Global optimization in variational quantum algorithms via dynamic tunneling method

Seung Park[1], Kyunghyun Baek[2], Seungjin Lee[*2], Mahn-Soo Choi[†1]

[1]Department of Physics, Korea University, Seoul 02841, Republic of Korea
[2]Electronics and Telecommunications Research Institute, Daejeon 34129, Republic of Korea

E-mail: [*]`seungjin.lee@etri.re.kr`, [†]`choims@korea.ac.kr`

Abstract.
    We present a global optimization routine for the variational quantum algorithms, which utilizes the dynamic tunneling flow. Originally designed to leverage information gathered by a gradient-based optimizer around local minima, we adapt the conventional dynamic tunneling flow to exploit the distance measure of quantum states, resolving issues of extrinsic degeneracy arising from the parametrization of quantum states. Our global optimization algorithm is applied to the variational quantum eigensolver for the transverse-field Ising model to demonstrate the performance of our routine while comparing it with the conventional dynamic tunneling method, which is based on the Euclidean distance measure on the parameter space.

Keywords: quantum computation, variational quantum algorithms, dynamic tunneling method, local minima, quantum machine learning

1. Introduction

The variational quantum algorithm (VQA) is a hybrid quantum-classical algorithm that has attracted significant attention in recent years owing to its potential to address complex problems in quantum chemistry [1, 2], material science [3, 4], physics [5], computational science [6, 7], and many other areas. By combining classical optimization techniques with quantum computation, the VQA leverages the variational principle of quantum mechanics to find the minimum of the cost function [8, 9].

In mathematical terms, the VQA can be framed as an optimization problem, where the cost function is defined as the expectation value of an observable with respect to a variational quantum state, referred to as an ansatz state, implemented through a parametric quantum circuit. Classical optimization techniques are then applied to iteratively update the parameters of the parametric quantum circuit until the expectation value of the observable converges to the lowest value.

As an optimization problem, the choice of classical optimizer plays a critical role in efficiently solving the VQA, along with the design of a suitable ansatz state. While well-established classical optimizers are directly utilized, several optimizers have been developed specifically to enhance VQA's performance, focusing on gradient-based optimization routines due to their proven convergence properties and the ability of quantum circuits to implement the gradient of VQA in parallel. Among the optimizers explored are the stochastic gradient descent [10, 11], the Riemannian gradient flow [12], and the natural gradient method [13, 14].

However, gradient-based optimizers for VQA often encounter the challenge of local minima, which is inherent both in gradient-based optimization and the variational method of quantum mechanics. Various designs for ansatz have been proposed to mitigate such local convergence, notably including [15, 16, 17, 18]. Among other things, it has been suggested to over-parameterize quantum state [19], introducing more quantum layers to transform local minima into saddle points, akin to strategies used in classical neural networks [20, 21]. However, over-parameterization presents practical challenges, including the limitations imposed by noise in NISQ devices and their restricted coherence time. Moreover, increasing circuit depth can exacerbate local minima issues, expanding the search space [22, 23] and potentially leading to barren plateaus [14, 24].

Consequently, global optimization routines have been proposed to complement the design of parametric quantum states, with most methods being gradient-free to avoid the drawbacks of gradient-based optimization. Techniques like the Nelder-Mead method [25, 26], bound optimization by quadratic approximation [27, 28], and the quantum kernel surrogate model-based method [29] have been explored, though they may require a significant number of function evaluations to converge, raising scalability concerns in the VQA context.

In this work, we propose a global optimization strategy specialized for VQA that incorporates a gradient-based optimizer called the dynamic tunneling method [30].

Similar to the conventional dynamic tunneling method, the optimizer uses the local minima detected by the optimizer to generate a dynamic flow towards a global minimum. However, our modified dynamic tunneling method enhances efficiency by addressing the limitations of the conventional method within the VQA framework.

This paper is organized as follows: In the next section, we first introduce the conventional dynamic tunneling method and then propose its modification for VQA, incorporating a distance measure between quantum states. In Section 3, we demonstrate the global optimization of the transverse field Ising model using the modified dynamic tunneling method, comparing its performance with the conventional dynamic tunneling method applied to the same problem. Section 4 is devoted to the conclusion and outlook of our work.

2. Dynamic tunneling method on VQA

2.1. Dynamic tunneling method

Dynamic tunneling method and its variants [30, 31] are global optimization algorithms implementing a dynamic flow to escape from a valley around a stable point to a valley around another stable point having the lower value of the cost function of an optimization problem. In its primitive form, this is achieved by constructing a so-called energy function from the cost function by exploiting the information of local minima of the cost function.

Explicitly, given an cost function $f(\boldsymbol{x})$, the dynamic tunneling flow is constructed by promoting optimization parameters $\boldsymbol{x}$ to be a flow $\boldsymbol{x}(t)$ generated by

$$\dot{\boldsymbol{x}} = -\frac{\partial E}{\partial \boldsymbol{x}} \tag{1}$$

where $E = E(\boldsymbol{x}; \bar{\boldsymbol{x}}, \lambda, k)$ is an enegy function given by [30]

$$E(\boldsymbol{x}; \bar{\boldsymbol{x}}, \lambda, k) = \frac{f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}})}{|\boldsymbol{x} - \bar{\boldsymbol{x}}|^{2\lambda}} + k \int_0^{f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}})} F_{\text{ReLU}}(z) dz. \tag{2}$$

