Quantum dynamics simulations beyond the coherence time on NISQ hardware by variational Trotter compression

Noah F. Berthusen,[1, 2, ∗] Thaís V. Trevisan,[1, 3] Thomas Iadecola[1, 3, †] and Peter P. Orth[1, 3, ‡]

[1]*Ames Laboratory, Ames, Iowa 50011, USA*
[2]*Department of Electrical and Computer Engineering, Iowa State University, Ames, Iowa 50011, USA*
[3]*Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA*
(Dated: December 24, 2021)

We demonstrate a post-quench dynamics simulation of a Heisenberg model on present-day IBM quantum hardware that extends beyond the coherence time of the device. This is achieved using a hybrid quantum-classical algorithm that propagates a state using Trotter evolution and then performs a classical optimization that effectively compresses the time-evolved state into a variational form. When iterated, this procedure enables simulations to arbitrary times with an error controlled by the compression fidelity and a fixed Trotter step size. We show how to measure the required cost function, the overlap between the time-evolved and variational states, on present-day hardware, making use of several error mitigation methods. In addition to carrying out simulations on real hardware, we investigate the performance and scaling behavior of the algorithm with noiseless and noisy classical simulations. We find the main bottleneck in going to larger system sizes to be the difficulty of carrying out the optimization of the noisy cost function.

I. INTRODUCTION

Simulating quantum dynamics of interacting many-body systems is one of the main potential applications of quantum computing, going back to Feynman's visionary work on simulating physics with quantum computers [1]. For quantum many-body systems evolving under a generic nonintegrable Hamiltonian H, such simulations are exponentially hard on classical devices due to the growth of entanglement during the simulation time [2, 3]. Quantum processing units (QPUs), on the other hand, can efficiently simulate quantum dynamics with resources scaling polynomially in the number of particles N and in the simulation time t. Since Lloyd's original proposal [4] to use a first-order Lie-Trotter product formula to decompose the unitary time-evolution operator into elementary gates, several other methods have been discovered that exhibit a more favorable asymptotic scaling of the required number of gates at or near the optimal scaling of $O(Nt)$ [5–12]. Concrete resource estimates, however, show that this often comes at the cost of requiring a large number of ancilla qubits or having a large constant overhead of gates [9]. Therefore, product formulas still remain a preferred choice on noisy intermediate-scale quantum (NISQ) hardware [13] due to their simplicity and competitive performance for physically relevant (e.g., local) Hamiltonians that fulfill additional properties [14].

Despite this progress, direct quantum simulation algorithms face a critical drawback on NISQ QPUs: the limited coherence time of the device imposes an upper bound on the depth of the quantum circuits that can be implemented with high fidelity. This in turn upperbounds the simulation time t that can be reached before the output is overwhelmed by errors—on current hardware, this time scale is roughly of order one in the natural units imposed by the Hamiltonian being simulated [15]. Quantum simulation beyond the coherence time using "fast-forwarding" algorithms [16–19] is possible, but only for nongeneric Hamiltonians, including those that can be mapped to free fermions [20].

Variational quantum algorithms (VQAs) provide a promising route to overcome the coherence-time obstacle for generic Hamiltonians [21, 22]. Their starting point is to represent the wavefunction by a variational ansatz $|\psi(\boldsymbol{\vartheta})\rangle = \prod_{j=1}^{\mathcal{N}} e^{-i\vartheta_j A_j} |\psi_0\rangle$ with $\mathcal{N}$ real parameters $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_{\mathcal{N}})$, initial state $|\psi_0\rangle$, and Hermitian generators A_j, which are often chosen to be single Pauli strings or sums of commuting Pauli strings. To simulate quantum dynamics, the parameters ϑ_j are updated in a way that allows the variational state to follow the exact dynamics; different algorithms to perform the update have been proposed in the literature.

In one class of VQAs, one derives an equation of motion for the variational parameters, $\sum_j M_{ij}\dot{\vartheta}_j(t) = V_i$ [23–25], by extremizing the distance between the variational state and the exact time-evolved state at every (infinitesimally small) time step in the evolution. Here, the matrix M_{ij} and vector V_j must be obtained at every time step by performing measurements on the QPU, and the main bottleneck of the algorithm is the large number of measurements. Since the number of components M_{ij} scales quadratically with the number of variational parameters $\mathcal{N}$, the number of measurements $\mathcal{M}$ scales as $\mathcal{M} \propto \mathcal{N}^2$.

An alternative approach that reduces the number of measurements is to determine the time-dependence of the variational parameters $\boldsymbol{\vartheta}(t)$ by optimizing the state overlap fidelity $\mathcal{F} = |\langle\psi(\boldsymbol{\vartheta})|\psi(t)\rangle|^2$ between the variational state and the exact time-evolved state $|\psi(t)\rangle$ [26, 27]. In

practice, the exact state can be approximated with high fidelity, for example, using Trotter evolution over a time interval τ sufficiently short to guarantee a desired high accuracy at finite gate depth. Such a "variational Trotter compression" (VTC) approach combines the desirable aspects of Trotter evolution with the shallow gate depth requirements of variational methods.

Here, we expand on previous works that explored VTC algorithms on classical computers [26, 27] by performing the first implementation and benchmarking of a VTC algorithm on real quantum hardware. We explicitly demonstrate a simulation of quantum quench dynamics in a Heisenberg spin chain beyond the coherence time on the IBM Santiago QPU. To achieve this goal, we employ several error mitigation strategies: we combine zero-noise extrapolation (ZNE) with Pauli twirling [23, 28] and symmetry-based post-selection. At the current levels of noise on the QPU, we find it advantageous to avoid the gradient based optimization used in the variant of the VTC algorithm described in Ref. [27], and instead choose the non-gradient based genetic optimization algorithm CMA-ES [29]. We compare different ways to compute the overlap fidelity on QPUs, and find that a method based on a "double time contour" circuit that foregoes the use of ancilla qubits and non-local SWAP gates [30, 31] is most suited for current NISQ hardware. Finally, we demonstrate the scalability of the VTC algorithm both on noisy and noiseless quantum simulators up to system sizes of $M = 6$ and $M = 11$ sites, respectively.

The remainder of the paper is organized as follows: in Sec. II we present the general VTC algorithm and point out differences with previous works [26, 27]. We also discuss different quantum circuit implementations of the optimization cost function. Then, in Sec. III, we apply VTC to simulate post quench dynamics in antiferromagnetic Heisenberg chains. In Sec. III A, we introduce the model and describe the variational ansatz we use, and in Sec. III B we benchmark the capability of the ansatz to capture the exact time-evolved state. We then execute VTC on different classical simulators and on real quantum hardware. In Sec. III C, we present results obtained on a statevector simulator. In Sec. III D, we use an ideal circuit simulator to consider sample noise due to a finite number of quantum measurements, and in Sec. III E, we show results from classical noisy circuit simulations, where the noise model parameters correspond to the IBM Santiago backend. Finally, in Sec. III F, we demonstrate our key result: a quantum dynamics simulation beyond the qubit coherence time on the real IBM Santiago device. We conclude and discuss future research directions in Sec. IV.

II. ALGORITHM

In this section we describe the different parts of the variational Trotter compression (VTC) algorithm.

