

PAPER • OPEN ACCESS


A fault-tolerant variational quantum algorithm with limited T -depth

To cite this article: Hasan Sayginel *et al* 2024 *Quantum Sci. Technol.* **9** 015015

View the [article online](#) for updates and enhancements.

You may also like

- [Sequence of penalties method to study excited states using VQE](#)
R Carobene, S Barison and A Giachero
- [Many-body calculations for periodic materials via restricted Boltzmann machine-based VQE](#)
Shu Kanno and Tomofumi Tada
- [Accelerating the variational quantum eigensolver using parallelism](#)
Lana Mineh and Ashley Montanaro

The logo for kiutra, featuring a stylized circular icon to the left of the word "kiutra" in a lowercase, sans-serif font.A photograph of a Kiutra cryogenic system, showing a vertical column with a central tube and a base unit with two circular ports.

Easy-to-use and Helium-3 free
cryogenics solutions

LEARN MORE

Quantum Science and Technology



PAPER

A fault-tolerant variational quantum algorithm with limited T -depth

OPEN ACCESS

RECEIVED
29 March 2023

REVISED
5 October 2023

ACCEPTED FOR PUBLICATION
20 October 2023

PUBLISHED
13 November 2023

Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Hasan Sayginel^{1,2,*} , Francois Jamet² , Abhishek Agarwal² , Dan E Browne¹  and Ivan Rungger^{2,*} 

¹ Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom

² National Physical Laboratory, Teddington TW11 0LW, United Kingdom

* Authors to whom any correspondence should be addressed.

E-mail: hasan.sayginel.17@ucl.ac.uk and ivan.rungger@npl.co.uk

Keywords: fault-tolerance, variational quantum eigensolver, T -depth, optimization

Abstract

We propose a variational quantum eigensolver (VQE) algorithm that uses a fault-tolerant (FT) gate-set, and is hence suitable for implementation on a future error-corrected quantum computer. VQE quantum circuits are typically designed for near-term, noisy quantum devices and have continuously parameterized rotation gates as the central building block. On the other hand, an FT quantum computer (FTQC) can only implement a discrete set of logical gates, such as the so-called Clifford+ T gates. We show that the energy minimization of VQE can be performed with such an FT discrete gate-set, where we use the Ross–Selinger algorithm to transpile the continuous rotation gates to the error-correctable Clifford+ T gate-set. We find that there is no loss of convergence when compared to the one of parameterized circuits if an adaptive accuracy of the transpilation is used in the VQE optimization. State preparation with VQE requires only a moderate number of T -gates, depending on the system size and transpilation accuracy. We demonstrate these properties on emulators for two prototypical spin models with up to 16 qubits. This is a promising result for the integration of VQE and more generally variational algorithms in the emerging FT setting, where they can form building blocks of the general quantum algorithms that will become accessible in an FTQC.

1. Introduction

Quantum computing has seen significant progress in the last decade, leading to the development of quantum processors with an increasing number of qubits. These noisy intermediate-scale quantum (NISQ) processors are currently being investigated for a possible advantage over classical computers [1–4]. However, anticipated general quantum algorithms with a proven exponential speed-up over their classical counterparts lie outside of the reach of NISQ processors, because of the limited physical qubit numbers as well as the accumulation of noise for deep circuits [5–7]. Such algorithms include Shor’s algorithm [8] and quantum phase estimation (QPE) [9]. Therefore, for the full potential of quantum computing to be achieved, fault-tolerant quantum computers (FTQCs) will be needed.

An essential part of developing an FTQC is the incorporation of quantum error correction (QEC) into the processors, which will not only increase the coherence times of the resulting logical qubits but also actively correct for errors in faulty operations as the information is being processed. To this end, there have been recent prototype demonstrations of QEC on superconducting [10] and ion-trap devices [11]. Furthermore, recently error suppression with increasing code-size was demonstrated for the surface code [12] despite an increased number of physical qubits and gate operations. While the gate error levels in current hardware are not yet sufficiently low for large-scale error correction, with rapid progress in the field, error-corrected quantum computers with a small number of logical qubits may emerge in the next few years.

The main difference between the quantum circuits achievable on near-term, so called NISQ computers, and future error-corrected quantum computers are the range of quantum gates available at the logical level. Typically, NISQ devices allow continuously parameterized single-qubit rotations, such as an arbitrary z -rotation, $R_z(\theta)$, where $\theta \in [0, 2\pi]$. On the other hand, FTQCs only allow a fixed set of discrete rotation angles, such as the $\pi/4$ rotation about the z -axis, also known as the T -gate. For this reason, any algorithm

must first be compiled into a code-specific fault-tolerant (FT) gate-set before it can be run on the logical qubits. FT gate synthesis has been investigated in depth, and a number of algorithms have been proposed [13–17].

In this article we study the variational quantum eigensolver (VQE) algorithm [18] in the context of FT quantum computation. VQE uses the variational principle to compute an upper bound for the ground state energy of a Hamiltonian by using a parameterized quantum circuit that approximates the ground state of the Hamiltonian after optimization of the parameters in the quantum circuit. Computing an approximation for the ground state and energy of a Hamiltonian is generally the first step in probing the energetic properties of physical systems, a problem that commonly arises in condensed matter physics and quantum chemistry [19]. As a result, VQE has potential practical applications in fields such as material science [20, 21] and drug discovery [22, 23]. Furthermore, general cost functions can typically be represented as a Hamiltonian, where a variational algorithm can then be used to obtain the ground state, which hence yields the optimized result for the given cost function. Therefore, results obtained for VQE are generally applicable also in applications in optimization, in particular using the quantum approximate optimization algorithm (QAOA) [24], and in quantum machine learning [25, 26].

Due to its partial resilience to noise for moderate numbers of qubits and circuit depths, VQE has been studied as a near-term algorithm and used to perform a number of proof of concept demonstrations on noisy quantum hardware. However, the scaling to larger systems is hindered by the noise in the hardware, since the noise mitigation methods used successfully in small experiments become very expensive to scale up to larger circuits [19, 27]. As the system sizes increase, it is still an open question as to whether the VQE algorithm by itself will be an efficient method to compute the ground state of a Hamiltonian, since a number of challenges need to be solved, as outlined in [19].

For such large scale systems there are a number of alternative algorithms requiring fault-tolerance, which in principle allow to obtain desired states on a quantum computer with very high quality. These target states are typically ground states of a Hamiltonian, which can either represent a physical system or a general cost function in optimization and machine learning tasks. Such methods with known scalability include the QPE algorithm and the so-called Rodeo algorithm [28]. However, all these methods rely on the necessary condition that the initial state, upon which the method is applied, has already a sufficient overlap with the target state. The problem is that as the systems scale up in size, the overlap can become exponentially small [20], in which case the use of these methods becomes impossible. As a solution to potentially overcome this problem we propose to integrate these methods with an FT implementation of VQE, which we call FT-VQE. For large systems FT-VQE may not directly obtain the final target state with very high accuracy, but it will likely allow to obtain sufficient overlap with the final target state to be used as initial state in the algorithms above (for example QPE or Rodeo algorithms). We therefore expect FT-VQE to become an integral part of such applications in fault-tolerant devices for this initial state preparation step. We note that FT-VQE would not solve all the intrinsic scaling challenges of VQE, such as the difficulty in the classical optimization of the circuit parameters for large systems [19, 29, 30]. However, in the FT setting, we expect VQE to be used alongside algorithms with known scalability such as QPE [9].

The hurdle which must be overcome is that current VQE quantum circuits, designed for near-term hardware, include continuously parameterized rotation gates. The rotation angles are optimized on a classical computer, and the energies used in the optimization process for a given set of angles are obtained from the quantum computer. It had not yet been investigated how well this classical optimization process can be performed given the limitations of an FT discrete gate-set, where a continuous range of parameters can only be achieved in the limit of large numbers of T -gates, and whether the necessary approximations in the gates for circuits with finite T -gate depth would affect the convergence of the algorithm. In this article, we address this question and show that by integrating the Ross–Selinger (RS) recompilation of a continuous rotation to an FT gate-set into our algorithm, we can obtain systematic convergence of the VQE algorithm. We find that there is no slowdown in convergence efficiency when compared to the conventional VQE circuits with parameterized hardware gates if an adaptive RS recompilation accuracy is used.