In (2), λ and k are hyperparameters depending on the problem, and $\bar{\boldsymbol{x}}$ is a stable point (or a local minimum) of the cost function. Also, $F_{\text{ReLU}}(z)$ is the rectified linear unit which is zero only when the $z \leq 0$. Together with an appropriately chosen λ, the first term in (2) amounts to induce a pole of the energy function at $\bar{\boldsymbol{x}}$ while violating the Lipschitz condition at $\bar{\boldsymbol{x}}$, which enables the flow to converge to a point outside of the valley around $\bar{\boldsymbol{x}}$ [30, 31]. On the other hand, the second term in (2) is a penalty term with weight k imposing the tunneling to seek a point having a lower cost value than $f(\bar{\boldsymbol{x}})$. Instead of constructing an energy function for a tunneling flow, one can introduce a dynamic flow directly via [30]

$$\dot{\boldsymbol{x}} = -\left(\frac{1}{|\boldsymbol{x} - \bar{\boldsymbol{x}}|^{2\lambda}} + k\, F_{\text{ReLU}}(f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}))\right) \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \tag{3}$$
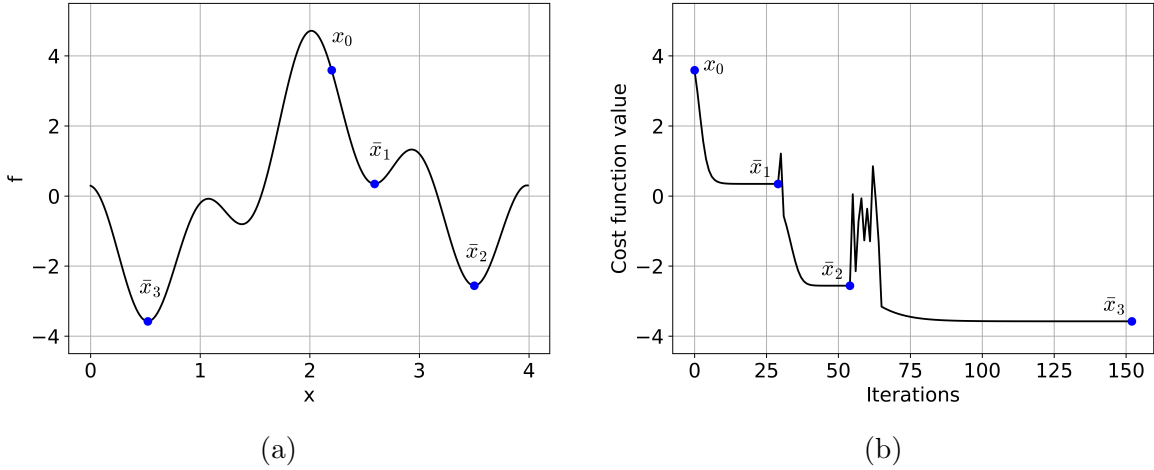
Figure 1: An illustration of the dynamic tunneling flow. (a) The landscape of a cost function. (b) The change of the cost function as iteration proceeds. The blue dots in (b) represent the corresponding points on the landscape in (a). The sharp peaks in (b) indicate the occurrence of dynamic tunneling.

which is more convenient for a practical implementation. Clearly, the dynamic flow in (3) has the same pole structure and penalty constraint as the energy function in (2).

Having a dynamic tunneling flow, the conventional dynamic tunneling method corresponds to a gradient-based optimization algorithm followed by an implementation of the dynamic tunneling flow in (3) for the stable point recognized by the optimization process. In turn, the dynamic tunneling method is a successive application of the optimization-tunneling pair until the tunneling flow can find no lower valley.

To illustrate the dynamic tunneling method and point out our method more clearly below, let us consider an optimization problem defined by the cost function

$$f(x) = \cos\left(\frac{\pi}{2}\left(x - \frac{1}{2}\right)\right) - \frac{1}{2}\cos\left(2\pi\left(x + \frac{3}{2}\right)\right) + \sin\left(\pi\left(x + \frac{1}{2}\right)\right) - \frac{3}{2}\sin\left(\frac{\pi}{2}(x+1)\right),$$
(4)

which is periodic in $x \in [0,4)$. The landscape of (4) is plotted in Figure 1(a). We first employ the usual gradient-descent optimizer with step size 0.005 and the initial parameter $x_0 = 2.2$ indicated as the point x_0 in Figure 1(a). The optimizer then ends up with a stable point $\bar{x}_1$ in Figure 1(a). Given a local minimum $\bar{x}_1$, which now becomes the initial point for the next optimization process, a dynamic tunneling flow (3) sets in and the optimizer escapes from $\bar{x}_1$, leading to the next stable point $\bar{x}_2$ depicted in Figure 1(a). By repeatedly alternating the relaxation to a local stable point by gradient-descent and the subsequent escape from it by dynamic tunneling, the whole dynamic tunneling routine traverses from the initial point x_0 through local stable points $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$, among which $\bar{x}_3$ is the global minimum of the problem, as illustrated in Figure 1(b). For the hyperparameters, in this particular example, we have chosen

$k = 50$ and set λ to be 1.5 times the lowest power of gradient of convergence to each stable point. The lowest power is extracted from the last two iterations reaching each local minimum.

2.2. Dynamic tunneling method with distance measure for quantum states

The VQA is a quantum algorithm that implements the variational principle of quantum mechanics by combining a parametric quantum circuit and a classical optimization algorithm [1, 32, 33]. In its simplest form, a VQA can be cast into the form of

$$f(\boldsymbol{x}; O) = \text{Tr}\left(O\rho(\boldsymbol{x})\right) \tag{5}$$

where O and ρ are a target observable and the density operator of a quantum state depending on a parameter $\boldsymbol{x}$, respectively. In practice, particularly for an N-qubit system, the quantum state in (5) is prepared by applying the parametrized quantum circuit implementing a unitary operator $U(\boldsymbol{x})$ to a fixed initial state ρ_0 so that the parametric quantum state has the form of

$$\rho(\boldsymbol{x}) = U(\boldsymbol{x})\rho_0 U(\boldsymbol{x})^\dagger, \tag{6}$$

often called an ansatz state.

The quantum function $f(\boldsymbol{x}; O)$ corresponds to a parametrized quantum expectation on the observable O, so an implementation of the variational method can be obtained by optimizing $f(\boldsymbol{x}; O)$ for $\boldsymbol{x}$ primarily by employing a classical optimizer. Consequently, the VQA is an optimization problem whose cost function corresponds to the quantum function of a parametric state,

By identifying a VQA as an optimization problem defined by a quantum function, we can apply the dynamic tunneling method in Section 2.1 by introducing a tunneling flow as in (1). However, apart from the intrinsic degeneracy of the target observable, the quantum function is not extrinsically injective since multiple parameters can induce the same quantum state due to the periodic feature of the unitary gates. In turn, the dynamic flow starting from a local minimum, say $\bar{\boldsymbol{x}}$, may converge to another point $\bar{\boldsymbol{x}}'$ with $\rho(\bar{\boldsymbol{x}}') = \rho(\bar{\boldsymbol{x}})$ which makes the algorithm computationally redundant.