A. Choice of ansatz

First, one needs to choose a particular way to build the variational ansatz $|\psi(\boldsymbol{\vartheta})\rangle$. The main part of the algorithm is independent of the form of the ansatz, but requires that it is able to faithfully represent the time-evolved state up to a desired fidelity. Since the entanglement content of the state increases during time evolution, the complexity of the ansatz and the number of required parameters grows with time t (and also with system size M). We characterize this growth in detail for a specific model in Sec. III. A key insight from previous works is that the growth with t of the number of variational parameters can be much slower (linear versus exponential) than in purely classical algorithms based on matrix product states (MPSs) [26], giving the VTC algorithm a potential quantum advantage.

The variational ansatz can either be chosen to have a fixed form or to be adaptively modified during the computation [25, 32, 33]. Here, we choose a fixed ansatz that is inspired by the Hamiltonian variational ansatz (HVA) [34, 35] and that takes a layered form,

$$|\psi(\boldsymbol{\vartheta})\rangle = U(\boldsymbol{\vartheta}) |\psi_0\rangle = \prod_{l=1}^{\ell} \prod_{i=1}^{N} e^{-i\vartheta_{l,i} A_i} |\psi_0\rangle . \quad (2.1)$$

Here, ℓ denotes the number of layers and N is the number of variational parameters per layer. We choose the Hermitian operators A_i to be sums of commuting Pauli strings that correspond to the terms in the Hamiltonian being simulated. Finally, $|\psi_0\rangle$ is an initial state that can be chosen arbitrarily; in particular, it does not need to correspond to the state of the system at the initial time t_i, which we denote by $|\psi_i\rangle$.

B. Variational Trotter compression

The central part of the algorithm is the variational compression step. The key idea is to accurately propagate the variational state over a short, finite time interval τ using Trotter evolution, and then to re-express the resulting state, $U_{\text{trot}}(\tau) |\psi(\boldsymbol{\vartheta}_t)\rangle$, in variational form via optimization of the overlap cost function

$$\mathcal{C} = |\langle\psi(\boldsymbol{\vartheta}_{t+\tau})|U_{\text{trot}}(\tau)|\psi(\boldsymbol{\vartheta}_t)\rangle|^2. \quad (2.2)$$

The cost function is measured on the QPU and feeds into a classical optimization routine that determines the updated variational parameters as $\hat{\boldsymbol{\vartheta}}_{t+\tau} = \arg\min_{\boldsymbol{\vartheta}_{t+\tau}} \mathcal{C}$. The complete compression step is then iterated until the final time t_f.

The maximal value of τ in a single compression step is determined by the number of Trotter steps $n_{\max}$ that can be executed on the noisy QPU, and by the desired Trotter error threshold ε, which is a function of $\tau/n_{\max}$ [4, 36]. There exist different methods to determine the overlap cost function on a QPU, which will be discussed in more

detail below in Sec. II C. The simplest method concatenates the Trotter and the variational ansatz circuits and calculates the probability that the system returns to the initial state, *i.e.*,

$$\mathcal{C} = |\langle\psi_0|U^\dagger(\boldsymbol{\vartheta}_{t+\tau})U_{\text{trot}}(\tau)U(\boldsymbol{\vartheta}_t)|\psi_0\rangle|^2. \qquad (2.3)$$

Here, $U_{\text{trot}}(\tau)$ is a (first-order) Trotter circuit using n steps. We find this straightforward method to be robust under noise and preferable on NISQ hardware with nearest-neighbor connectivity compared to alternative methods that use ancilla qubits or multiple SWAP operations [30, 31]. Since the variational and Trotter circuits are applied consecutively in Eq. (2.3), the depth of the ansatz, which can be expressed as the number of layers ℓ, must be taken into account when choosing the number of Trotter steps n. Loosely speaking, one must have that $2\ell + n < n_{\max}$, where $n_{\max}$ is the maximal number of Trotter steps that can be executed within the finite coherence time of the device.

Different methods can be used for the classical optimization of the overlap cost function $\mathcal{C}$. Ref. [27] proposes a gradient-based optimizer, where the gradient is measured directly using a quantum circuit, while Ref. [26] performs the optimization purely classically using tensor network based methods. We have compared the performance of gradient- and non-gradient-based methods and find that non-gradient-based methods such as CMA-ES [29] are preferable for noisy cost functions when considering realistic noise levels present on current hardware.

C.   Quantum circuit implementation of cost function measurement

As mentioned previously, when implementing the VTC algorithm on a real quantum device one needs to decide how to evaluate the overlap cost function $\mathcal{C}$ in Eq. (2.2), since the full wavefunction is not accessible. We consider two methods to evaluate $\mathcal{C}$ that are physically realizable on current NISQ devices, and we compare their resource scaling. Since the fidelities of two-qubit entangling gates are much lower than those of single-qubit gates on current hardware, we focus on the scaling of the number of two-qubit gates.

1.   SWAP test circuit

One common method for computing the overlap of two quantum states is the SWAP test [30, 37], whose circuit implementation is shown in Fig. 1. Here, the two crosses denote swap operations between pairs of individual qubits in the two quantum registers containing the states $|\Phi\rangle$ and $|\Psi\rangle$. The ancilla expectation value $\langle Z\rangle = |\langle\Psi|\Phi\rangle|^2$ is equal to the overlap of the states and thus serves as a cost function for VTC. Specifically, we initialize the states $|\Phi\rangle = U_{\text{trot}}(\tau)U(\boldsymbol{\vartheta}_t)|\psi_0\rangle$
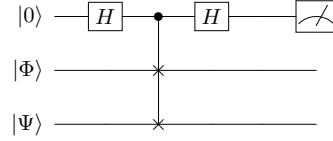

FIG. 1. SWAP-test circuit diagram for computing the overlap between two quantum states $|\Psi\rangle$ and $|\Phi\rangle$ via the expectation value $\langle Z\rangle = |\langle\Psi|\Phi\rangle|^2$ of an ancilla qubit.

and $|\Psi\rangle = U(\boldsymbol{\vartheta}_{t+\tau})|\psi_0\rangle$ and find the optimal parameters $\boldsymbol{\vartheta}_{t+\tau}$ by maximizing $\langle Z\rangle$ for the ancilla.

Let us now estimate the number of two-qubit gates required to implement this circuit on two quantum registers of length M, each corresponding to a system of M qubits. Swapping arbitrary M-qubit states $|\Phi\rangle$ and $|\Psi\rangle$ requires M controlled-SWAP gate operations, each of which can be broken down into two controlled-NOT (CNOT) gates and one Toffoli gate. The Toffoli gate can be further decomposed into a circuit containing 6 CNOTs. Therefore, the swap test incurs a cost of $8M$ CNOTs in total. One major drawback of using the SWAP test on QPUs with local qubit connectivity, however, is that the ancilla must be coupled to every other qubit during the cascading controlled-SWAPs. As the current generation of superconducting QPUs lacks such nonlocal connectivities, one would need to compile a nonlocal SWAP into many local CNOTs, which significantly increases the resource cost of the SWAP test.

