conditioned MPF	
$[6,7]$	$[-6, 7]$	13
$[3,4,5,6,7]$	$[27/8, -128/3, 375/4, -216, 2401/24]$	518.33
$[1,2,3,4,5,6,7]$	$[1/720, -8/15, 243/16, -1024/9, 15625/48, -3888/10, 117649/720]$	1007.22

Table 4: Total number of repetitions of the different circuits in the deepest Trotter circuit needed to simulate spin-boson systems of different sizes $(M, n^{\max})$ to achieve the target accuracy $\varepsilon_t \in \{10^{-2}, 10^{-3}, 10^{-4}\}$. Here, M is the number of bosons and $n^{\max}$ is the maximum occupation number in each bosonic mode. We refer to Ref. [66] for details on the quantum implementation. However, compared to Ref. [66] we count the number of different circuits $e^{-iH_j t}$ instead of the number of first-order Trotter layers. For example, if $H = H_1 + H_2$ we count $e^{-iH_1 t}e^{-iH_2 t}$ as two repetitions while Miessen *et al.* count it as one. The number in parenthesis in the S_2 and S_4 columns represent l in the MPF, i.e. in column 4, row 7, 13(2) means that an MPF with S_2 requires two circuits with the deepest circuit repeating $e^{-iH_j t}$ 13 times to achieve an error below 10^{-2}. The values for S_1 and VQA are taken from Ref. [66].

		$\varepsilon_t = 10^{-2}$			$\varepsilon_t = 10^{-3}$			$\varepsilon_t = 10^{-4}$			
N_q	$(M,\, n^{\max})$	S_1	S_2	S_4	S_1	S_2	S_4	S_1	S_2	S_4	VQA
3	(1,1)	48	11(2)	31(2)	170	23(2)	41(2)	552	43(2)	81(2)	2
4	(1,2)	69	21(2)	61(3)	240	29(3)	101(2)	777	45(3)	121(2)	3
5	(1,3)	69	25(3)	61(3)	219	41(2)	101(2)	690	113(1)	101(2)	6
	(2,1)	54	13(2)	51(2)	196	23(2)	71(3)	646	57(2)	81(3)	4
7	(2,2)	102	13(2)	121(2)	333	77(2)	181(2)	1068	89(2)	201(2)	9
	(3,1)	64	21(2)	51(2)	228	21(2)	51(2)	750	21(2)	181(3)	6
11	(2,4)	117	41(3)	101(4)	372	81(1)	161(2)	1167	145(1)	401(1)	12
	(5,1)	96	13(2)	71(2)	314	49(2)	91(2)	1008	135(1)	211(3)	10

Table 5: Total number of repetitions of the different circuits in the deepest Trotter circuit needed to simulate spin-boson systems of different sizes $(M, n^{\max})$ to achieve the target accuracy $\varepsilon_t \in \{10^{-2}, 10^{-3}, 10^{-4}\}$. The setting is the same as in Table 4 but here we compare only different PFs.

		$\varepsilon_t = 10^{-2}$			$\varepsilon_t = 10^{-3}$			$\varepsilon_t = 10^{-4}$		
N_q	$(M,\, n^{\max})$	S_1	S_2	S_4	S_1	S_2	S_4	S_1	S_2	S_4
3	(1,1)	48	27	111	170	45	121	552	79	171
4	(1,2)	69	37	221	240	93	241	777	165	261
5	(1,3)	69	37	221	219	65	121	690	113	261
	(2,1)	54	37	111	196	63	121	646	111	171
7	(2,2)	102	61	161	333	109	221	1068	189	361
	(3,1)	64	37	81	228	63	121	750	111	201
11	(2,4)	117	57	221	372	81	241	1167	145	401
	(5,1)	96	45	101	314	77	141	1008	135	241

bound to $\mathcal{O}\left(N_q/(2l+1)!\right)$ for which we also assumed $t \leq 1$. Thus, for a target error tolerance ε_t