The structure of the paper is as follows. Section 2 discusses the methods, introducing the FT Clifford+ T gate-set, the RS algorithm used for FT gate synthesis, and an FT implementation of the VQE algorithm. Section 3 introduces the two spin models that we use to test the method, and then the results for both fixed angles of a quantum circuit and for a full VQE loop performed using our implementation with an FT gate-set. In section 4 we discuss the conclusions.

2. Methods

2.1. FT gate-set

The universal FT gate-set used in this study is the Clifford+ T gate-set. The Clifford set in this study consists of Hadamard, phase-gate, and controlled-NOT gates, $\{H, S, \text{CNOT}\}$, which can generate any N -qubit Clifford operation. With the addition of the T -gate, any unitary lying in $SU(2^N)$ can be approximated up to any target accuracy by a sequence of Clifford+ T gates [31].

The T -gate is expressed in matrix form as

$$T = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}. \quad (1)$$

There are diverse ways of achieving FT logic gates depending on the underlying code. However, in many codes, including the surface code [32] (currently the most promising quantum error correcting code due to its high threshold and simple two-dimensional modular implementation), Clifford group gates can be performed in a direct and fast method, while T -gates are more expensive.

In order to implement the T -gates in these codes, a non-unitary technique such as magic-state distillation and injection is required. This procedure has a spatial and temporal overhead due to the additional ancillary qubits and physical gates needed [33]. Therefore, an important metric in the analysis of the performance of an FT quantum algorithm implemented with Clifford+ T gates is the number of T -gates.

In our work, we are, therefore, interested in the two following metrics.

- **T -count:** T -count is defined as the total number of T -gates in the FT quantum circuit.
- **T -depth:** T -depth is defined as the total number of layers of T -gates, where within each layer parallel execution of the T -gates on different logical qubits is possible.

2.2. RS algorithm

The RS algorithm computes approximations of arbitrary single qubit z -rotations, $R_z(\theta)$ over the Clifford+ T gates. It achieves this by approximating $R_z(\theta)$ with another unitary U , where U has an exact decomposition over Clifford+ T gates up to the single-qubit global phase, $\omega = e^{i\frac{\pi}{4}}$ [14]

$$R_z(\theta) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix} \xrightarrow{\text{RS}} U = \prod_m U_m \in \{\omega, H, S, T\}. \quad (2)$$

The algorithm then outputs the Clifford+ T decomposition of the unitary U . The error in this approximation is given as

$$\|R_z(\theta) - U\| \leq \epsilon = 10^{-d} \quad (3)$$

where $\|\cdot\|$ is the operator norm bounded by ϵ , and d is the digit accuracy. The accuracy of the decomposition can be systematically improved with increasing T -depth. Furthermore, the RS algorithm is efficient in the number of T -gates, which in the typical case scales with respect to ϵ as $4\log_2(\frac{1}{\epsilon}) + O(\log(\log(\frac{1}{\epsilon})))$. This is in contrast to the Solovay–Kitaev algorithm [13], which achieves T -counts of $O(\log^c(\frac{1}{\epsilon}))$, where c is a constant, making RS more favorable for $c > 3$.

Rotations about other axes, such as R_x and R_y , can be achieved with the corresponding R_z rotation and additional Clifford gates as

$$R_x(\theta) = HR_z(\theta)H, \quad R_y = SHR_z(\theta)HS^\dagger. \quad (4)$$

Therefore, the RS transpilation can be used to find the Clifford+ T approximation of any $SU(2)$ unitary.

The software implementation of the RS algorithm is available in an open-source software package [34]. Figure 1 shows practical numbers for the number of T gates in the RS decomposition of $R_z(\theta)$ rotations in the range $\theta \in [0, \pi]$ at various d , obtained using this software package. It can be observed that for a given d , the number of T gates required to realize different rotations is similar for all rotations. Note that multiples of $\pi/2$ can be carried out by S gates and hence have a T -count of 0.

2.3. FT implementation of VQE

The VQE is a hybrid quantum–classical algorithm, where the ground state wave function of a Hamiltonian \mathcal{H} is expressed on the quantum computer by executing a parametric quantum circuit, denoted as ansatz

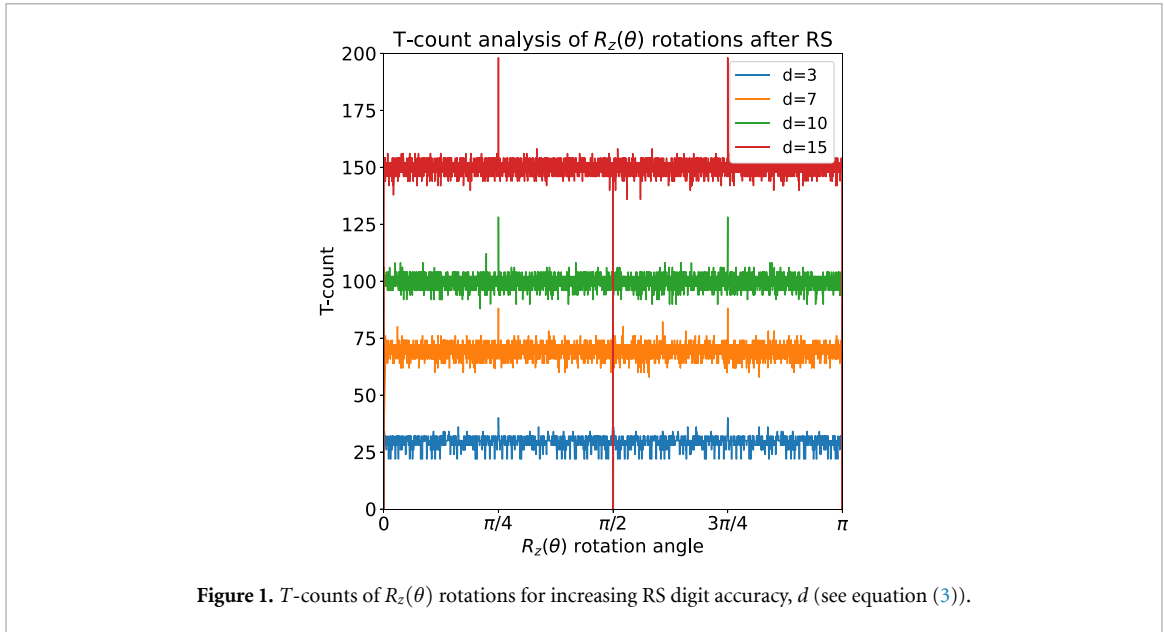


Figure 1. T-counts of $R_z(\theta)$ rotations for increasing RS digit accuracy, d (see equation (3)).

$V(\theta)|0\rangle$. $V(\theta)$ is a sequence of parametric gates, and θ is a vector of parameters that is optimized to find the ground state. The energy of the parametric wave function,

$$\mathcal{E}(\theta) = \langle 0|V(\theta)^\dagger \mathcal{H}V(\theta)|0\rangle, \quad (5)$$

can be obtained on a quantum processor as expectation value.

If the quantum circuit can span the entire Hilbert space, the minimum value of $\mathcal{E}(\theta)$ will correspond to the ground state of \mathcal{H} . However, practical implementations typically use a circuit with limited parameters, leading to an approximation of the ground state. The minimum value of $\mathcal{E}(\theta)$ is found via classical optimization.

2.3.1. Classical optimization

The task of finding the minimum value of $\mathcal{E}(\theta)$ is challenging, because it is usually a non-convex function. In this paper we focus on gradient-based minimization methods, specifically the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [35–38]. In BFGS, the procedure assumes a continuous optimization landscape as a function of the parameters. On the contrary, in FT quantum computation, one will necessarily be restricted to a discrete gate-set, typically Clifford+ T . Here, we therefore substitute the parametric gate-set $R_x(\theta), R_y(\theta), R_z(\theta)$ by their RS decompositions (equations (2) and (4)).

Two strategies are usually used to obtain the gradient in VQE: the finite-difference rule and the parameter-shift rule [39–41]. In this paper, we use both strategies and evaluate their relative performance.

Finite-difference rule: this method requires two evaluations of a given circuit with an infinitesimally shifted parameter. The required derivatives can be approximately calculated as

$$\partial_{\theta_\mu} \mathcal{E}(\theta_\mu) \approx \frac{\mathcal{E}(\theta_\mu + \frac{1}{2}\Delta\theta_\mu) - \mathcal{E}(\theta_\mu - \frac{1}{2}\Delta\theta_\mu)}{\Delta\theta_\mu}. \quad (6)$$

For ease of notation in the equation we only specify the shifted parameter θ_μ as argument in the function; all other parameters in the vector θ are kept constant. For the FT-VQE algorithm, the finite-shift rule requires two runs of the RS algorithm for calculating the Clifford+ T decompositions of the shifted angles, followed by two expectation value evaluations.