2.   Double time contour circuit

The simplest way to evaluate the overlap cost function, which does not require any qubit overhead or ancillas, is to implement a "double time contour" circuit, corresponding to the direct implementation of the overlap matrix element in Fig. 2. The key idea is that the inverse of the updated variational circuit $U^\dagger(\boldsymbol{\vartheta}_{t+\tau})$ effectively unwinds the evolution induced by $U_{\text{trot}}(\tau)U(\boldsymbol{\vartheta}_t)$. The probability $p_{|\psi_0\rangle}$ for the system to end up in its initial state $|\psi_0\rangle$ is maximal for the optimal parameters $\hat{\boldsymbol{\vartheta}}_{t+\tau}$. We find this cost function to be robust to noise under realistic conditions. The circuit is shown in Fig. 2 and consists of a consecutive application of three unitary circuits onto a fixed initial state $|\psi_0\rangle$. For convenience, we choose the initial state $|\psi_0\rangle$ to be a Z-basis state. This method to evaluate $\mathcal{C}$ does not require any additional overhead and the number of required qubits is equal to M, the number of qubits in the system. Note, however, that the depth of the circuit $U^\dagger(\boldsymbol{\vartheta}_{t+\tau})U_{\text{trot}}(\tau)U(\boldsymbol{\vartheta}_t)$ is about 1.5 times larger than that required for the SWAP test.
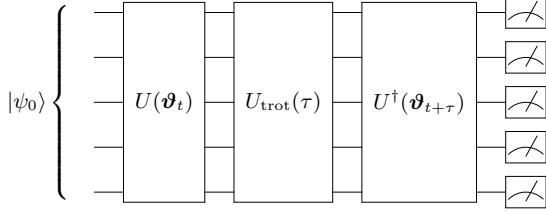
FIG. 2. Double time contour quantum circuit for ancilla-free evaluation of the overlap cost function $\mathcal{C}$ in Eq. (2.3). Circuit is shown for $M = 5$.

III. APPLICATION TO HEISENBERG SPIN CHAIN DYNAMICS

In this section, we apply the variational Trotter compression algorithm to investigate post-quench dynamics in antiferromagnetic Heisenberg chains. We first describe the variational ansatz we employ and then present results using a statevector simulator, ideal and noisy circuit simulators, and real quantum hardware.

A. Model Hamiltonians and variational ansatz

To benchmark the VTC algorithm and to study its robustness with respect to noise, we apply it to investigate quantum quench dynamics in integrable and nonintegrable Heisenberg spin chains. The Hamiltonian of the pure Heisenberg chain, which is integrable, is given by

$$H_0 = \frac{J}{4} \sum_{i=1}^{M} (X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1}) . \quad (3.1)$$

Here, X_i, Y_i, Z_i are Pauli operators at site i, and M denotes the total number of sites of the chain. For concreteness, we will focus in the following on the antiferromagnetic model with $J > 0$, and consider the quantum spin dynamics that arises when initially preparing the system in the classical Néel ground state $|\psi_i\rangle = |0101\cdots\rangle$. The classical Néel state is not an eigenstate of H_0, leading to nontrivial dynamics of $|\psi(t)\rangle = e^{-iH_0 t} |\psi_i\rangle$. To study the difference between simulations of integrable versus nonintegrable dynamics, we also consider an integrability-breaking next-neighbor interaction in the model,

$$H_1 = H_0 + \frac{J}{4} \sum_{i=1}^{M} Z_i Z_{i+2} . \quad (3.2)$$

Unless explicitly stated otherwise, we employ periodic boundary conditions in the following.

Let us now discuss our choice of variational ansatz. For the integrable case, we consider a brickwall-type quantum circuit with ℓ layers (shown in Fig. 3) that we apply to

an initial product state $|\psi_0\rangle = |0101\cdots\rangle$:

$$|\psi(\boldsymbol{\vartheta}^{(\ell)})\rangle = \prod_{l=1}^{\ell} U_{\text{even}}(\boldsymbol{\phi}_l) U_{\text{odd}}(\boldsymbol{\theta}_l) |\psi_0\rangle . \quad (3.3)$$

Here, $\boldsymbol{\vartheta}^{(\ell)} \equiv (\boldsymbol{\theta}_1, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\theta}_\ell, \boldsymbol{\phi}_\ell)$ are $M\ell$ variational parameters and

$$U_{\text{odd}}(\boldsymbol{\theta}_l) = \prod_{j \text{ odd}} e^{-i\,\theta_{l,j}\,(X_j X_{j+1} + Y_j Y_{j+1} + Z_j Z_{j+1})} \quad (3.4\text{a})$$

$$U_{\text{even}}(\boldsymbol{\phi}_l) = \prod_{j \text{ even}} e^{-i\,\phi_{l,j}\,(X_j X_{j+1} + Y_j Y_{j+1} + Z_j Z_{j+1})} \quad (3.4\text{b})$$

are the unitary operators acting on the odd and the even bonds of the chain, respectively. Here, $\theta_{l,j}$ is the jth entry of the parameter vector $\boldsymbol{\theta}_l$, and $\phi_{l,j}$ is the jth entry of $\boldsymbol{\phi}_l$. The quantum circuit has depth 2ℓ, corresponding to the number of unitaries applied to every qubit. Note that the length of the vector $(\boldsymbol{\theta}_l, \boldsymbol{\phi}_l)$ is equal to the number of sites M and $\boldsymbol{\vartheta}^{(\ell)}$ has $M\ell$ components. Additionally, for odd M, the ansatz contains a boundary term U_{boundary} that does not commute with either U_{odd} or U_{even}. A compact quantum circuit representation of the unitary operators in Eqs. (3.4) has been given in Ref. 38 and is shown in Fig. 4.

For the nonintegrable model H_1, we add an additional unitary to each brickwall layer l:

$$U_Z(\boldsymbol{\gamma}_l) = \prod_{j=1}^{M} e^{-i\,\gamma_{l,j}\,Z_j Z_{j+2}} . \quad (3.5)$$

This doubles the number of parameters in the ansatz, which is given by $2M\ell$ in the nonintegrable case.
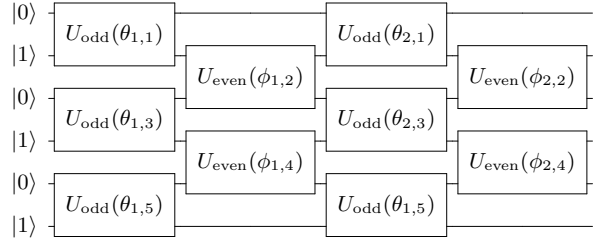

FIG. 3. Quantum circuit implementing the brickwall ansatz $|\psi(\boldsymbol{\vartheta}^{(\ell)})\rangle$ in Eq. (3.3) for $\ell = 2$ and $M = 6$. For simplicity, we show the circuit for OBC. For PBC, there exists an additional even layer gate $U_{\text{even}}(\phi_{l,M})$ between the first and the last qubit.