the asymptotic scaling of l satisfies the inequality

$$\frac{N_q}{(2l+1)!} < \varepsilon_t. \tag{22}$$

For simplicity, we write $y \coloneqq N_q/\varepsilon_t$ and $x \coloneqq (2l+1)$ such that Eq. (22) reads $x! > y$. By Stirling's approximation, i.e. $x! \sim \sqrt{2\pi x}(x/e)^x$, the inequality $x! > y$ holds if we set $y = (x/e)^x$. Rearranging this equation yields

$$\frac{1}{x} y^{\frac{1}{x}} = e^{-1}, \tag{23}$$

which has the solution

$$2l + 1 = x = \frac{\ln y}{W\left(\frac{\ln y}{e}\right)}, \tag{24}$$

where $W(\cdot)$ denotes the Lambert function [29]. For $z > e$ it holds that $\ln z - \ln\ln z < W(z) < \ln z$ [50]. We can thus substitute $W(e^{-1}\ln y)$ by $\ln(e^{-1}\ln y) - \ln\ln(e^{-1}\ln y) = \mathcal{O}(\ln\ln y)$ to find that the number of extrapolation points l scales as

$$\mathcal{O}\left(\frac{\log(N_q/\varepsilon_t)}{\log\log(N_q/\varepsilon_t)}\right). \tag{25}$$

Equation (25) holds for $t \leq 1$. For an arbitrary time, similarly to the main text, we rescale $\vec{k}$ by $\alpha = t$ to ensure that $t/k_j \leq 1$. Furthermore, for $t < 1$ the number of Trotter steps of the deepest circuit k_l is bounded by l^2 for the sequences in Ref. [46] which becomes $k_l \leq \alpha l^2$ when we rescale $\vec{k}$ to accommodate for $t > 1$. We thus conclude that k_l scales as

$$\mathcal{O}\left(\frac{t\log^2(N_q/\varepsilon_t)}{\log^2\log(N_q/\varepsilon_t)}\right) \tag{26}$$

as presented in the main text. We further emphasize that Eq. (26) is a worst case upper bound which may yet be improved, for instance, by better choices of α.

G Dynamical Decoupling

Dynamical decoupling (DD) applies carefully chosen gate sequences on idle qubits to suppress, e.g., decoherence [69] and cross-talk [82]. These sequences are chosen to cancel system-environment interactions to a given order in time-dependent perturbation theory [62, 69]. Since, in practice, the duration of the two-qubit cross-resonance gates on IBM hardware varies across qubit pairs and because the circuit structure has layers of one- and two-qubit gates there are often periods where some qubits idle and therefore suffer decoherence and energy relaxation. The simplest DD sequence replaces sufficiently long idle periods of length τ with the gate sequence $\frac{\tau'}{4} - X - \frac{\tau'}{2} - X - \frac{\tau'}{4}$, as done in Ref. [17, 53]. Here, the idle time τ' is the idle time without DD less the duration of the X gate. We did not see a significant impact of DD on the hardware results and therefore we did not use it. We conjecture this is because single-qubit gates have a $4{\cdot}10^{-4}$ error and that the idle times are small. Indeed, the duration of the average delay which can accommodate the DD sequence in the circuit without the added ZNE gates is 252 ns. Under a T_1 time of 125 μs, the average T_1 across *ibmq_montreal*, this delay can be seen as an error of $1-e^{-0.292/125} \approx 2{\cdot}10^{-3}$, i.e. comparable to twice the single-qubit gate error. If the idle times in the circuit were longer we believe that it would have been advantageous to use DD.