Parameter-shift rule: the parameter-shift rule is a method of computing the gradient of a parameterized gate by running the same quantum circuit twice with a finite shift in the gate parameter [39–41]. It states that if the generator G of the gate $\mathcal{G}(\theta_\mu)$ (with a single parameter θ_μ) has two unique eigenvalues denoted as e_0 and e_1 , then the derivative of the circuit expectation with respect to the gate parameter θ_μ is given by the difference in expectation value of two circuits. The two circuits are run with shifted parameters scaled by a parameter r given as $r = \frac{e_1 - e_0}{2}$. The gradient is obtained as

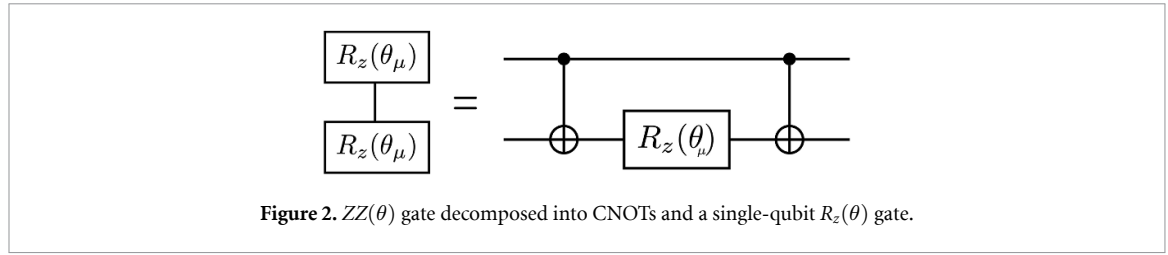


Figure 2. ZZ(θ) gate decomposed into CNOTs and a single-qubit $R_z(\theta)$ gate.

$$\partial_{\theta_\mu} \mathcal{E}(\theta_\mu) = r \left[\mathcal{E} \left(\theta_\mu + \frac{\pi}{4r} \right) - \mathcal{E} \left(\theta_\mu - \frac{\pi}{4r} \right) \right] \tag{7}$$

where r is the scaling parameter and $\frac{\pi}{4r}$ is a finite shift.

The main gate of interest in this work is $R_z(\theta) = e^{-i\frac{\theta}{2}Z}$. The Z -gate has eigenvalues ± 1 , hence the value of r for $R_z(\theta)$ is $r = \frac{1}{2}$. The parameter-shift rule for $R_z(\theta)$ is, therefore, expressed as

$$\partial_{\theta_\mu} \mathcal{E}(\theta_\mu) = \frac{1}{2} \left[\mathcal{E} \left(\theta_\mu + \frac{\pi}{2} \right) - \mathcal{E} \left(\theta_\mu - \frac{\pi}{2} \right) \right]. \tag{8}$$

We note that the required parameter-shift can be exactly implemented with S gates together with the appropriate multiple of the phase ω : $R_z(\pi/2) = \omega^7 S$ and $R_z(-\pi/2) = \omega SSS$.

We remark that the two-qubit gate $ZZ(\theta) = e^{i\frac{\theta}{2}Z \otimes Z}$, which is required in the Hamiltonian variational ansatz (HVA) used in the results section of this article, can be decomposed into CNOTs and a single-qubit rotation as shown in figure 2. The gradient of the $ZZ(\theta)$ can therefore be calculated with equation (8).

Therefore, the parameter-shift rule for $ZZ(\theta)$ -like gates does not have an additional FT overhead, i.e. no extra T -gates are needed for the required shifts because S gates can realize the desired shifts.

3. Results

To demonstrate the method, we apply the FT-VQE to two spin models, namely the transverse-field Ising model (TFIM) and the XXZ model, using an HVA [42, 43]. Within the HVA each ansatz is model-specific. For a general Hamiltonian, \mathcal{H} , which is a linear sum of not-necessarily commuting terms $\mathcal{H} = \sum_i \mathcal{H}_i$, an HVA is expressed as

$$|\psi_L\rangle = \prod_{l=1}^L \left(\prod_i \exp(-i\theta_{l,i} \mathcal{H}_i) \right) |\psi_0\rangle \tag{9}$$

where L is the total number of layers (layer depth). $|\psi_0\rangle$ is the ground state of one of the individual terms \mathcal{H}_i of the Hamiltonian [42, 43]. HVA is motivated by the QAOA [24], and the choice of $|\psi_0\rangle$ is similar, which is a state that is easy to prepare in hardware. For our investigation, HVA is chosen because of its linear scaling in the number of parameters with increasing number of layers.

TFIM. The TFIM Hamiltonian is a prototype model of quantum magnetism, and is given by

$$\mathcal{H}_{\text{TFIM}} = - \sum_{i=1}^N \left[\hat{Z}_i \hat{Z}_{i+1} + g \hat{X}_i \right], \tag{10}$$

where we consider periodic boundary conditions ($\hat{Z}_{N+1} = \hat{Z}_1$). The value of g affects the magnetic phase, where $g < 1$ gives a ferromagnetic phase as ground state, $g > 1$ corresponds to a paramagnetic phase, and $g = 1$ is a critical point. The system is gapless at $g = 1$ in the thermodynamic limit where N goes to infinity. For our FT-VQE simulations, we consider the critical $g = 1$ point for a system of $N = 16$ qubits. The initial state is taken as $|\psi_0\rangle = |+\rangle^{\otimes 16}$. For the HVA we use a layer depth of $L = 8$, leading to 16 parameters. The corresponding quantum circuit ansatz is shown in figure 3. This system size and ansatz choice are based on those of [43], where the entanglement and optimization properties of this system have been first investigated using the HVA. In [43] it is shown that barren plateaus are expected at $L = 8$ with a random parameter initialization. We, therefore, choose this setup as one of our test systems to evaluate the performance of FT-VQE in presence of barren plateaus.

XXZ model. The XXZ model is another prototypical model of quantum magnetism. In the one-dimensional case its Hamiltonian is given by

$$\mathcal{H}_{\text{XXZ}} = \sum_{i=1}^N \hat{X}_i \hat{X}_{i+1} + \hat{Y}_i \hat{Y}_{i+1} + \Delta \hat{Z}_i \hat{Z}_{i+1}, \tag{11}$$

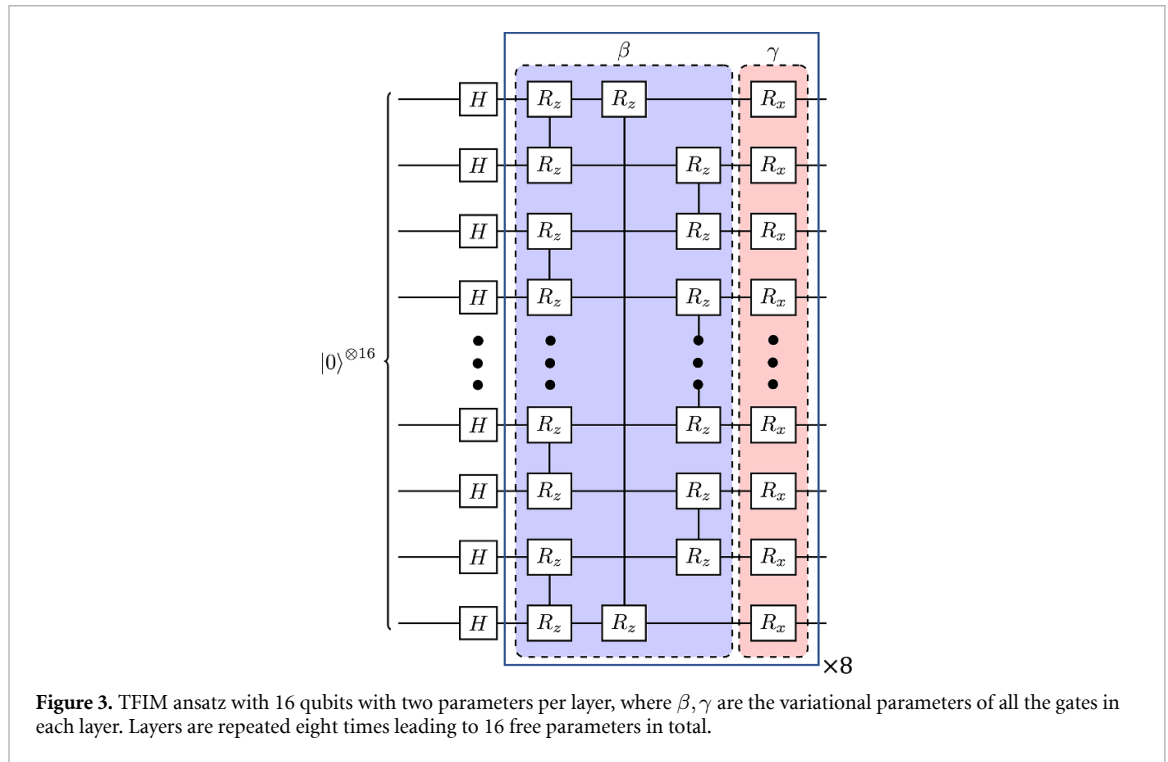


Figure 3. TFIM ansatz with 16 qubits with two parameters per layer, where β, γ are the variational parameters of all the gates in each layer. Layers are repeated eight times leading to 16 free parameters in total.