Such extrinsic degeneracy thus not only makes the algorithm computationally inefficient but also prevents the termination criterion of the algorithm from working properly since we cannot exploit the divergence of the tunneling flow to ensure the global minimum. Therefore, instead of using the parameter-space distance $|\boldsymbol{x} - \bar{\boldsymbol{x}}|$, we employ a distance measure on the space of quantum states in the first term of (1). Explicitly, the tunneling flow for VQA is given by

$$\dot{\boldsymbol{x}} = -\left(\frac{1}{\mathcal{D}(\rho(\boldsymbol{x}), \rho(\bar{\boldsymbol{x}}))^{2\lambda}} + k\, F_{\text{ReLU}}(f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}))\right)\frac{\partial f}{\partial \boldsymbol{x}} \tag{7}$$

where $\mathcal{D}(\rho, \rho')$ is a distance measure on states ρ and ρ'.

By introducing the distance measure on quantum states, one can avoid all other local minima corresponding to the same quantum state since the modified dynamic tunneling flow completely excludes the quantum state itself due to the first term inside the parenthesis in (7). It is emphasized that the exclusion globally occurs over the whole parameter space, enabling the flow to circumvent a whole set of local minima corresponding to each quantum state.

There are several distance measures for quantum states [34], and most of them have a similar effect when they are cast into the dynamic tunneling method. Here, we utilize the Hilbert-Schmidt (HS) distance as the distance measure for the dynamic tunneling method. The Hilbert-Schmidt distance $\mathcal{D}_{\mathrm{HS}}$ between two density matrices ρ, σ is given by

$$\mathcal{D}_{\mathrm{HS}}(\rho, \sigma) = \sqrt{\mathrm{Tr}\left((\rho - \sigma)^{\dagger}(\rho - \sigma)\right)}. \tag{8}$$

If the two density matrices are of pure states, denoted by $\rho \equiv |\psi\rangle\langle\psi|$ and $\sigma \equiv |\phi\rangle\langle\phi|$, the Hilbert-Schmidt distance between them can be expressed as

$$\mathcal{D}_{\mathrm{HS}}(|\psi\rangle\langle\psi|, |\phi\rangle\langle\phi|) = \sqrt{2 - 2|\langle\psi|\phi\rangle|^2}. \tag{9}$$

Consequently, in most cases of the VQA, the distance can be readily obtained by measuring the fidelity between the two states [35].

In addition to resolving the extrinsic degeneracy of the quantum function, the Hilbert-Schmidt distance is practically appreciable since it is bounded by $\sqrt{2}$, regardless of the size of the parameter space. In particular, due to the upper bound, the denominator of the first term in (7) does not significantly affect the gradient of the cost function, thereby ensuring the convergence of the tunneling process.

Moreover, local minima with different values of quantum function correspond to nearly orthogonal quantum states, and hence one can expect that the distance at the end of the tunneling process is close to the upper bound, $\sqrt{2}$. This facilitates the estimation of appropriate values for the hyperparameters of the optimization, such as the learning rate, in contrast to the tunneling algorithm using the parameter-space distance in which the distance between any two local minima is unpredictable.

3. Case study: transverse-field Ising model

We test the performance of the method with numerical simulation on the variational quantum eigensolver (VQE), one of the main applications of the VQA that seeks the ground-state energy of the given Hamiltonian. In this test, we simulate the VQE on the one-dimensional (1D) transverse-field Ising model.

The transverse-field Ising model, consisting of spins arranged in a lattice, is a prototype model exhibiting various correlation effects in many-body physics and has been extensively studied in condensed matter physics [36, 37, 38]. Furthermore, the one-dimensional version allows for an exact solution [36]. This makes the model an excellent benchmark for many quantum algorithms, including the VQE.
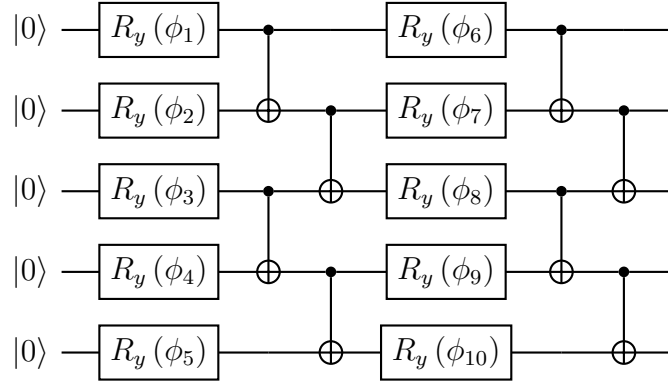
Figure 2: A parameterized quantum circuit corresponding to a variational ansatz for 5 qubits. $R_y(\phi_i) := e^{-i\sigma^y \phi_i/2}$ denotes the single-qubit rotation around the y-axis.

The transverse-field Ising model is described by the Hamiltonian

$$H = J \sum_{j=1}^{N-1} \sigma_j^z \sigma_{j+1}^z + B \sum_{j=1}^{N} \sigma_j^x, \tag{10}$$

where N is the number of spins (i.e. qubits), $J > 0$ is the coupling strength between the nearest-neighbor spins and B is the strength of the external magnetic field in the transverse direction. We set $J = 1$, therefore all the energy will be measured in units of J. In the tests below, we naturally choose the Hamiltonian as a target observable.

Note that the model exhibits a quantum phase transition from the ordered (i.e., anti-ferromagnetic) to disordered (i.e., paramagnetic) states. This phase transition may be understood in terms of spontaneous breaking of Z_2 symmetry: To see this, note that the Hamiltonian is invariant under the symmetry transformation $X \equiv \prod_j \sigma_j^x$, that is, $[H, X] = 0$. In the $B \to 0$ limit (ordered phase), the ground states of the model are two-fold degenerate and are related to each other by the symmetry transformation X. On the other hand, in the $B \to \infty$ limit (disordered phase), the ground state is non-degenerate; the symmetry transformation X maps the ground state to itself. In this work, we focus on the disordered (paramagnetic) phase to avoid unnecessary complications due to the ground-state degeneracy.