B. Required number of layers ℓ

To benchmark the ability of the variational ansatz in Eq. (3.3) to represent the exact wavefunction within a desired accuracy ϵ, we determine the minimal number
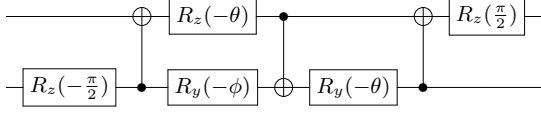
FIG. 4. Quantum circuit implementing the unitary operator $U(\alpha) = \exp\left[-i\alpha(X_j X_{j+1} + Y_j Y_{j+1} + Z_j Z_{j+1})\right]$, where $\theta = \frac{\pi}{2} - 2\alpha$ and $\phi = 2\alpha - \frac{\pi}{2}$.

of layers $\ell_{\min}(t, \epsilon, M)$ needed such that the variational ansatz can represent the exact time-evolved state $|\psi(t)\rangle$ at time t up to an infidelity $1 - \mathcal{F} < \epsilon$. To do this, for each t we numerically minimize the infidelity

$$1 - \mathcal{F}\left(t, \boldsymbol{\vartheta}^{(\ell)}\right) = 1 - |\langle \psi(\boldsymbol{\vartheta}^{(\ell)})|\psi(t)\rangle|^2 \qquad (3.6)$$

over the parameters $\boldsymbol{\vartheta}^{(\ell)}$ to obtain the optimal parameters $\hat{\boldsymbol{\vartheta}}^{(\ell)}$. Repeating this for different values of ℓ, we find $\ell_{\min}(t, \epsilon, M)$ as the smallest ℓ for which the minimal infidelity $1 - \mathcal{F}(t, \hat{\boldsymbol{\vartheta}}^{(\ell)})$ falls below the desired error threshold ϵ. Here, we obtain the exact state $|\psi(t)\rangle = e^{-iHt}|\psi_i\rangle$ via exact diagonalization with $H = H_0$ (or $H = H_1$) and initial state $|\psi_i\rangle = |0101\cdots\rangle$. The optimization is performed using the gradient-based optimizer L-BFGS-B [39].

We first study the time-dependence of the minimal infidelity for a fixed number of layers ℓ. As shown in Fig. 5, $1 - \mathcal{F}(t, \hat{\boldsymbol{\vartheta}}^{(\ell)})$ first increases rapidly after the quench (unless ℓ is sufficiently large) and then saturates at a value that decreases with increasing layer number. This can be understood by comparison to the behavior of the entanglement entropy when tracing out half the system's degrees of freedom [see inset of Fig. 5(b)]: it grows linearly over time until it reaches saturation due to the finite system size. Since each layer of the brickwall quantum circuit couples only nearest-neighbor qubits, the spread of entanglement across the system in the variational state is limited by the total number of layers ℓ. As a result, for a fixed layer number ℓ, the infidelity $1 - \mathcal{F}$ grows as a function of time as the variational ansatz is unable to capture the entanglement that builds up in the system during time evolution.

In the bottom panel of Fig. 5, we show the growth with time of the layer number $\ell^*(t)$ required to keep the infidelity at time t below a fixed threshold that we set to $\epsilon = 1 \times 10^{-4}$ (see also dashed line in the top panel). Like the entanglement entropy, the required layer number grows linearly in time before it reaches saturation. This linear growth in the number of layers and variational parameters provides an opportunity for quantum advantage, since one generically expects state-of-the-art classical techniques based on matrix product techniques to exhibit an exponential scaling with time of the number of parameters [2, 26]. Finally, due to finite system size, the required layer number saturates at long times. We denote the saturation value of the layer number by $\ell^* \equiv \lim_{t \to \infty} \ell^*(t, \epsilon, M)$. In practice, we choose a time

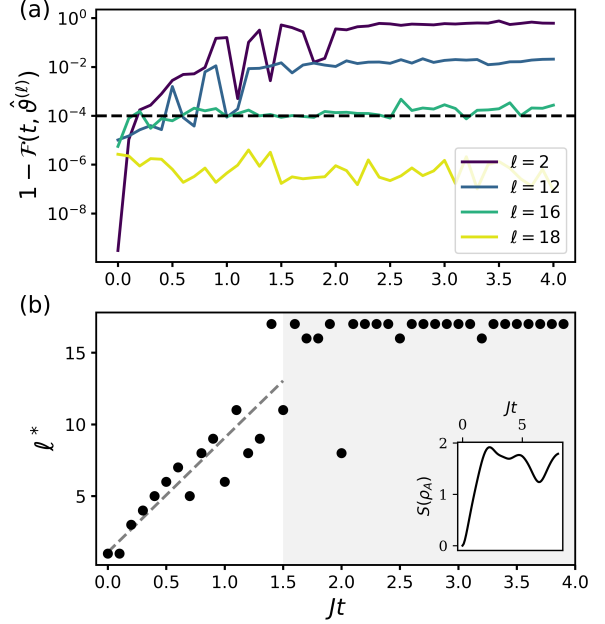FIG. 5. (a) Minimal infidelity $1 - \mathcal{F}(t, \hat{\boldsymbol{\vartheta}}^{(\ell)})$ as a function of time t for different values of ℓ. Results are for the $M = 8$ pure Heisenberg model H_0, Eq. (3.1), using PBC. We obtain the infidelity by performing a numerical optimization of Eq. (3.6) every $0.1Jt$. (b) Required layer number $\ell^*(t) \equiv \ell_{\min}(t, \epsilon = 1 \times 10^{-4}, M = 8)$ as a function of time. The inset shows the dynamics of the half-chain entanglement entropy for comparison.

t_f much larger than the saturation timescale and define $\ell^* \equiv \ell^*(t_f)$. Note that an ansatz with ℓ^* layers is able to represent the time-evolved state out to arbitrary times (see also top panel of Fig. 5).

We now systematically study the dependence of the saturation layer number ℓ^* on system size M and desired error threshold ϵ. We therefore fix a final time $t_f = 50J^{-1}$, which is larger than the saturation time scale of the entanglement entropy for all system sizes we consider. In Fig. 6 we show the minimal infidelity, $1 - \mathcal{F}(t_f, \hat{\boldsymbol{\vartheta}}^{(\ell)})$, as a function of layer number ℓ. We present results for the integrable Heisenberg model H_0 in panel (a) and for the nonintegrable model H_1 in panel (b). As a function of ℓ, the infidelity curves are first relatively flat until a characteristic M-dependent value, where they start to plunge to much smaller values. For definiteness, we set the desired error threshold to $\epsilon = 5 \times 10^{-3}$, which is indicated by the black dashed line in Fig. 6. The insets show the resulting $\ell^*(\epsilon = 5 \times 10^{-3}, M)$ as a function of system size M. We find that ℓ^* grows exponentially with M, which implies an exponential growth of the number of variational parameters $\mathcal{N} \propto M\ell$ that are needed to describe the long time dynamics. For $M = 10$ we find $\ell^*(M = 10) \approx 40$ in the integrable model and $\ell^*(M = 10) \approx 20$ in the non-
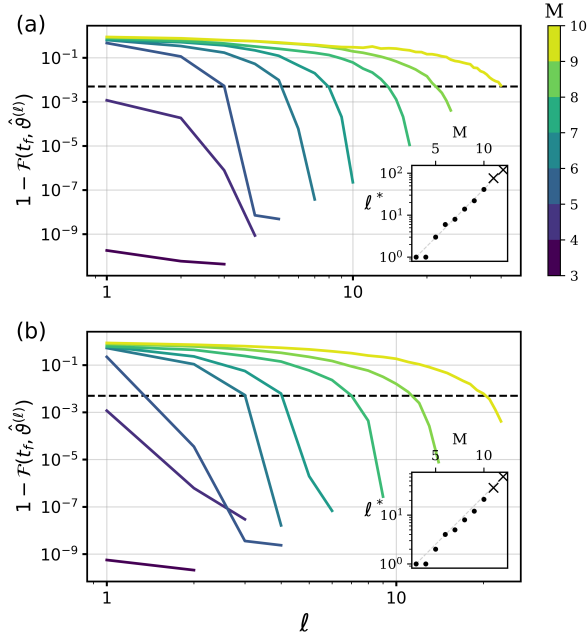
FIG. 6. Minimal infidelity $1 - \mathcal{F}(t_f, \hat{\boldsymbol{\vartheta}}^{(\ell)})$ at late time $t_f = 50J^{-1}$ as a function of layer number ℓ. Different curves are for different system sizes M. Panel (a) is for the pure Heisenberg model H_0, Eq. (3.1), and panel (b) is for the nonintegrable model H_1, Eq. (3.2). Both results are for PBC. The black dashed line shows a fixed error threshold $\epsilon = 5 \times 10^{-3}$, which defines ℓ^*. The insets show that ℓ^* exhibits exponential scaling with system size M. Since we find that $1 - \mathcal{F}$ decays exponentially with increasing ℓ, the values of ℓ^* denoted by a cross are obtained by exponential extrapolation of $1 - \mathcal{F}(\ell)$.