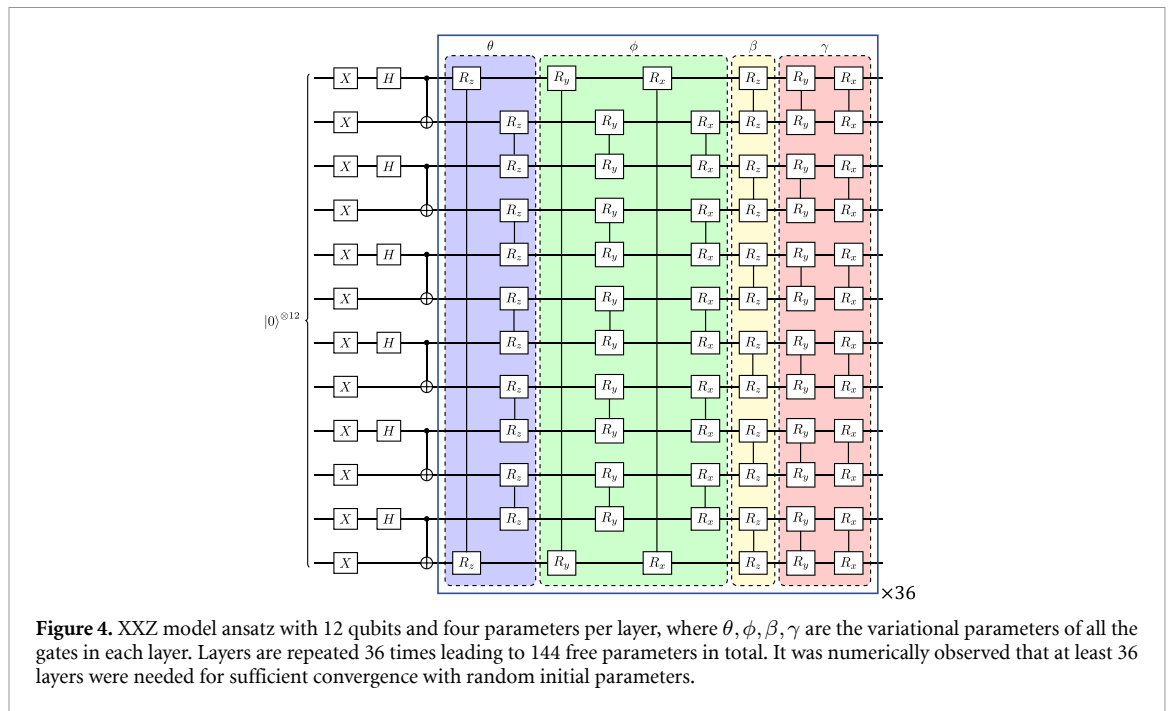
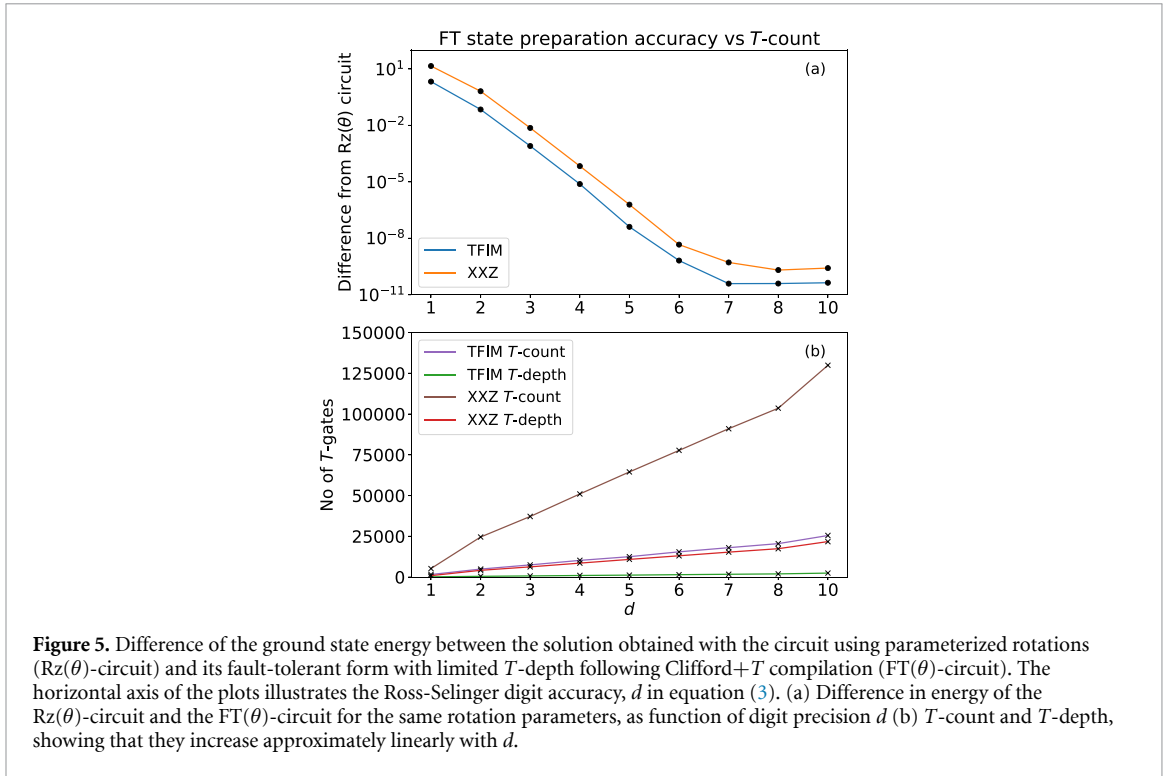


Figure 4. XXZ model ansatz with 12 qubits and four parameters per layer, where $\theta, \phi, \beta, \gamma$ are the variational parameters of all the gates in each layer. Layers are repeated 36 times leading to 144 free parameters in total. It was numerically observed that at least 36 layers were needed for sufficient convergence with random initial parameters.

where again periodic boundary conditions are applied. Δ represents the spin anisotropy of the model. For our FT-VQE simulations, we take $\Delta = 1$, at which there is a phase transition to the Néel ordered state. The system size we consider is $N = 12$ qubits, and for the HVA we use a layer depth $L = 36$, leading to 144 free parameters. The initial state is taken as $|\psi_0\rangle = \bigotimes_{n=1}^{N=6} |\Psi^-\rangle$, where $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$ is a Bell state as in [43]. In [43] it is shown that such ansatz achieves good VQE convergence at $L = 36$, which is attributed to the system being in the over-parameterization limit for the XXZ model, and hence avoids a barren plateau. Therefore, this second ansatz illustrated in figure 4 allows us to evaluate the performance of FT-VQE for a system that has no barren plateaus, but a large number of variational parameters and layers.

Finally, all the calculations performed in this paper have been done on the quantum emulator Qulacs [44] as state vector emulations.