Here, we consider the case of $N = 10$ and $B = 5$, in which the system is in the nondegenerate, paramagnetic phase. The ground-state energy is about -50.45. Note that the Hamiltonian is real-valued in the computational basis. This implies that the wave function of any eigenstate of the Hamiltonian can be chosen to be real in the same basis. We exploit this fact to reduce the variational ansatz subspace. Figure 2 shows the parameterized quantum circuit to generate our variational ansatz for $N = 5$; for larger systems, the quantum circuit can be extended in a similar pattern. Each R_y involves the trainable parameter ϕ_i, while the structure of the CNOT gates mirrors the nearest-neighbor connectivity of the system, thereby creating entanglement.

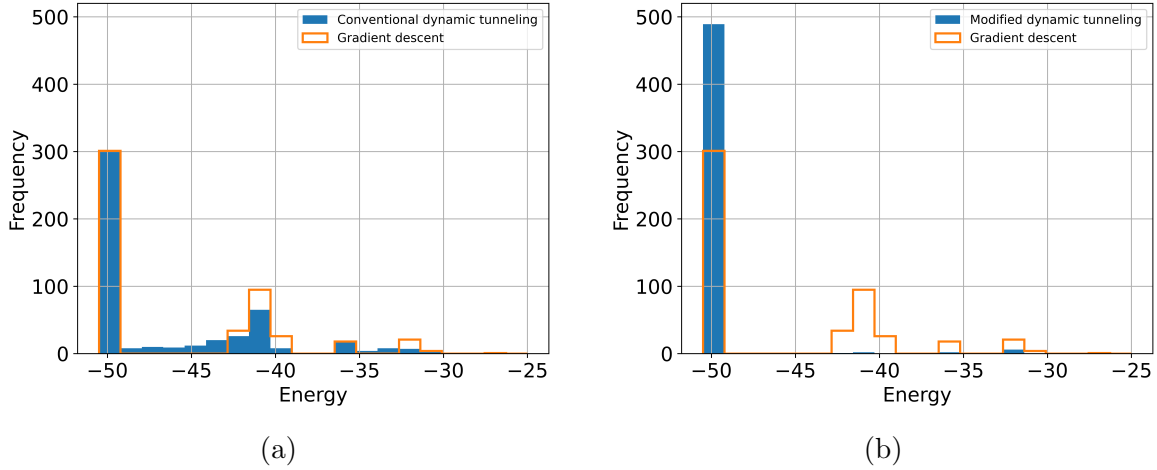(a)   (b)

Figure 3: A comparison of the overall convergence of the (a) conventional and (b) modified dynamic tunneling methods. Each histogram shows the final converged values of the cost function over 500 random samples. For a reference, we also show the histogram with empty bars and orange boundaries from the simple gradient-descent method. The global minimum is approximately -50.45 in this example.

Figure 3 summarizes our main results, comparing the performances of the simple gradient-descent optimization algorithm (empty boxes with orange boundaries, in both (a) and (b)), (a) the conventional dynamic tunneling algorithm based on the parameter-space distance and (b) the modified dynamic tunneling algorithm based on the distance measure on quantum states.

We have set the hyperparameters for our simulation as follows: λ for each tunneling is chosen to be 1.5 times the lowest power of gradient of convergence to each stable point. $k = 1.125$, and the learning rate is 0.01, fixed over the whole process. For practical reasons, we limit the number of execution of the dynamic tunneling algorithm to 6 times, with the iteration limit of 2000 for each run.

In Figure 3(b), we can see that the modified dynamic tunneling method successfully escapes the local minima and reaches the global minimum, while the conventional one fails to converge into the global minimum as shown in Figure 3(a). A few results that remained in local minima in Figure 3(b) can also reach the global minimum by increasing the iteration limit.

Figure 4 presents typical profiles of the cost function over the entire optimization process. It is interesting to note that in Figure 4(a), each optimization procedure stops even on a slope rather than at local minima on the landscape of the cost function in the parameter space, indicating that the tunneling process does not converge within the given iteration limit. This is the reason why the histograms in Figure 3(a) are smoothly distributed over the global and local minima. On the other hand, the iterations for the modified dynamic tunneling converge well first to local minima and eventually to the global minimum, as shown in Figure 4(b).

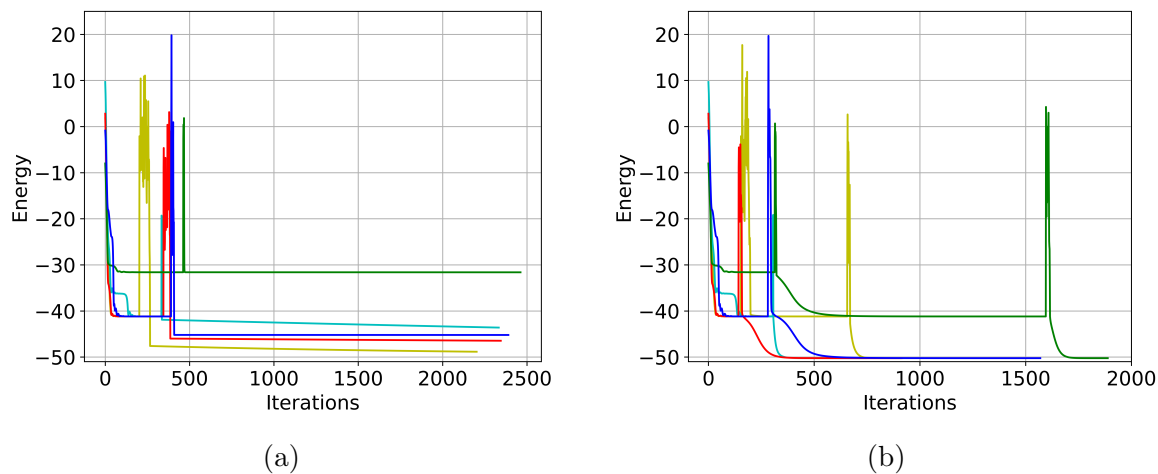(a)   (b)