integrable one. In both cases, the number of variational parameters $\mathcal{N} \approx 350$, since the nonintegrable ansatz has twice as many parameters per layer as the integrable one. This exponential scaling behavior at large times constitutes a bottleneck of the approach when considering larger system sizes. Note, however, that at short times we find a favorable linear scaling of ℓ and $\mathcal{N}$ with time t, which provides an opportunity for quantum advantage. Such behavior was also reported previously using a sequential quantum circuit ansatz [26].

C. Statevector simulator results

We now discuss the performance of the VTC algorithm in simulating post-quench dynamics of the antiferromagnetic Heisenberg chain. In the following we focus on the integrable model H_0, since the results of Fig. 6 indicate that the algorithm's performance will not differ substantially between the two cases. In this section, we discuss results obtained using an exact statevector simulator, where we have direct access to the overlap cost function $\mathcal{C}$ in Eq. (2.2). Recall that the VTC algorithm consists of two steps: a propagation step, where we use a first-order Trotter product formula to evolve the state from time t to $t + \tau$, and a compression step, where the time-evolved state is compressed into variational form by numerical optimization of the overlap cost function $\mathcal{C}$.

In Fig. 7, we present results for the state overlap fidelity of the VTC-evolved state with the exact state: $\mathcal{F}(t) = |\langle \psi(\hat{\boldsymbol{\vartheta}}_t) | \psi(t) \rangle|^2$, where $|\psi(t)\rangle = e^{-iH_0 t} |\psi_i\rangle$ and $\hat{\boldsymbol{\vartheta}}_t$ are the optimal variational parameters maximizing the cost function $\mathcal{C}$ at time t. Note that in this and the following sections, we drop the (ℓ) superscript when referring to the variational parameters. Instead, we always use $\ell^* \equiv \lim_{t\to\infty} \ell^*(t, \epsilon, M)$ for a specified ϵ and M. The VTC fidelity $\mathcal{F}(t)$ is shown as a green dashed line in Fig. 7. Here, we focus on simulations at system size $M = 11$ over a long time $t_f = 140J^{-1}$ using a variational ansatz with $\ell = 76$ layers that is able to represent the exact time-evolved state over the full time interval. We find a sizable fidelity $\mathcal{F}(t_f) = 0.83$ at the end of the simulation after performing 9 propagation and compression steps at times $t = m\tau$ with $\tau = 15.2$ and $m = 1, \ldots, 9$. Here, we have chosen $n = \ell$ Trotter steps to propagate the state from time t to $t + \tau$.

This value of $\mathcal{F}(t_f)$ should be contrasted with the fidelity obtained by direct Trotter simulation using the same resources as VTC: the Trotter simulation fidelity $\mathcal{F}_{\text{Trot}}(t) = |\langle \psi_{\text{Trot}}(t) | \psi(t) \rangle|^2$ falls to zero already at time $t = 120J^{-1}$ (grey curve). We note that the direct Trotter simulation uses a fixed number of 3ℓ Trotter steps. This is because the double time contour circuit used to measure the VTC cost function $\mathcal{C}$ is of length 3ℓ if one chooses $n = \ell$ during the VTC propagation step [see Eq. (2.3)].

During the VTC algorithm there are two independent sources of error that occur during the propagation and the compression step, respectively. First, a Trotter error occurs during state propagation that is controlled by the Trotter step size τ/n. Second, a compression error arises that is given by the final (i.e., minimal) value of $1 - \mathcal{C}$ at the end of the numerical optimization. This error is controlled by the optimization parameter ϵ that sets the threshold for convergence. Of course, a smaller value of ϵ makes the optimization more difficult and time consuming. To contrast these two sources of error, we include in Fig. 7 results for a perfect compression (orange line), corresponding to the Trotter simulation fidelity $\mathcal{F}_{\text{Trot}}(t)$ with fixed step size, where the only source of error is the finite, fixed Trotter step size τ/n. In the ideal case of perfect compression, this error could be brought down arbitrarily by reducing τ. However, since we have to perform more compression steps for smaller τ, this will increase the compression error in practice. This leads to our choice of parameters for the VTC simulation in Fig. 7, where the two errors are comparable.

Finally, the main bottleneck of the algorithm in going to larger system sizes is the time needed to perform the compression optimization. Due to the large number of parameters, e.g. $\mathcal{N} = M\ell = 836$ for $M = 11$ and $\ell = 76$, the optimization of the variational ansatz to produce a
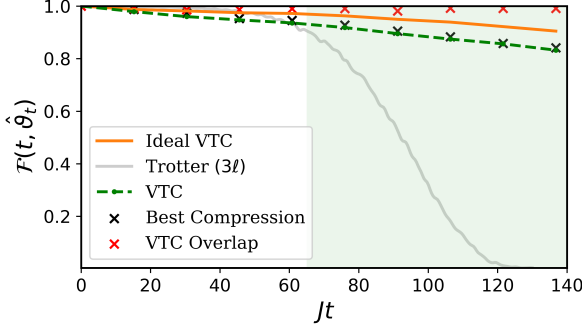
FIG. 7. Statevector simulator results for the VTC state over-lap fidelity $\mathcal{F} = |\langle\psi(\hat{\boldsymbol{\vartheta}}_t)|\psi(t)\rangle|^2$ (green dashed) with the exact state $|\psi(t)\rangle$ during post-quench dynamics in the $M = 11$ pure Heisenberg chain H_0, Eq. (3.1). The system is initially pre-pared in the classical Néel state $|\psi_i\rangle$. We use $\ell = 76$ layers in the variational ansatz, and $n = 76$ Trotter steps in the VTC propagation step from $t \to t + \tau$ with $\tau = 15.2$. The error threshold during compression is set to $\epsilon = 5 \times 10^{-3}$ and we employ the gradient based optimizer L-BFGS-B. The or-ange curve depicts the Trotter simulation fidelity $\mathcal{F}_{\mathrm{Trot}}$ with fixed step size $\tau/n = 0.2$, corresponding to VTC with perfect compression. The grey curve are results from direct Trotter simulations, where we use a fixed number of $3 \times 76 = 228$ Trotter steps for a fair comparison. In the region highlighted in green, $t > 60J^{-1}$, the state fidelity is larger for VTC than for direct Trotter simulation, showing an advantage at larger times. The black crosses denote the fidelity at time t of the best possible Trotter compression state $U_{\mathrm{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle$, i.e., $|\langle\psi(t)|U_{\mathrm{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle|^2$. Red crosses denote the overlap be-tween the best possible compression state and the state to which the algorithm converged during the compression opti-mization step, i.e., $|\langle\psi(\hat{\boldsymbol{\vartheta}}_t)|U_{\mathrm{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle|^2$, which is equal to $\mathcal{C}$ at the end of the optimization.

single compression point in Fig. 7 takes a few hours of CPU time, preventing long-time simulations for larger systems. On the other hand, if one limits the simulation to shorter times, once can reduce the number of layers ℓ in the variational ansatz while keeping the product $\mathcal{N} = M\ell$ constant, and thereby simulate larger systems. This was demonstrated for a mixed-field Ising model using a different classical optimization method based on tensor network techniques in Ref. [26].