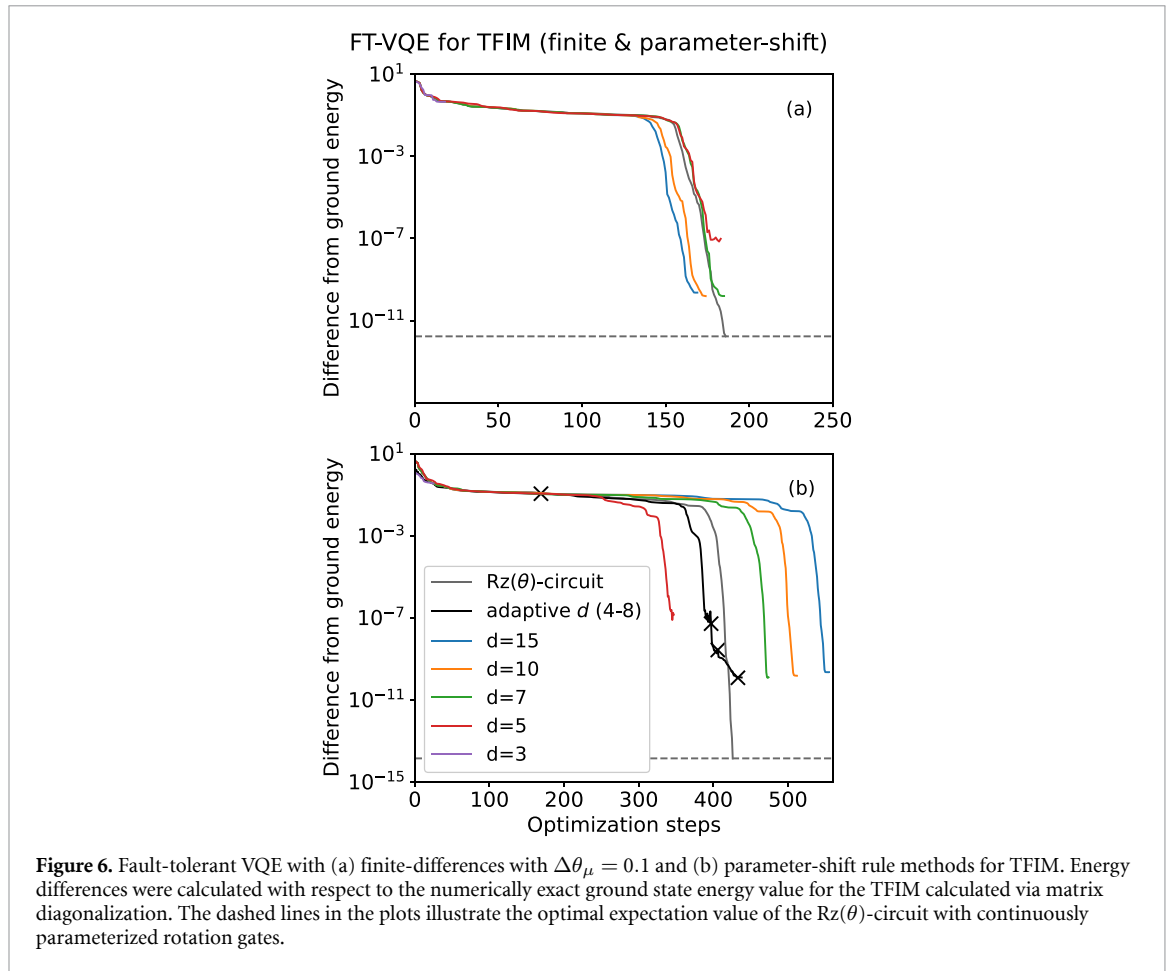


3.1. State preparation accuracy analysis with limited T -depth

First, we perform the VQE minimization using the parameterized rotation gates. We denote this circuit as the $R_z(\theta)$ -circuit due to the continuous parameterization of the rotation angles, and due to the decomposition of each parameterized gate into single-qubit $R_z(\theta)$ rotations. We then fix the rotation gates at the optimized values, perform the RS decomposition at different d for each rotation gate, and then replace all parameterized $R_z(\theta)$ rotations with their RS decompositions. We denote this circuit as the $FT(\theta)$ -circuit. We evaluate the energy difference between the $R_z(\theta)$ -circuit and $FT(\theta)$ -circuits for different digital accuracy d (equation (3)). Figure 5 shows this comparison for both the TFIM and XXZ models, together with the resulting T -count and T -depth. The energy difference decreases exponentially with d , until d is larger than about six, above which the energy difference becomes approximately constant. We verified that this tail-off is due to the finite numerical accuracy of the emulator software. In appendix we show that when performing the numerical evaluations at higher precision the energy difference decays exponentially up to digit accuracy of $d = 16$. Our results therefore show that one can systematically increase the accuracy of the energy of the $FT(\theta)$ -circuit by increasing d up to the numerical accuracy. Importantly, the circuit depth only increases linearly with d , as guaranteed by the RS decomposition.

For $d = 4$ the error in the energy expectation value is less than 10^{-4} for both the TFIM and XXZ models. For the TFIM model the corresponding T -count is 10 232, and the T -depth is 974; for the XXZ model the T -count is 50 976 and the T -depth is 8554. The energy difference becomes approximately constant for d larger than six, where it reaches values of the order of 10^{-10} to 10^{-9} . The T -count for $d = 6$ is 15 480, with a T -depth of 1474, for the TFIM; for the XXZ model the T -count is 77 680, with a T -depth of 13 056. TFIM with $N = 16$ qubits has a total number of 256 single-qubit $R_z(\theta)$ gates, whereas the XXZ model has a total of 1296 single-qubit $R_z(\theta)$ gates. The total T -count for both models at arbitrary d is approximately equal to the number of $R_z(\theta)$ gates multiplied by the average number of T -gates per RS decomposition of a single $R_z(\theta)$ gate, which can be extracted from figure 1. Note that in practice also the $R_z(\theta)$ gates have finite accuracy when implemented in hardware, so that the hardware result will also deviate from the ideal exact circuit results.

The above T -count estimates are modest when compared with recent resource count estimates for other FT algorithms. For example, for the problem of integer factoring, it is expected that at least 10^9 T -gates and 10^5 logical qubits will be needed [5]. In [45, 46] the required T -counts of various quantum simulation problems are investigated, and estimates range between 10^7 and 10^{12} T -gates depending on the complexity of the system. Particularly, in [46], authors estimate the T -counts of various spin systems. The Hamiltonian studied in [46] is similar to our XXZ model with $\Delta = 1$ (a Heisenberg chain), and an additional transverse-magnetic field. Their lower estimates are of the order of 10^7 T -gates for a spin-system with 14 qubits for an error level of 10^{-3} . For the XXZ model of a similar size, our estimate with FT-VQE is 10^4 – 10^5 T -gates.



3.2. VQE using the RS decompositions

We now evaluate and optimize the convergence of the full VQE minimization using the circuit with the RS decomposed rotation gates. As starting point for the minimization process we use random initial rotation parameters, and compile these into their RS decomposition for a given d . We use gradient-based optimizers. To obtain the gradients we use two different methods, and compare the respective convergence behavior. In the first method we evaluate the energy for the parameters shifted by a small finite difference from their previous values (see section 2.3 for details), and we compute the gradient from the resulting energy difference. We choose $\Delta\theta_\mu = 0.1$ for this small shift, because the shift needs to be orders of magnitude larger than the error made in the RS decomposition. In general, the precision of gradients that can be obtained using $FT(\theta)$ -circuits depends on d . In the second method we use the parameter-shift rule outlined in section 2.3 to obtain the gradients. Once the gradients are obtained, the optimizer updates the rotation angles as part of the energy minimization process. The updated rotation parameters are again recompiled into $FT(\theta)$ -circuits. This process is iterated until convergence of the VQE optimization. The optimizer used for both methods is the BFGS algorithm [35–38]. The stopping criterion for the minimization is when the difference energy between the new iteration and the previous one is less than 10^{-14} . The same criterion is taken for both the $R_z(\theta)$ -circuit and the $FT(\theta)$ -circuits.