H Calibration of R_{ZX} pulses

Computing the dynamics of a quantum system with a Trotter expansion requires many high accuracy small rotations. To avoid decoherence the R_{ZX} gates are built from scaled cross-resonance gates following the linear methodology in Ref. [32, 76]. Scaling calibrated CNOT gates to implement $R_{ZX}(\theta)$ with small values of θ, i.e. up to ~ 50 mrad, results in Gaussian cross-resonance pulses without a flat-top and a θ-dependent amplitude. Since the rotation angle θ implemented by the cross-resonance drive depends non-linearly on the pulse amplitude [4, 32, 64], we mitigate rotation errors with a fine amplitude calibration experiment [81]. Here, the gate sequence $[R_{ZX}(\theta)]^n|00\rangle$ is repeated for a variable n and a desired θ after which the target qubit is measured and both qubits are reset, see Fig. 7(a). The underlying pulse sequences, implemented by linearly scaling the cross-resonance amplitude, creates an effective rotation $[R_{ZX}(\mathrm{d}\theta+\theta)]^n$ in the measured population, see Fig. 7(b) which we fit to the function

$$y(n) = \frac{A}{2} e^{-n/\tau}\cos\left(2\pi n[\mathrm{d}\theta + \theta] + \pi\right) + B. \tag{27}$$

We chose the maximum n as $3/\theta$ to have approximately three full oscillations. The small rotation angles we are dealing with thus require pulse sequences with a duration which can exceed the T_1 times of the qubits. To produce good fits, as measured by χ^2, we add the damping factor $e^{-n/\tau}$ in Eq. (27). We run the fine amplitude calibration for all qubit pairs and 20 values of θ linearly spaced between 0 and the largest θ for which the scaled CR gates have no flat-top. The measured $\mathrm{d}\theta$ yields a correction factor $\theta/(\mathrm{d}\theta+\theta)$ that we multiply with the amplitude of the scaled CR gates, see Fig. 8. We observe that this amplitude calibration improves the value of the measured expectation values, see Fig. 9.

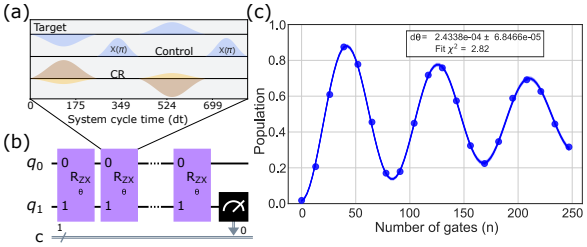


Figure 7: Fine amplitude calibration of a R_{ZX} gate with a small rotation angle. (a) Pulse schedule of a single R_{ZX} gate. (b) Gate sequence designed to amplify rotation errors. (c) Population of the target qubit as a function of the number of gates fit to a damped cosinusoidal oscillation.

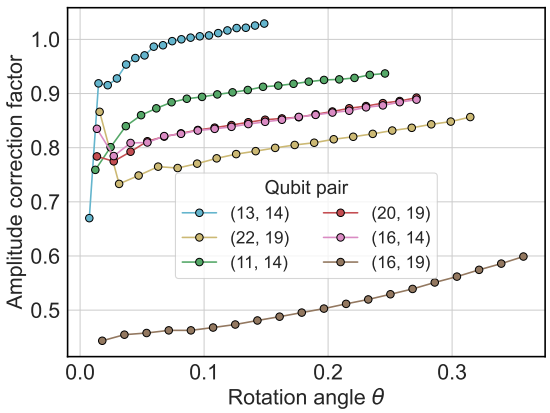


Figure 8: Small R_{ZX} rotation angle amplitude correction factors $\theta/(\mathrm{d}\theta+\theta)$ obtained with fine amplitude calibrations on *ibmq_montreal* on May 19th 2022.

I Comparison with Mitiq

Ref. [17] simulates an Ising model using a similar approach to error mitigation, that is, Pauli Twirling, a pulse efficient transpilation of the R_{ZZ} gates and dynamical decoupling. By contrast, we do not apply dynamical decoupling and implement ZNE by linearly scaling cross-resonance pulses. Ref. [17] uses the Mitiq package [58] to implement random gate folding, i.e. repeating random layers of R_{ZZ} gates. Here, we use the example described in Sec. 4.4 as a benchmark to compare both approaches in Fig. 10. We observe that ZNE implemented by folding gates improves the data with respect to the unmitigated results in some cases. Instead, implementing ZNE by scaling CR pulses improves the results in almost all cases.