Figure 4: Iteration profiles of the cost function (energy) over the entire optimization process in the (a) conventional and (b) modified dynamic tunneling methods for 5 representative samples out of the whole in Figure 3. Each sharp peak indicates the occurrence of dynamic tunneling.
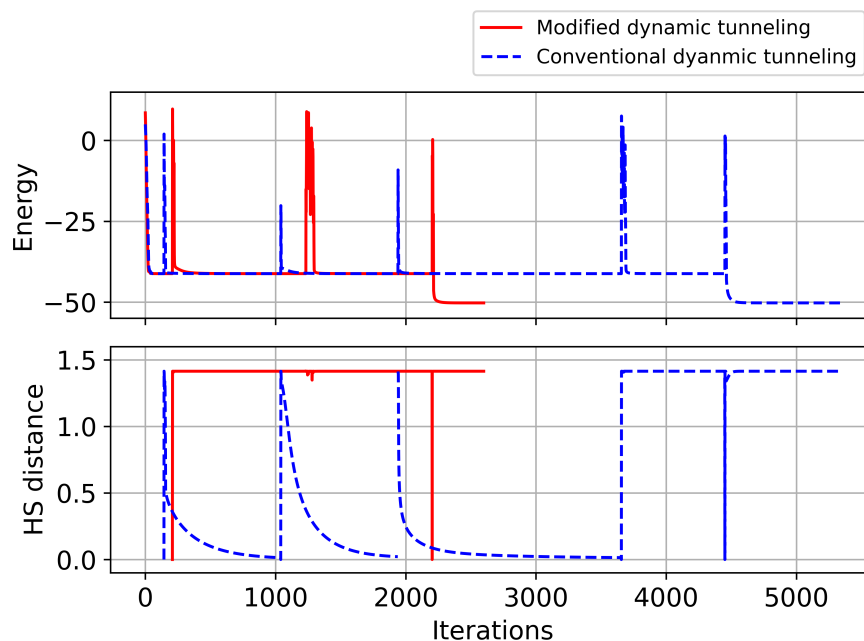

Figure 5: A detailed comparison of the iteration profiles of the conventional (blue dashed curves) and modified (red curves) dynamic tunneling methods for selected worst-case samples. For the conventional method, we have put an ad-hoc bound 2^λ on the λ term to avoid the convergence problem. Even with such a recipe, the conventional method converges much slower (almost twice as slow) than the modified method due to the extrinsic degeneracy.
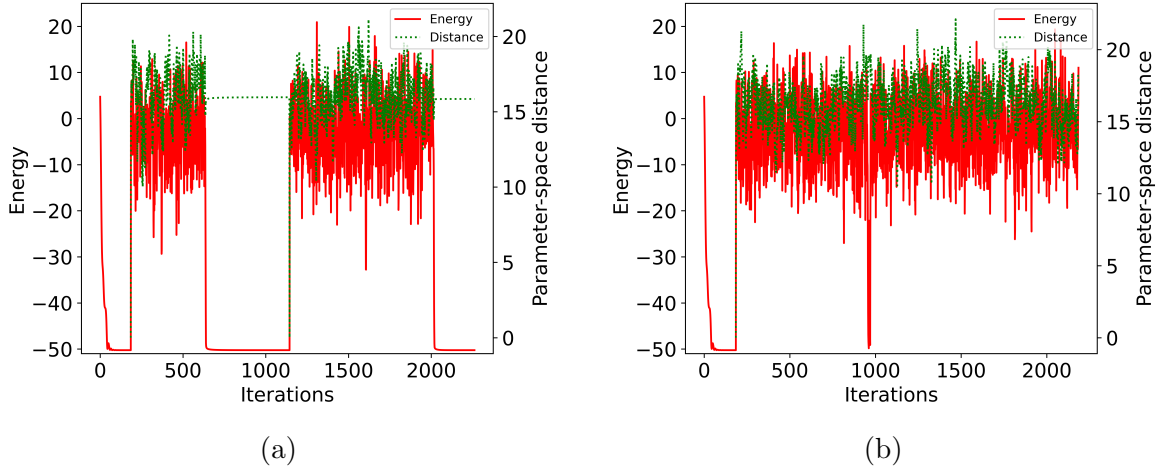
(a)   (b)

Figure 6: The iteration profiles around the *global* minimum of the (a) conventional and (b) modified dynamic tunneling methods. The green dashed lines show the value of the parameter-space distance at the corresponding iteration steps, and we have put an ad-hoc bound 2^λ on the λ term for the conventional method. The conventional dynamic tunneling flow never converges to a single global minimum, and the iteration must be terminated manually.

Further details of the dynamical aspects of the conventional and modified dynamical tunneling algorithms are compared in Figure 5. As one can see from the blue curves in Figure 5, the conventional dynamic tunneling method often encounters local minima of the same quantum state. In our simulation, this happened 57 times per 431 samples which start the tunneling process from local minima. Figure 5 shows the worst-case scenario observed during the modified tunneling optimization over the 431 samples. One can see that even when the flow converges into the local minimum of the same energy level again, the two quantum states corresponding to the local minima are different. Such instances are rare and occur only when there is intrinsic degeneracy of the local minimum.

Figure 6 highlights the worst-case examples that are not featured in Figure 4 or 5. Recall that there are multiple global minima corresponding to identical ground states but located at different points in the parameter space. As shown in Figure 6(a), the flow in the conventional tunneling method may jump between these alternative global minima, never converging to a single minimum. In our simulation, this happened 23 times out of 300 samples starting from the global minimum. In these instances, the termination criterion fails to determine the global minimum, even with an increase in the iteration limit. On the other hand, when using the modified tunneling method, the flow cannot converge into other global minima, as shown in Figure 6(b). This ensures the termination criterion works well, allowing for the safe determination of the global minimum.

4. Conclusion and outlook

In this work, we have proposed a global optimization algorithm for VQA by employing the dynamic tunneling flow generated by (7). In contrast to the conventional dynamic flow, our dynamic flow by (7) exploits the distance measure of quantum states to resolve the extrinsic degeneracy on the quantum function arising from the parametrization of the ansatz state.

The modified dynamic tunneling method has been applied to the VQE for the transverse-field Ising model. Our simulation results demonstrate the enhanced performance of the modified dynamic tunneling method as a global optimization algorithm, in terms of the convergence to the global solution, the analytic behavior of the method during the process, and the number of iterations to reach the global solution.

As indicated in (7), the dynamic tunneling flow for VQA heavily relies on a distance measure on quantum states. Therefore, a further investigation of the distance measure concerning a practical implementation of the flow is required in addition to the direct adoption of known efficient measures computed by the quantum computation framework [35, 39, 40].