We now analyze the convergence behavior of this FT-VQE method by evaluating the energy difference between the energy at each minimization step and the energy obtained by direct numerical diagonalization of the Hamiltonian matrix, which is exact up to numerical precision. This data is obtained from the TensorFlow Quantum dataset available at [47]. Figure 6 shows the results for the TFIM model, for both the finite-difference based gradient (figure 6(a)) and for the gradient obtained using the parameter-shift rule (figure 6(b)). We also show the convergence behavior for the $R_z(\theta)$ -circuit (gray solid line) as reference. Both in figures 6(a) and (b) we observe that the convergence behavior of the $R_z(\theta)$ -circuit and $FT(\theta)$ -circuits is similar. Up to about 150–300 optimization steps the energy decreases very slowly, and then rather abruptly the decrease of energy becomes much larger until it converges. This initial slow convergence is caused by small gradients in large parts of the energy landscape, which are commonly referred to as barren plateaus. It can be seen upon close inspection that the light purple $d = 3$ plot gets stuck in the barren plateau, where the energy error remains high at a value of the order of 10^{-1} . The gradient in the barren plateau is characterized

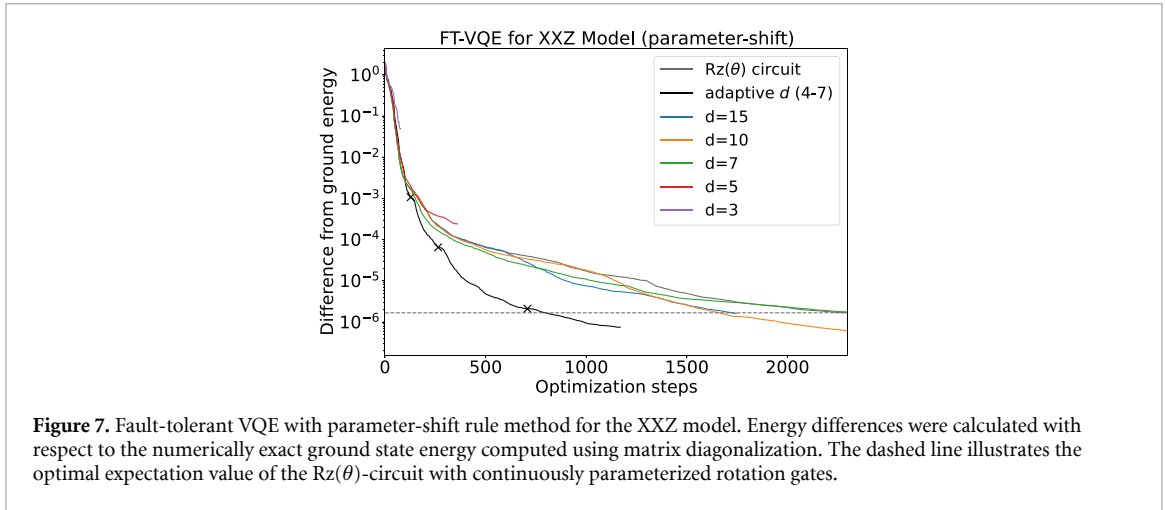


Figure 7. Fault-tolerant VQE with parameter-shift rule method for the XXZ model. Energy differences were calculated with respect to the numerically exact ground state energy computed using matrix diagonalization. The dashed line illustrates the optimal expectation value of the $Rz(\theta)$ -circuit with continuously parameterized rotation gates.

by its small magnitude, which leads to correspondingly small updates in the optimization process. In this case, a higher d is needed to update the angle with enough accuracy. For $d \geq 5$, sufficiently high accuracy has been reached allowing the convergence of the VQE. For $d \geq 5$ the convergence behavior in presence of barren plateaus is similar for both the $Rz(\theta)$ -circuit and $FT(\theta)$ -circuits.

We note that the finite-differences method escapes the barren plateau at around 150 steps, while for the parameter-shift rule this happens above 300 steps. We attribute this behavior to large shift of $\Delta\theta_\mu = 0.1$ in the angles for the gradient calculation. The rather large $\Delta\theta_\mu$ can lead to random errors of that order in the gradient. A stochastic component during optimization resulting from such random errors in finite precision gradients can speed up the convergence by bringing the system out of a barren plateau. These random errors can also help the system escape out of a local minimum and hence improve convergence. The $Rz(\theta)$ -circuit converges to an error of 10^{-12} with the finite-shift rule and to 10^{-14} with the parameter-shift rule. On the other hand, the $FT(\theta)$ -circuits converge to a similar error level of $10^{-10} - 10^{-9}$ with both finite-differences and parameter-shift rule for $d \geq 7$. $FT(\theta)$ -circuits use RS decompositions, and as a result involve many more multiplications of matrices as compared to the $Rz(\theta)$ -circuits. The higher number of multiplications amplifies the numerical precision errors in Qulacs matrix multiplications. This is consistent with the results of figure 5, and with the numerical precision analysis presented in appendix.

As explained in section 2.3, in the finite-shift rule one has to recompute the RS decompositions for the shifted angles, whereas in the parameter-shift rule the required angle shifts can be realized with S -gates. Therefore, in this setting the parameter-shift rule has less classical overhead for the computation of the gradient. Furthermore, since the required shifts are multiples of the S -gate, the gradient calculations do not contribute to the T -count. Finally, due to its resilience to noise, as outlined in [25], overall we expect that the parameter-shift rule is more efficient for FT-VQE.

Figure 7 shows the convergence behavior for the XXZ model, where we only use the parameter-shift method. One can see that there are no barren plateaus, differently to what was found for the TFIM system. One possible reason for this can be the larger number of variational parameters in the XXZ model ansatz. Such over-parameterization can help avoid barren plateaus, as discussed in [43]. The system converges rather rapidly to errors below 10^{-2} and 10^{-3} , which can be typical accuracy thresholds for practical applications. While for $d = 3$ the system only converges down to 10^{-1} , for $d = 5$ one reaches an accuracy of below 10^{-3} , and hence $d = 5$ can be the appropriate digit precision of the RS decomposition for practical applications. If one further increases d and allows for more optimization steps, the energy difference keeps decreasing down to values of 10^{-6} and below. Importantly, the $FT(\theta)$ -circuit convergence at these high d values is very similar to the one for the $Rz(\theta)$ -circuit. Note that even the $Rz(\theta)$ -circuit ansatz has an inherent accuracy limit compared to the ground energy calculated from matrix diagonalization, which is the reason for the small but finite remaining energy difference.

The results show that the small values of d can achieve convergence to moderate errors in the energy, and that with increasing d one can systematically improve the accuracy of the final energy up to a numerical accuracy limit determined by the precision of the software. Furthermore the results also show that when barren plateaus are encountered, random errors in the gradients can help the system escape from the barren plateau. For example, in figure 6(b) the $d = 5$ $FT(\theta)$ -circuit escapes the barren plateau at significantly lower optimization step number when compared to larger d . We also note that *a priori* it is not known which d will allow the system to reach a target accuracy. We therefore propose an adaptive algorithm, where a small d is

used to start the minimization process, which is then progressively increased during the convergence until a target accuracy is reached. We implement this adaptive approach in our software, where we minimize the energy for a given d until the criterion $|\mathcal{E}(\boldsymbol{\theta}_{t+1}) - \mathcal{E}(\boldsymbol{\theta}_t)| = 1 \times 10^{-14}$ is reached. Once this is the case, we proceed by increasing the accuracy d by one. We show the results for the TFIM model in figure 6(b), and for the XXZ model in figure 7 (black curves). We set the initial d to $d = 3$, and then allow d to increase up to a maximum of $d = 8$ within the adaptive algorithm. Each ‘×’ marker represents the point where d is incremented. The results show that the adaptive approach leads to efficient convergence, while allowing to minimize the T -depth of the circuit. We therefore expect a generally improved performance of adaptive FT-VQE when compared to FT-VQE at fixed d .

We note that while our work considered an HVA, the methods used are applicable to any variation of VQE. Since the Clifford+ $R_z(\theta)$ gate-set is universal, any quantum circuit can be decomposed first into this form, and then into the Clifford+ T set via RS decomposition. In practice, this technique will be efficient for a large number of parametric quantum circuits, where the performance of FT-VQE will mostly depend on the performance of the underlying variational algorithm with continuously parameterized circuits. For example, the ADAPT-VQE ansatz [48] can be used as an alternative to the fixed HVA.