D. Ideal quantum circuit simulator results

Next, we investigate the performance of the VTC al-gorithm on a noiseless quantum circuit simulator, where we evaluate the overlap cost function using either the SWAP-test (Fig. 1) or the double time contour circuit (Fig. 2) using a finite number of quantum measurements, or "samples." The fact that the overlap cost function now exhibits sample noise increases the difficulty of the numerical optimization during the compression step. In particular, we find that the non-gradient-based optimizer
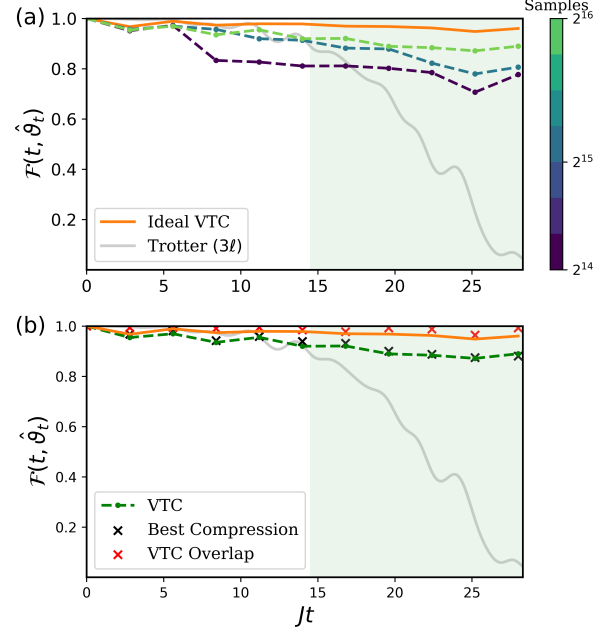

FIG. 8. Optimal fidelity $\mathcal{F}(t, \hat{\boldsymbol{\vartheta}}_t)$ for VTC simulations on an ideal quantum circuit simulator. (a) Dependence of the fi-delity on the number of circuit samples. Curves of different colors correspond to executions of the algorithm using dif-ferent numbers of quantum measurements (samples) for each cost function evaluation, as indicated by the color bar. Re-sults are for $M = 6$, $\ell = n = 7$, $\tau = 2.8$, and compression error threshold $\epsilon = 5 \times 10^{-3}$. We use CMA-ES for the compres-sion optimization. The direct Trotter simulations (grey) use $3\ell = 21$ Trotter steps and are run on a statevector simulator. (b) Analysis of the fidelity data for 2^{16} samples in the vein of Fig. 7. The data points represented by black and red crosses are calculated as described in the caption of Fig. 7.

CMA-ES is more reliable and converges much faster to a minimum of the noisy overlap cost function compared to a gradient-based optimizer such as L-BFGS-B. While the sample noise can in principle be brought arbitrarily low by increasing the number of samples, we here focus on a realistic number of samples, between 2^{14} and 2^{16}, that could be executed on current NISQ hardware.

The presence of sample noise limits the largest sys-tem size we are able to simulate using VTC to $M = 6$. We do not observe any significant difference in the per-formance of the algorithm depending on whether we use the SWAP-test or the double time contour circuit to mea-sure the cost function. As shown in Fig. 8(a), the fidelity for the VTC simulation is consistently larger than that for the noiseless direct Trotter simulation at large times. As expected, the VTC fidelity increases with the num-ber of circuit samples as the sample noise is reduced. Specifically, we find a fidelity larger than 0.88 at times $t > 25J^{-1}$, where the fidelity using direct Trotter sim-ulation with 3ℓ steps has already fallen to zero due to

the accumulation of Trotter error. Note that we use a statevector simulator for the direct Trotter calculations.

To further analyze the algorithm's performance for the maximal number (2^{16}) of samples, we include in Fig. 8(b) the fidelity (with respect to the exact time-evolved state) of the best possible compression during the next propagation period τ. Explicitly, this is given by the overlap $|\langle\psi(t)|U_{\text{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle|^2$; the data are represented as black crosses in the figure. The red crosses indicate the overlap of the best possible compression state, $U_{\text{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle$, with the state to which the algorithm converged; explicitly, this is given by $|\langle\psi(\hat{\boldsymbol{\vartheta}}_t)|U_{\text{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle|^2$. Note that this is equal to the value of the cost function $\mathcal{C}$ at the end of the optimization. An interesting feature relative to the noiseless simulation considered in Fig. 7 is that the VTC algorithm can converge to a state whose fidelity with the exact time-evolved state exceeds that of the "best compression" state $U_{\text{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle$—this is visible in the last data point in Fig. 8(b). This feature is a result of the sample noise impacting the optimization during the compression step. Indeed, for the noiseless statevector simulation, the actual fidelity of the converged VTC state is upper bounded by the "best-compression" value, as expected.

A single compression step for the parameters in Fig. 8 takes up to ten hours of CPU time. The difficulty of the noisy optimization during the compression step constitutes the main bottleneck of the circuit simulations in going to larger system sizes. This bottleneck could be mitigated by parallelizing the circuit evaluations across different CPUs (or indeed QPUs), which would allow the accumulation of more samples to reduce the sample noise. Another possibility worth exploring is to replace the overlap cost function by, e.g., a reduced density matrix fidelity [40] defined only over a subset of the full system where local quantities of interest are to be computed.

E. Noisy circuit simulator results

We now discuss VTC simulation results on a noisy quantum circuit simulator that takes gate imperfections and finite qubit coherence times into account. To connect with the simulations on the real IBM QPU Santiago, which will be discussed below, we choose a noise model with parameters drawn from that chip, as implemented in Qiskit Aer [41]. Due to the increased noise in the simulations arising from gate errors, readout errors, and finite qubit coherence times T_1 and T_2, the numerical optimization during the compression step is even more challenging than for the ideal quantum circuit simulator. We use the non-gradient-based method CMA-ES for the classical optimization, which we find to be more reliable than gradient-based approaches.