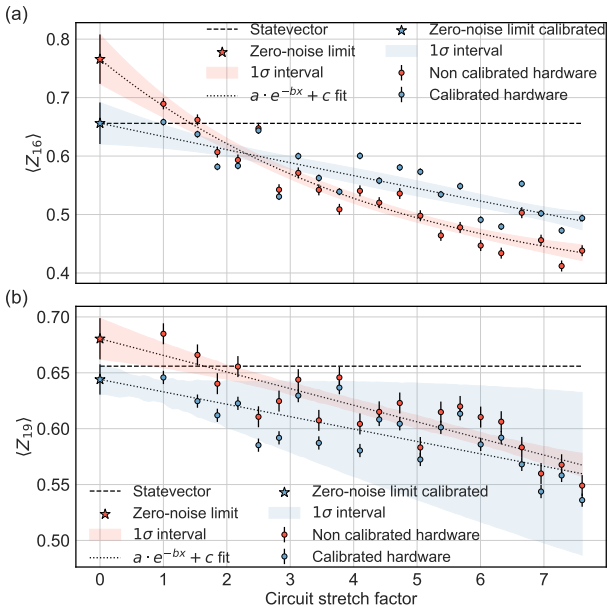


Figure 9: Comparison of the local ZNE magnetization $\langle Z_i\rangle$ using amplitude calibrated (blue markers) versus uncalibrated (red markers) R_{ZX} gates. That data is for the five-spin Ising model described in Sec. 4.4 simulated on *ibmq_montreal*. Pannels (a) and (b) correspond to qubit 16 and 19, respectively. Each point is the average of ten different runs and the error bars show twice the standard deviation.