3.3. Proposed early hardware demonstration

The current state-of-the-art quantum hardware with its significant noise levels is not yet suitable for executing extended circuits with many gates, which are required to perform the FT-VQE computations presented in our work. The earliest experimental demonstration of FT-VQE on partially FT quantum hardware, as a proof-of-principle rather than for practical purposes, can be done using one or two-qubit Hamiltonians. For the one-qubit case, a general single-qubit unitary $U \in \text{SU}(2)$ comprising three single qubit rotations can be used. Assuming an adaptive RS accuracy of $d = 3 - 7$, figure 1 suggests that between 25 and 75 T -gates for each of the single-qubit rotations should be sufficient. Therefore, this demonstration would require between 75 and 225 T -gates at each VQE iteration. For a demonstration with two qubits, one may consider a general two-qubit unitary $U \in \text{SU}(4)$ from [49]. This circuit requires three CNOTs and 15 single-qubit rotations. Assuming again an adaptive RS accuracy of $d = 3 - 7$, we estimate 375–1125 T -gates to be needed at each VQE iteration for successful convergence.

4. Conclusions

Our results demonstrate that variational quantum algorithms, such as VQE, show promise for practical applications on error-corrected quantum computers with limited T -depth. Our findings suggest that VQE is viable for running on an FTQC. We expect that FT-VQE alone cannot achieve quantum advantage, but that it will be a required component of a number of quantum algorithms that combine different methods. For example, VQE can be the first step of an algorithm to compute the ground state, where VQE is used to prepare a state with finite overlap with the ground state, which is then used as a starting state for other algorithms such as QPE.

We presented the FT-VQE algorithm, which we demonstrated on a quantum emulator for 12 and 16 qubits for two prototypical spin systems. Our study shows that FT-VQE convergence behavior is analogous to standard VQE, especially when we use our proposed adaptive setting of the RS circuit re-compilation accuracy. The presented data show that the discretization of arbitrarily parameterized rotation angles with a finite T -gate depth does not negatively affect the convergence of the VQE algorithm, and also that the required T -gate depth for such good convergence is moderate. The main limitation of FT-VQE when compared to traditional VQE is that the circuit depth is significantly increased due to the restriction to the Clifford+ T gate-set, the main advantage is that it allows for FT execution, where noise will not accumulate despite the increased circuit depth.

This work highlights the potential of FT-VQE as a powerful tool for practical quantum applications, with particular relevance in quantum chemistry simulations and optimization problems. These findings contribute to the ongoing development of quantum computing technologies and underscore the importance of continued research in this area.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

H S is supported by the Engineering and Physical Sciences Research Council [Grant Number EP/S021582/1]. H S also acknowledges support from the National Physical Laboratory. H S acknowledges the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), and associated support services, in the completion of this work. F J, A A, and I R acknowledge the support of the UK Department for Science, Innovation & Technology through the UK National Quantum Technologies Programme. DEB acknowledges the support of the Engineering and Physical Sciences Research Council [Grant Numbers EP/S005021/1 and EP/T001062/1] and InnovateUK.

After the completion of this work, we became aware of [50], which investigates state preparation with the unitary coupled cluster ansatz in circuits with limited T -depth. This work is complementary to the analysis of convergence behavior of the HVA decomposition investigated here.

Appendix. Numerical precision of the classical emulation

To evaluate the numerical precision of the fault-tolerant variational quantum eigensolver (FT-VQE) circuits we first evaluate the precision of the Ross–Selinger (RS) circuit synthesis for a single rotation gate, and then evaluate how the implementation of a sequence of gates in the used emulators affects the overall numerical precision.

A.1. RS error analysis

We carry out error analysis for the RS algorithm in order to assess whether the z -rotation realized by the approximate Clifford+ T unitary U is close to the expected rotation θ , and whether the error is compatible with the operator norm given in equation 3. Furthermore, the non-zero off-axis rotation realized by the unitary U is also calculated. For this purpose, we consider the Euler angle decomposition of a general $SU(2)$ matrix in the ZXZ convention, given by

$$U(\theta_1, \phi, \theta_2) = e^{-i\frac{\theta_1}{2}Z} e^{-i\frac{\phi}{2}X} e^{-i\frac{\theta_2}{2}Z} = \begin{bmatrix} e^{-i(\frac{\theta_1+\theta_2}{2})} \cos(\frac{\phi}{2}) & -ie^{i(\frac{\theta_1-\theta_2}{2})} \sin(\frac{\phi}{2}) \\ -ie^{-i(\frac{\theta_1-\theta_2}{2})} \sin(\frac{\phi}{2}) & e^{i(\frac{\theta_1+\theta_2}{2})} \cos(\frac{\phi}{2}) \end{bmatrix}. \quad (\text{A1})$$

We calculate the z -rotation angles θ_1 and θ_2 and the off-axis x -rotation ϕ .

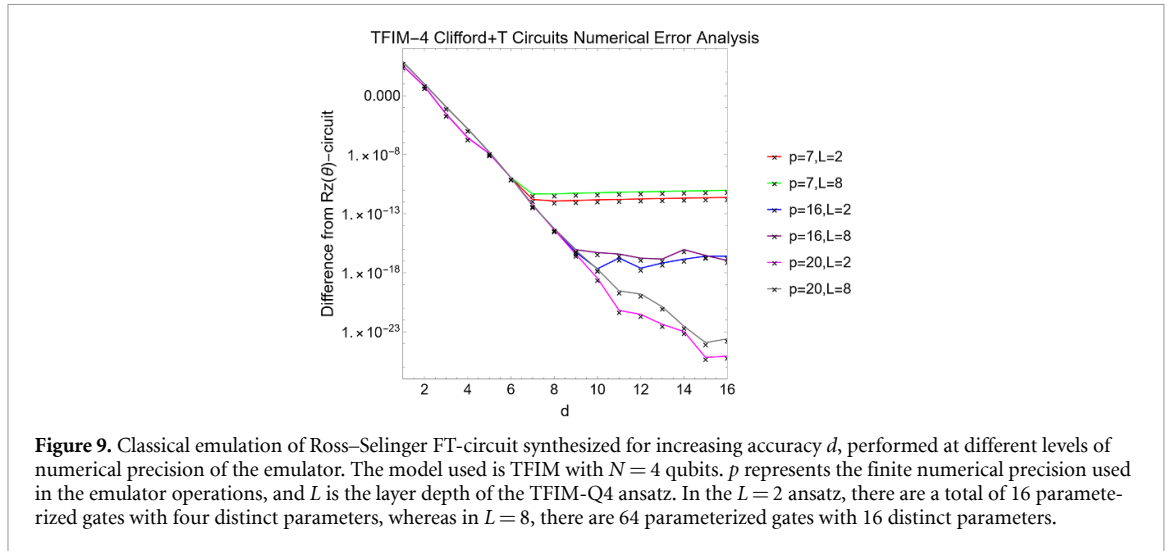
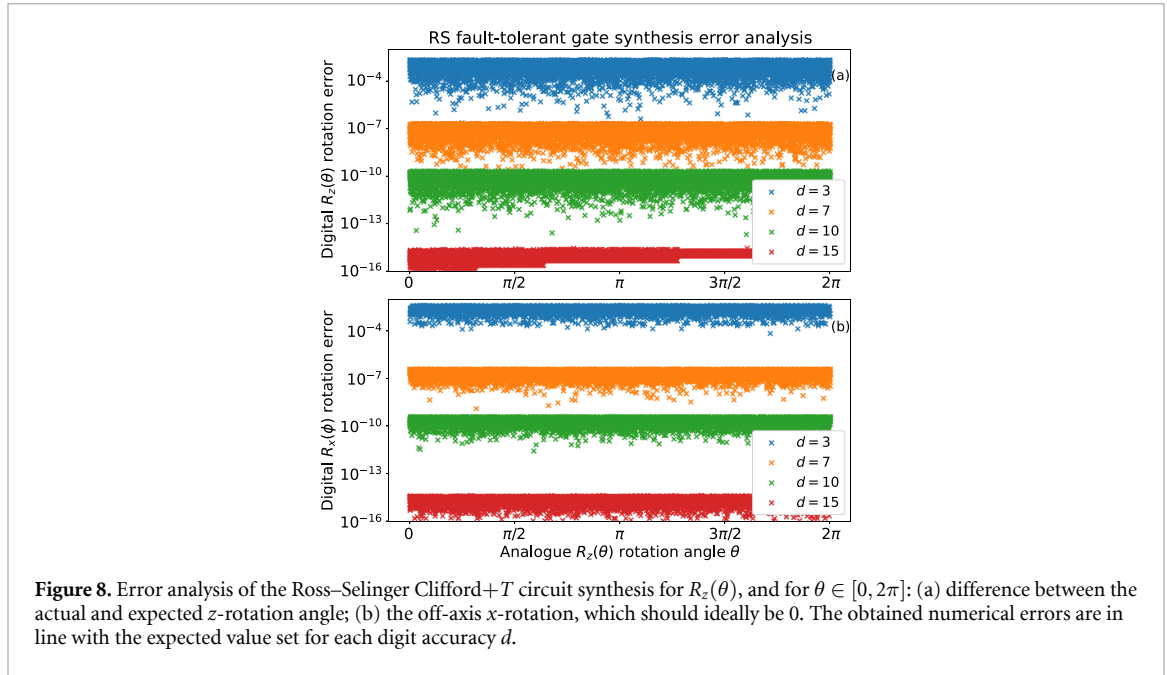
In the ideal case, θ in equation 3 should be $\theta \approx \theta_1 + \theta_2$, whereas the off-axis rotation should be $\phi \approx 0$. Figure 8 shows the results for the computed angles. Note that we take the expression of U for different angles directly from the RS software, so that the only error in consideration is from the RS approximation of the $R_z(\theta)$ rotation. We compute the absolute difference of the Euler angles of U from their expected value. In figure 8 it can be seen that both the z -rotations and the undesired x -rotations have similar levels of errors, which can be systematically reduced by increasing d .