To reduce the noise in the cost function, it is absolutely essential to exploit a combination of standard and specifically tailored error mitigation techniques. We employ standard readout error mitigation as built into Qiskit
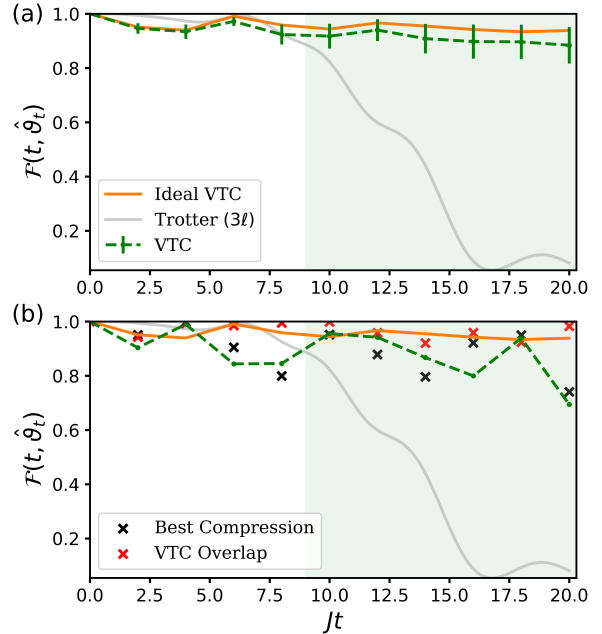

FIG. 9. (a) VTC results on a noisy circuit simulator using error model parameters corresponding to the IBM Santiago backend. We use $M = 3$, $\ell = n = 2$, $\tau = 2$, $\epsilon = 5 \times 10^{-3}$, and 2^{13} samples. The numerical optimization is done using CMA-ES. The green dashed VTC curve shows the mean fidelity found after 50 runs of the algorithm with the standard deviation shown as error bars. The orange curve shows results of noiseless VTC with perfect compression. The grey curve is a noiseless Trotter simulation using $3\ell = 6$ steps. (b) VTC results on the real IBM Santiago device for the same parameters as in panel (a). Each data point on the green VTC curve represents a single run of the compression algorithm. The region highlighted in green demonstrates quantum dynamics simulations beyond the finite coherence time of the QPU. The data points represented by black and red crosses are calculated as described in the caption of Fig. 7.

Ignis, and zero-noise extrapolation (ZNE) with a linear fit, which we implement using the software package Mitiq [42]. We also apply a specific postselection protocol to the results: since the Heisenberg model preserves the magnetization operator $S^z = \sum_{i=1}^{M} Z_i$, only computational basis states with the same expectation value of S^z as the initial Néel state are physically allowed. We thus discard any counts of computational basis states that do not fulfill the total S^z conservation law after performing readout error mitigation.

In Fig. 9(a) we present VTC results for the post-quench dynamics in an $M = 3$ spin chain with open boundary conditions using the noisy quantum circuit simulator. The noise parameters are drawn from the IBM QPU Santiago. We use $\ell = 2$ layers in the variational ansatz, $n = 2$ Trotter steps during the VTC propagation step, and average over 8192 circuit samples. Since the calcula-

tion of each compression point only takes about a minute, we are able to run the noisy simulation 50 times and average over the resulting VTC fidelities (green dashed line). The error bars on the green dashed line correspond to the standard deviation over these 50 runs. We observe that the noisy VTC simulations agree with the ideal VTC results (orange line) within error bars. Importantly, for times larger than $t > 9J^{-1}$ the VTC fidelity is larger than the fidelity of a noiseless direct Trotter simulation with $3\ell = 6$ Trotter steps. While the Trotter fidelity drops to zero around $t \approx 16J^{-1}$, the VTC fidelity remains above 0.9 throughout the full simulation until $t = 20J^{-1}$.

F. Results on IBM QPU Santiago

We now demonstrate quantum dynamics simulations beyond the qubit coherence time on real IBM quantum hardware. This is achieved by running VTC simulations on the IBM QPU Santiago for a system of size $M = 3$ with open boundary conditions. We use a variational ansatz with $\ell = 2$ layers, employ $n = 2$ Trotter steps during the state propagation step of VTC, and average over 2^{13} circuit samples. The complete algorithm is executed on real quantum hardware, i.e., all circuits during the compression optimizations are being executed on the QPU. In addition to the error mitigation techniques described in Sec. III E, we found it essential to apply a Pauli twirling technique, which dresses the two-qubit CNOT gates with random Pauli gates [43]. This is important in order to convert the noise on the real device into a stochastic form, which is necessary to justify ZNE. Without the additional Pauli twirling protocol, we observed that the ZNE error mitigation method was not reliable.

In Fig. 9(b), we present VTC simulation results on the IBM Santiago backend. These results were collected over the course of three days, since calculation of each compression point takes about 2-3 hours, which includes waiting time in the IBM execution queue. Most importantly, we find that at times $t > 9J^{-1}$ the VTC fidelity largely exceeds the fidelity obtained from noiseless direct Trotter simulations that use the same circuit resources. Note that the direct Trotter simulations are performed using statevector simulations and thus do not contain any of the noise that is present in the VTC simulations. While the direct Trotter fidelity vanishes for times larger than $t > 16J^{-1}$, the VTC fidelity remains at a value of 0.8 at that time. This explicitly demonstrates that even though the impact of noise on the real QPU is more severe compared to the noisy simulator, we are able to achieve quantum dynamics simulations beyond the coherence of the device using the VTC algorithm.

In Fig. 9(b) we also analyze the data from the QPU using the methodology of Figs. 7 and 8. We find that the fidelity of the state to which the compression optimization converges can again exceed that of the best possible Trotter compression, i.e., $|\langle\psi(t)|U_{\text{Trot}}|\psi(\hat{\boldsymbol{\vartheta}}_{t-\tau})\rangle|^2$ (black

crosses). This effect, which was also apparent in the results of Fig. 8, is much more pronounced on the real QPU due to the presence of both sample and gate noise. The red crosses again represent the fidelity between the best possible compression and the state found by the algorithm. Over the 10 compression steps, the algorithm was able to recreate the Trotter evolution of the state from the previous compression step with a mean fidelity of 0.969.

IV. CONCLUSION

We have demonstrated a simulation of the post-quench dynamics in a three-site antiferromagnetic Heisenberg chain beyond the qubit coherence time on real quantum hardware. This was achieved by compressing the Trotter time-evolved state at intermediate time steps into a variational form. The state overlap served as a cost function for the compression optimization step, which was executed on NISQ hardware without any additional overhead using a double time contour circuit. We have further benchmarked this variational Trotter compression algorithm on larger Heisenberg chains using statevector as well as noiseless and noisy circuit simulators. While most of our results were obtained for an integrable model, we also tested the ability of a variational ansatz to capture the dynamics generated by the Heisenberg Hamiltonian in the presence of an integrability-breaking term. These tests revealed that the performance of the VTC algorithm does not depend strongly on the presence or absence of integrability—indeed, an ansatz with the same number of parameters captures the wavefunction to the same accuracy in both cases.

The main goal moving forward is to increase the system sizes accessible to VTC simulations, for which several key issues need to be addressed. First, the noise level of the overlap cost function needs to be reduced in order to accelerate the classical optimiziation that is performed during the compression step. In addition to reducing device errors by hardware improvements, one can devise several error mitigation strategies [44] such as probabilistic error cancellation [28], virtual distillation [45, 46], Clifford data regression [47] or a combination thereof [48–52]. Sample noise can be reduced by increasing the number of circuit shots beyond the current limit of 2^{13} on the IBM backend we used. The fidelity of the Trotter time evolution during the propagation step can be increased by using pulse level control of the gates, which can often lead to shorter gate times. Finally, other future directions would be to explore alternative variational ansaetze such as hardware efficient ones, and to employ other classical optimizers for noisy cost functions.