Finally, the success of the algorithm is also sensitive to an appropriate choice of the value of the hyperparameters. While the hyperparameter λ in (7) can be determined from the history of gradients of the quantum function, the tunneling penalty k of the flow remains to be suitably chosen to ensure that the flow converges to the lower valley. If the target system is scalable, one can devise a strategy to estimate the appropriate value of k based on the dimension of the space of variational parameters. During the simulation, we extrapolate the value of k by gradually increasing the system size N, starting from a sufficiently small size, $N = 6$. In general, the tunneling penalty term arises from the constraint imposing the feasible space of the flow to be below the known local minima. Thus, one can regard the flow as constrained dynamics, where optimization routines for constrained problems, such as the interior-point method [41], can be applied to improve the convergence of the flow instead of delicately choosing k.
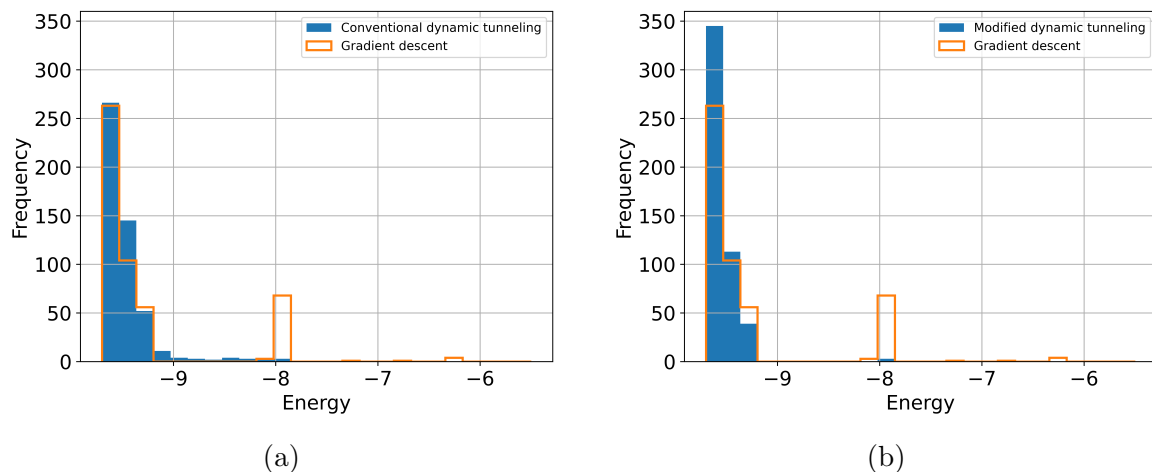
Figure A1: A comparison of the overall convergence of the (a) conventional and (b) modified dynamic tunneling method. Each histogram shows the final converged values of the cost function over 500 random samples. The histogram with empty bars and orange boundaries from the simple gradient-descent method is included as a reference. The global minimum is approximately -9.76 in this example.

Appendix A. Anti-ferromagnetic phase

In the main text, we mainly focused on the paramagnetic phase of the transverse-field Ising model to simplify the discussion and make the main points clearer. In this appendix, for reference, we provide the simulation results in the anti-ferromagnetic phase. Exhibiting an *intrinsic* ground-state degeneracy and long-range correlations, the anti-ferromagnetic phase sets a further testing ground compared to the paramagnetic phase.

We set $N = 10$ and $B = 0.5$ in the model (10), and use the same ansatz as in the paramagnetic phase, as described in Figure 2. Due to the finite size, the ground states are nearly degenerate with energies around -9.76, as explained in Section 3.

Figure A1 shows the overall convergence of different methods. As in the paramagnetic phase (Figure 3), the modified dynamic tunneling method better escapes local minima and reaches the subspace of the nearly degenerated ground states faster. Note that the few distinct bins near the global minimum are different linear combinations of the nearly degenerate ground states, which are difficult to distinguish within the ansatz subspace.

Technically, during the optimization process, we used FISTA[42] (a variant of gradient descent optimizer) to accelerate the update of the parameters for both the conventional and modified dynamic tunneling methods. This is necessary because in the anti-ferromagnetic phase, the gradient is significantly smaller than in the paramagnetic case across the entire cost function landscape. On the other hand, we keep the same hyperparameters except for the value of k, which is set to 16 in this case.

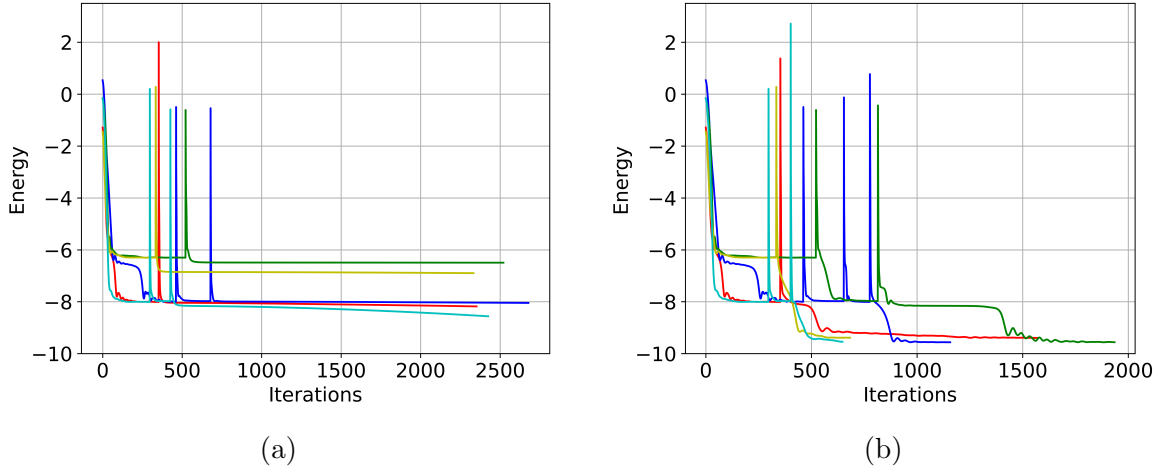Figure A2 presents the profiles of the cost function of some representative samples

Figure A2: Iteration profiles of the cost function (energy) over the entire optimization process in the (a) conventional and (b) modified dynamic tunneling methods for 5 representative samples out of the whole in Figure A1.

throughout the entire optimization process. As in the paramagnetic phase, the optimization process of the conventional dynamic tunneling method in Figure A2(a) stops on a slope, failing to converge into any minima. In contrast, as shown in Figure A2(b), the modified dynamic tunneling method successfully finds the global minimum within fewer iteration steps, despite the intrinsic degeneracy causing the iterations to encounter local minima with the same energy level multiple times.