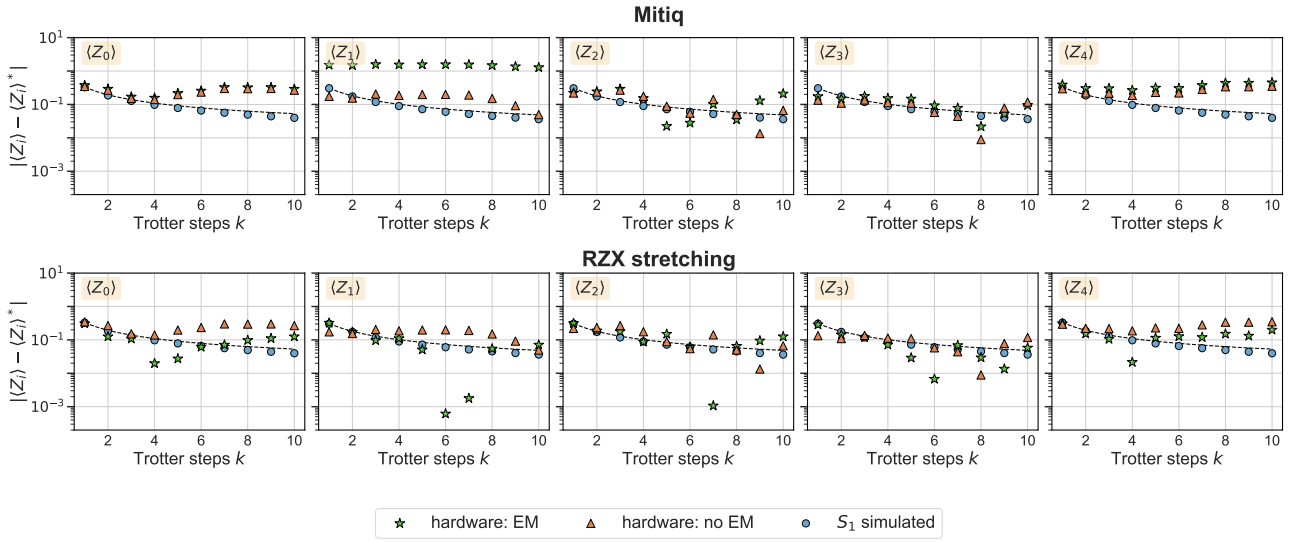


Figure 10: Comparison of two ZNE methodologies on *ibmq_montreal*. The top row compares simulation, hardware with no error mitigation (hardware: no EM) and hardware with readout error mitigation, PT, DD and ZNE as implemented by Mitiq using the code publicly available from Ref. [88]. In the bottom row the hardware error mitigated results were obtained using readout error mitigation, PT and the ZNE approach described in Sec. 4.3. In all cases the R_{ZZ} gates are obtained by pulse-efficient transpilation as described in Sec. 4.2.

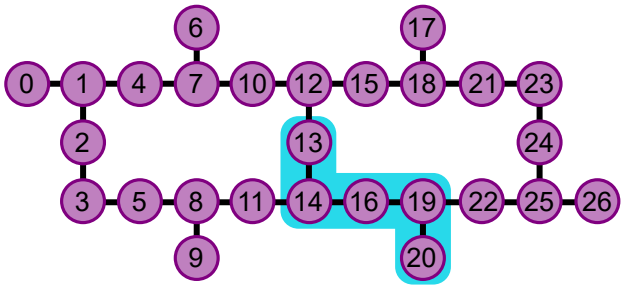


Figure 11: Layout of *ibmq_montreal*. We choose five qubits in a line such that the fidelity decay resembles the T_1 decay. We chose to work with qubits $\{13, 14, 16, 19, 20\}$.

Table 6: Properties of the qubits and gates used to simulate the Ising Hamiltonian as reported by *ibmq_montreal* on May 29th 2022. The T_1 and T_2 times are reported in μs. The CNOT error in each row is the error between the qubit of that row and the qubit of the next row.

			Error rate (%) of		
Qubit	T_1	T_2	$\sqrt{X}$	CNOT	Measure
13	118	122	0.018	0.57	0.74
14	133	256	0.019	0.79	1.09
16	82	22	0.052	1.68	1.83
19	160	167	0.023	0.86	1.21
20	85	133	0.035	–	1.50
average	115.6	140	0.029	0.97	1.27

J Hardware properties

Error mitigation techniques can significantly decrease the physical noise in the measured observables. However, their efficacy is limited and error rates are mostly hardware related. In particular, choosing *good* qubits is a simple but worthwhile strategy towards achieving accurate results. Here *good* qubits refers to low error rates in the readout, single- and two-qubit gates. Most of these values are provided by the IBM Quantum backends [1]. However, our circuits comprise mostly $R_{ZX}(\theta)$ gates, which are the main source of noise. The fidelity of these gates is not provided by the backends as they are built from scaled-CR gates [32]. Therefore, we perform quantum process tomography of scaled-CR $R_{ZX}(\theta)R_{ZX}(-\theta)$ gates using Qiskit Experiments [55] on *ibmq_montreal*, see Fig. 11.

For each qubit pair in the device we calculate the fidelity of the gate sequence for 20 values of $\theta \in (0.5, 20)$, shown as dots in Fig. 12, and plot it together with the T_1 decay of the two qubits, represented as solid lines. We then inspect the plots and define as *good* qubit pairs those for which the gate fidelity decay resembles the T_1 decay. Finally, we select 4 good pairs that form five linearly connected qubits, depicted in Fig. 11. The properties of *ibmq_montreal* are shown in Table 6.