A.2. Circuit emulation at finite numerical precision

RS decompositions lead to a large number of Clifford+ T gates. When using an emulator, these are implemented as classical matrix multiplications. Due to the large number of matrix multiplications, finite numerical precision errors accumulate and become observable. This behavior can be seen in figures 5(a) and 6(a), (b). To analyze accumulation of precision errors, and to explain why the error levels off when d is increased above some threshold in the FT-circuit emulator runs, we implement an emulator in Mathematica, which allows to perform the computations at arbitrary numerical precision.

Since the arbitrary precision calculations are computationally expensive for large number of qubits, we use a scaled-down version of the transverse-field Ising model (TFIM). We find that for $N = 4$ qubits we can observe the same leveling off of the error with increasing d as for the larger systems. We therefore use this scaled-down system as our test system. In analogy to figure 5, we first calculate the expectation value of the $R_z(\theta)$ -circuits with optimized parameters. Then, we perform the RS circuit recompilation for the $R_z(\theta)$ -circuits at various d , and compute the expectation values. The results are presented in figure 9. The axes of figures 9 and 5(a) are the same.

We take two different layer depths of the TFIM four-qubit ansatz in order to investigate the effect of increasing number of matrix multiplications at various numerical precision values, which we indicate by p . In the $L = 2$ ansatz, there are a total of 16 parameterized gates, with four distinct parameters, whereas for $L = 8$ there are 64 parameterized gates, with 16 distinct parameters. At the limited $p = 7$ precision, a similar pattern as that of figure 5(a) is observed, where the error plateaus despite increasing d . Importantly, when we



increase the precision of our Mathematica emulator to $p = 16$ and $p = 20$, the leveling off threshold of the error systematically decreases. For $p = 20$ no leveling off is observed. This confirms that the leveling off of the error in the emulator runs for the FT-circuits is due to the inherent numerical precision of the emulator.

ORCID iDs

Hasan Sayginel <https://orcid.org/0009-0006-8850-4044>

Francois Jamet <https://orcid.org/0000-0002-5679-209X>

Abhishek Agarwal <https://orcid.org/0000-0002-5881-7182>

Dan E Browne <https://orcid.org/0000-0003-3001-158X>

Ivan Rungger <https://orcid.org/0000-0002-9222-9317>

References

- [1] Arute F et al 2019 *Nature* **574** 505
- [2] Daley A J, Bloch I, Kokail C, Flannigan S, Pearson N, Troyer M and Zoller P 2022 *Nature* **607** 667
- [3] Madsen L S et al 2022 *Nature* **606** 75
- [4] Wu Y et al 2021 *Phys. Rev. Lett.* **127** 180501
- [5] Gidney C and Ekerå M 2021 *Quantum* **5** 433
- [6] Kivlichan I D and et al 2020 *Quantum* **4** 296
- [7] Campbell E T 2021 *Quantum Sci. Technol.* **7** 015007

- [8] Shor P W 1997 *SIAM J. Comput.* **26** 1484–509
- [9] Kitaev A Y 1995 *Electron. Coll. Comput. Complex.* **TR96** 003
- [10] Krinner S et al 2022 *Nature* **605** 669
- [11] Postler L et al 2022 *Nature* **605** 675
- [12] Acharya R et al 2023 *Nature* **614** 676
- [13] Glaudell A N, Ross N J and Taylor J M 2021 *npj Quantum Information* **7** 103
- [14] Ross N J and Selinger P 2016 *Quantum Inf. Comput.* **16** 901
- [15] Kliuchnikov V, Maslov D and Mosca M 2016 *IEEE Trans. Comput.* **65** 161
- [16] Gheorghiu V, Mosca M and Mukhopadhyay P 2022 *npj Quantum Inf.* **8** 141
- [17] Amy M, Glaudell A N and Ross N J 2020 *Quantum* **4** 252
- [18] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 *Nat. Commun.* **5** 4213
- [19] Tilly J et al 2022 *Phys. Rep.* **986** 1
- [20] Bauer B, Bravyi S, Motta M and Chan G K-L 2020 *Chem. Rev.* **120** 12685
- [21] Lordi V and Nichol J M 2021 *MRS Bull.* **46** 589
- [22] Cao Y, Romero J and Aspuru-Guzik A 2018 *IBM J. Res. Dev.* **62** 1
- [23] Blunt N S et al 2022 *J. Chem. Theory Comput.* **18** 7001
- [24] Farhi E, Goldstone J, Gutmann S and Zhou L 2022 *Quantum* **6** 759
- [25] Cerezo M, Verdon G, Huang H-Y, Cincio L and Coles P J 2022 *Nat. Comput. Sci.* **2** 567
- [26] Benedetti M, Lloyd E, Sack S and Fiorentini M 2019 *Quantum Sci. Technol.* **4** 043001
- [27] Guo Y and Yang S 2022 *PRX Quantum* **3** 040313
- [28] Choi K, Lee D, Bonitati J, Qian Z and Watkins J 2021 *Phys. Rev. Lett.* **127** 040505
- [29] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 *Nat. Commun.* **9** 4812
- [30] Wang S, Fontana E, Cerezo M, Sharma K, Sone A, Cincio L and Coles P J 2021 *Nat. Commun.* **12** 6961
- [31] Nebe G, Rains E M and Sloane N J 2006 *Self-Dual Codes and Invariant Theory* (Springer)
- [32] Bravyi S, Englbrecht M, König R and Peard N 2018 *npj Quantum Inform.* **4** 55
- [33] Litinski D 2019 *Quantum* **3** 128
- [34] Selinger P and Ross N J 2019 Exact and approximate synthesis of quantum circuits (available at: <https://hackage.haskell.org/package/newsynth>) (Accessed 25 January 2023)
- [35] BROYDEN C G 1970 *IMA J. Appl. Math.* **6** 76
- [36] Fletcher R 1970 *Comput. J.* **13** 317
- [37] Goldfarb D 1970 *Math. Comput.* **24** 23
- [38] Shanno D F 1970 *Math. Comput.* **24** 647
- [39] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev. A* **98** 032309
- [40] Wierichs D, Izaac J, Wang C and Lin C Y-Y 2022 *Quantum* **6** 677
- [41] Crooks G E 2019 arXiv:1905.13311 [quant-ph]
- [42] Ho W W and Hsieh T H 2019 *SciPost Phys.* **6** 029
- [43] Wiersema R, Zhou C, de Sereville Y, Carrasquilla J F, Kim Y B and Yuen H 2020 *PRX Quantum* **1** 020319
- [44] Suzuki Y et al 2021 *Quantum* **5** 559
- [45] Babbush R et al 2018 *Phys. Rev. X* (<https://doi.org/10.1103/PhysRevX.8.041015>)
- [46] Childs A M, Maslov D, Nam Y, Ross N J and Su Y 2018 *Proc. Natl Acad. Sci.* **115** 9456
- [47] Tensorflow Tensor Flow Quantum datasets 2022 (available at: www.tensorflow.org/quantum/api_docs/python/tfq/datasets) (Accessed 25 January 2023)
- [48] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 *Nat. Commun.* **10** 3007
- [49] Vatan F and Williams C 2004 *Phys. Rev. A* **69** 032315
- [50] Kuroiwa K and Nakagawa Y O 2023 arXiv:2301.04150 [quant-ph]