Appendix B. Higher-dimensional models

In the main text, we have discussed 1D transverse-field Ising model to compare the performances of the conventional and modified dynamic tunneling methods. In this appendix, for completeness, we provide some simulation results for the transverse-field Ising model on a two-dimensional square lattice. From the results in the main text and below, we expect the same conclusion in three dimensions.

We consider a 3×4 square lattice ($N = 12$) and set $B = 5$. The ansatz is slightly modified to the form depicted in Figure B1 [on a 2×4 lattice ($N = 8$) for illustration] in order to account for the connectivity between qubits on the lattice.

The hyperparameters for this simulation are set to be the same as those for the 1D cases, except for the value of k, which is set to 0.35 now. We again use the accelerated gradient-descent method to enhance the overall convergence rate for both the conventional and modified dynamic tunneling methods.

Figure B2 compares the performance of our method with the conventional dynamic tunneling method. The modified dynamic tunneling method better escapes local minima and converges faster to the global minimum than the conventional one. Through this example, we expect that our method can be applied to higher-dimensional problems, which have an even more complicated landscape of the cost function.
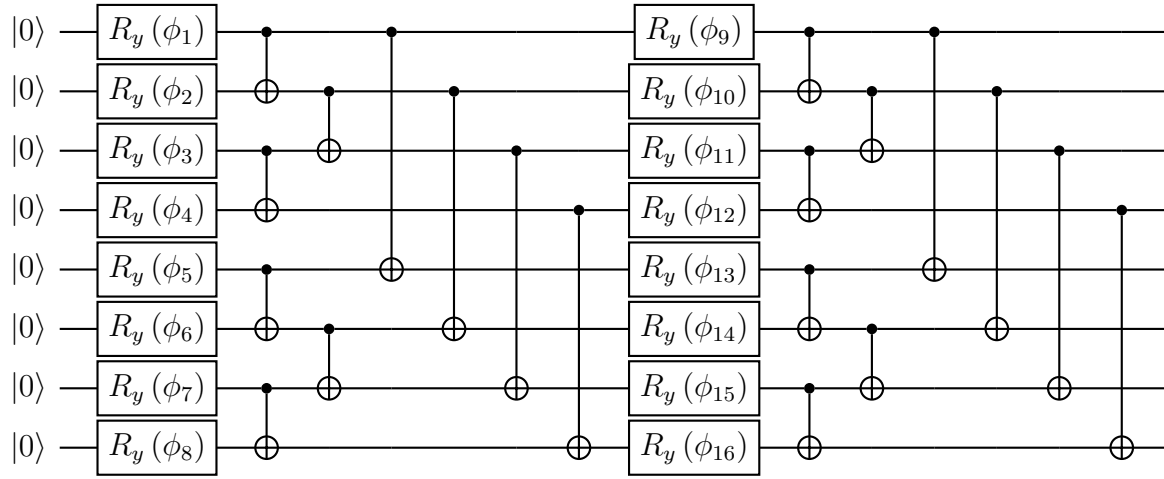
Figure B1: A parameterized quantum circuit corresponding to a variational ansatz for the transverse-field Ising model on a 2×4 ($N = 8$) lattice. The ansatz may be constructed in a similar fashion for larger lattices. $R_y(\phi_i) := e^{-i\sigma^y \phi_i/2}$ denotes the single-qubit rotation around the y-axis.
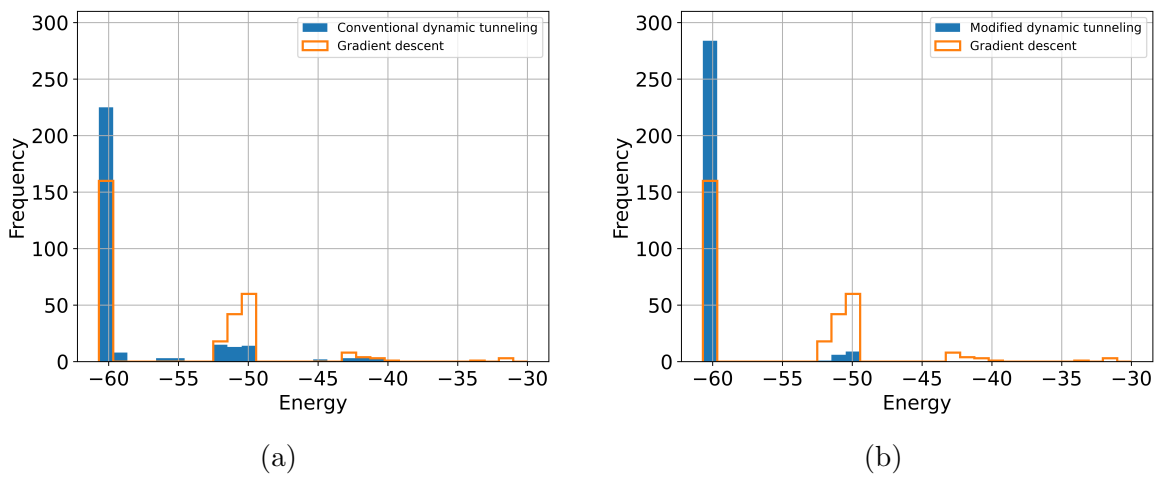

Figure B2: Overall convergence of the (a) conventional and (b) modified dynamic tunneling methods. Each histogram shows the final converged values of the cost function over 300 random samples. For a reference, we also show the histogram with empty bars and orange boundaries from the simple gradient-descent method. The global minimum is approximately -60.87 in this example.