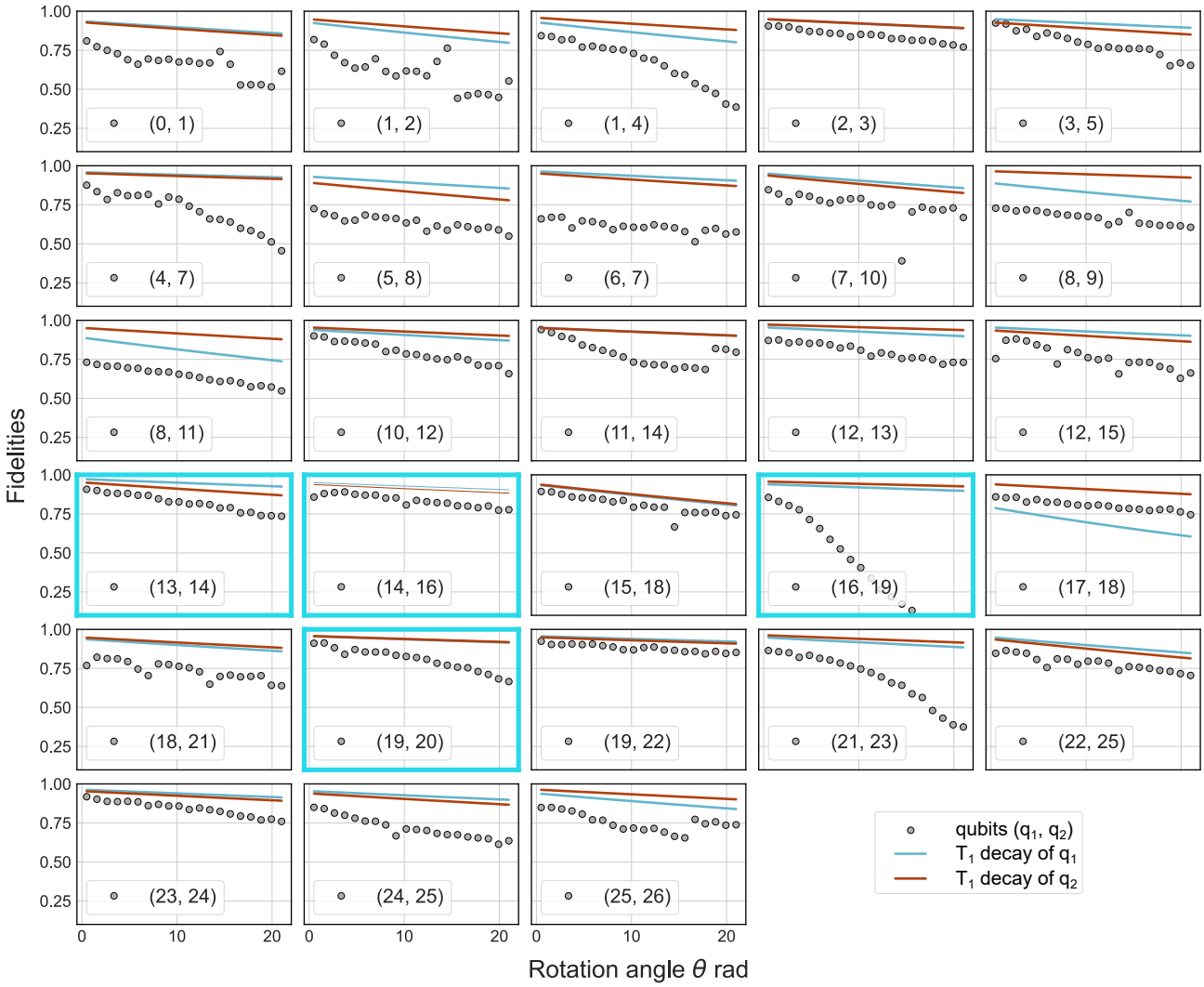


Figure 12: Fidelity of $R_{ZX}(\theta)R_{ZX}(-\theta)$ on *ibmq_montreal* on May 29th 2022 for different rotation angles $\theta \in (0.5, 21)$. Each panel corresponds to a different qubit pair (q_1, q_2). The dots represent the fidelity of the gates, while the solid lines the T_1 decay of the qubits. Here we chose to work with qubits $\{13, 14, 16, 19, 20\}$ and the corresponding panels are highlighted in blue.