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

[2] T. Prosen and M. Žnidarič, Is the efficiency of classical simulations of quantum dynamics related to integrability?, Phys. Rev. E **75**, 015202 (2007).

[3] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, Ann. Phys. January 2011 Special Issue, **326**, 96 (2011).

[4] S. Lloyd, Universal Quantum Simulators, Science **273**, 1073 (1996).

[5] A. M. Childs and N. Wiebe, Hamiltonian simulation using linear combinations of unitary operations, Quantum Information and Computation **12**, 901 (2012).

[6] D. W. Berry, G. Ahokas, R. Cleve, and B. C. Sanders, Efficient Quantum Algorithms for Simulating Sparse Hamiltonians, Commun. Math. Phys. **270**, 359 (2007).

[7] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma, Simulating Hamiltonian Dynamics with a Truncated Taylor Series, Phys. Rev. Lett. **114**, 090502 (2015).

[8] G. H. Low and I. L. Chuang, Optimal hamiltonian simulation by quantum signal processing, Phys. Rev. Lett. **118**, 010501 (2017).

[9] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Toward the first quantum simulation with quantum speedup, PNAS **115**, 9456 (2018).

[10] A. M. Childs and Y. Su, Nearly Optimal Lattice Simulation by Product Formulas, Phys. Rev. Lett. **123**, 050503 (2019).

[11] G. H. Low, V. Kliuchnikov, and N. Wiebe, Well-conditioned multiproduct Hamiltonian simulation, arXiv:1907.11679 (2019), arXiv:1907.11679 [physics, physics:quant-ph].

[12] J. Haah, M. B. Hastings, R. Kothari, and G. H. Low, Quantum Algorithm for Simulating Real Time Evolution of Lattice Hamiltonians, SIAM J. Comput. , FOCS18 (2021).

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

[14] M. C. Tran, S.-K. Chu, Y. Su, A. M. Childs, and A. V. Gorshkov, Destructive error interference in product-formula lattice simulation, Phys. Rev. Lett. **124**, 220502 (2020).

[15] A. Smith, M. S. Kim, F. Pollmann, and J. Knolle, Simulating quantum many-body dynamics on a current digital quantum computer, npj Quantum Information **5**, 1 (2019).

[16] A. Cervera-Lierta, Exact ising model simulation on a quantum computer, Quantum **2**, 114 (2018).

[17] 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 Information **6**, 82 (2020).

[18] J. Gibbs, K. Gili, Z. Holmes, B. Commeau, A. Arrasmith, L. Cincio, P. J. Coles, and A. Sornborger, Long-time simulations with high fidelity on quantum hardware (2021), arXiv:2102.04313 [quant-ph].

[19] S. Gulania, B. Peng, Y. Alexeev, and N. Govind, Quantum time dynamics of 1D-Heisenberg models employing the Yang-Baxter equation for circuit compression, arXiv:2112.01690.

[20] Y. Atia and D. Aharonov, Fast-forwarding of hamiltonians and exponentially precise measurements, Nature Communications **8**, 1572 (2017).

[21] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[22] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum (nisq) algorithms (2021), arXiv:2101.08448 [quant-ph].

[23] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Physical Review X **7**, 10.1103/physrevx.7.021050 (2017).

[24] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, Quantum **3**, 191 (2019).

[25] Y.-X. Yao, N. Gomes, F. Zhang, C.-Z. Wang, K.-M. Ho, T. Iadecola, and P. P. Orth, Adaptive variational quantum dynamics simulations, PRX Quantum **2**, 030307 (2021).

[26] S.-H. Lin, R. Dilip, A. G. Green, A. Smith, and F. Pollmann, Real- and imaginary-time evolution with compressed quantum circuits, PRX Quantum **2**, 010342 (2021).

[27] S. Barison, F. Vicentini, and G. Carleo, An efficient quantum algorithm for the time evolution of parameterized circuits, Quantum **5**, 512 (2021).

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

[29] N. Hansen, yoshihikoueno, ARF1, K. Nozawa, M. Chan, Y. Akimoto, and D. Brockhoff, Cma-es/pycma:r3.1.0 (2021).

[30] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Quantum fingerprinting, Physical Review Letters **87**, 10.1103/physrevlett.87.167902 (2001).

[31] B. F. Schiffer, J. Tura, and J. I. Cirac, Adiabatic spectroscopy and a variational quantum adiabatic algorithm (2021), arXiv:2103.01226 [quant-ph].

[32] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nat. Commun. **10**, 3007 (2019).

[33] H. L. Tang, V. O. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, qubit-ADAPT-VQE: An adaptive algorithm for con-

structing hardware-efficient ansatze on a quantum processor, arXiv:1911.10205.

[34] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[35] R. Wiersema, C. Zhou, Y. de Sereville, J. F. Carrasquilla, Y. B. Kim, and H. Yuen, Exploring entanglement and optimization within the hamiltonian variational ansatz, PRX Quantum **1**, 020319 (2020).

[36] A. M. Childs, Y. Su, M. C. Tran, N. Wiebe, and S. Zhu, Theory of trotter error with commutator scaling, Phys. Rev. X **11**, 011020 (2021).

[37] S. Foulds, V. Kendon, and T. Spiller, The controlled SWAP test for determining quantum entanglement, Quantum Sci. Technol. **6**, 035002 (2021).

[38] F. Vatan and C. Williams, Optimal quantum circuits for general two-qubit gates, Phys. Rev. A **69**, 032315 (2004).

[39] F. Gerber, florafauna/optimparallel-python: test zenodo (2020).

[40] A. Bolens and M. Heyl, Reinforcement learning for digital quantum simulation, Phys. Rev. Lett. **127**, 110502 (2021).

[41] G. Aleksandrowicz et al., Qiskit: An open-source framework for quantum computing, http://doi.org/10.5281/zenodo.2562111.

[42] R. LaRose, A. Mari, P. J. Karalekas, N. Shammah, and W. J. Zeng, Mitiq: A software package for error mitigation on noisy quantum computers (2020), arXiv:2009.04417 [quant-ph].

[43] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, Phys. Rev. X **7**, 021050 (2017).

[44] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid Quantum-Classical Algorithms and Quantum Error Mitigation, J. Phys. Soc. Jpn. **90**, 032001 (2021).

[45] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual Distillation for Quantum Error Mitigation, Phys. Rev. X **11**, 041036 (2021).

[46] B. Koczor, Exponential Error Suppression for Near-Term Quantum Devices, Phys. Rev. X **11**, 031057 (2021).

[47] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Error mitigation with Clifford quantum-circuit data, Quantum **5**, 592 (2021).

[48] Z. Cai, Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications, npj Quantum Inf **7**, 1 (2021).

[49] A. Mari, N. Shammah, and W. J. Zeng, Extending quantum probabilistic error cancellation by noise scaling, Phys. Rev. A **104**, 052607 (2021).

[50] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unified approach to data-driven quantum error mitigation, Phys. Rev. Research **3**, 033098 (2021).

[51] D. Bultrini, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, Unifying and benchmarking state-of-the-art quantum error mitigation techniques, arXiv:2107.13470 [quant-ph] (2021), arXiv:2107.13470 [quant-ph].

[52] C. Piveteau, D. Sutter, and S. Woerner, Quasiprobability decompositions with reduced sampling overhead, arXiv:2101.09290 [quant-ph] (2021), arXiv:2101.09290 [quant-ph].