References

[1] Peruzzo A, McClean J, Shadbolt P, Yung M H, Zhou X Q, Love P J, Aspuru-Guzik A and O'brien J L 2014 *Nature communications* **5** 4213 URL https://doi.org/10.1038/ncomms5213

[2] McArdle S, Endo S, Aspuru-Guzik A, Benjamin S C and Yuan X 2020 *Rev. Mod. Phys.* **92**(1) 015003 URL https://link.aps.org/doi/10.1103/RevModPhys.92.015003

[3] Ma H, Govoni M and Galli G 2020 *npj Computational Materials* **6** 85 URL https://doi.org/10.1038/s41524-020-00353-z

[4] Endo S, Kurata I and Nakagawa Y O 2020 *Phys. Rev. Res.* **2**(3) 033281 URL https://link.aps.org/doi/10.1103/PhysRevResearch.2.033281

[5] Bauer C W, Davoudi Z, Balantekin A B, Bhattacharya T, Carena M, de Jong W A, Draper P, El-Khadra A, Gemelke N, Hanada M, Kharzeev D, Lamm H, Li Y Y, Liu J, Lukin M, Meurice Y, Monroe C, Nachman B, Pagano G, Preskill J, Rinaldi E, Roggero A, Santiago D I, Savage M J, Siddiqi I, Siopsis G, Van Zanten D, Wiebe N, Yamauchi Y, Yeter-Aydeniz K and Zorzetti S 2023 *PRX Quantum* **4**(2) 027001 URL https://link.aps.org/doi/10.1103/PRXQuantum.4.027001

[6] Jäger J and Krems R V 2023 *Nature Communications* **14** 576 URL https://doi.org/10.1038/s41467-023-36144-5

[7] Amaro D, Rosenkranz M, Fitzpatrick N, Hirano K and Fiorentini M 2022 *EPJ Quantum Technology* **9** 5 URL https://doi.org/10.1140/epjqt/s40507-022-00123-4

[8] Bharti K, Cervera-Lierta A, Kyaw T H, Haug T, Alperin-Lea S, Anand A, Degroote M, Heimonen H, Kottmann J S, Menke T, Mok W K, Sim S, Kwek L C and Aspuru-Guzik A 2022 *Rev. Mod. Phys.* **94**(1) 015004 URL https://link.aps.org/doi/10.1103/RevModPhys.94.015004

[9] Cerezo M, Arrasmith A, Babbush R, Benjamin S C, Endo S, Fujii K, McClean J R, Mitarai K, Yuan X, Cincio L and Coles P J 2021 *Nature Reviews Physics* **3** 625–644 URL https://doi.org/10.1038/s42254-021-00348-9

[10] Kingma D P and Ba J 2017 Adam: A method for stochastic optimization (*Preprint* 1412.6980)

[11] Sweke R, Wilde F, Meyer J, Schuld M, Fährmann P K, Meynard-Piganeau B and Eisert J 2020 *Quantum* **4** 314 URL https://doi.org/10.22331/q-2020-08-31-314

[12] Wiersema R and Killoran N 2023 *Phys. Rev. A* **107**(6) 062421 URL https://link.aps.org/doi/10.1103/PhysRevA.107.062421

[13] Stokes J, Izaac J, Killoran N and Carleo G 2020 *Quantum* **4** 269 ISSN 2521-327X URL https://doi.org/10.22331/q-2020-05-25-269

[14] Wierichs D, Gogolin C and Kastoryano M 2020 *Phys. Rev. Res.* **2**(4) 043246 URL https://link.aps.org/doi/10.1103/PhysRevResearch.2.043246

[15] Choquette A, Di Paolo A, Barkoutsos P K, Sénéchal D, Tavernelli I and Blais A 2021 *Phys. Rev. Res.* **3**(2) 023092 URL https://link.aps.org/doi/10.1103/PhysRevResearch.3.023092

[16] Wang X, Okyay M S, Kumar A and Wong B M 2023 *AVS Quantum Science* **5** 043801 ISSN 2639-0213 (*Preprint* https://pubs.aip.org/avs/aqs/article-pdf/doi/10.1116/5.0162455/18150186/043801_1_5.0162455.pdf) URL https://doi.org/10.1116/5.0162455

[17] Park C Y and Killoran N 2024 *Quantum* **8** 1239 ISSN 2521-327X URL https://doi.org/10.22331/q-2024-02-01-1239

[18] Park C Y 2024 *APL Quantum* **1** 016101 ISSN 2835-0103 (*Preprint* https://pubs.aip.org/aip/apq/article-pdf/doi/10.1063/5.0186205/19692925/016101_1_5.0186205.pdf) URL https://doi.org/10.1063/5.0186205

[19] Larocca M, Ju N, García-Martín D, Coles P J and Cerezo M 2023 *Nature Computational Science* **3** 542–551 URL https://doi.org/10.1038/s43588-023-00467-6

[20] Livni R, Shalev-Shwartz S and Shamir O 2014 On the computational efficiency of training neural networks *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1* NIPS'14 (Cambridge, MA, USA: MIT Press) p 855–863

[21] Li D, Ding T and Sun R 2018 *ArXiv* **abs/1812.11039**

[22] Sim S, Johnson P D and Aspuru-Guzik A 2019 *Advanced Quantum Technologies* **2** 1900070 URL https://onlinelibrary.wiley.com/doi/abs/10.1002/qute.201900070

[23] Holmes Z, Sharma K, Cerezo M and Coles P J 2022 *PRX Quantum* **3**(1) 010313 URL https://link.aps.org/doi/10.1103/PRXQuantum.3.010313

[24] Cerezo M, Sone A, Volkoff T, Cincio L and Coles P J 2021 *Nature communications* **12** 1791 URL https://doi.org/10.1038/s41467-021-21728-w

[25] Nelder J A and Mead R 1965 *The computer journal* **7** 308–313

[26] Guerreschi G G and Smelyanskiy M 2017 *arXiv preprint arXiv:1701.01450* URL https://doi.org/10.48550/arXiv.1701.01450

[27] Powell M J *et al.* 2009 *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge* **26** 26–46

[28] Lavrijsen W, Tudor A, Müller J, Iancu C and De Jong W 2020 Classical optimizers for noisy intermediate-scale quantum devices *2020 IEEE international conference on quantum computing and engineering (QCE)* (IEEE) pp 267–277

[29] Smith A W, Paige A and Kim M 2023 *Quantum Science and Technology* **8** 045016 URL https://doi.org/10.1088/2058-9565/aceb87

[30] Yao Y 1989 *IEEE Trans. Syst. Man Cybern.* **19** 1222–1230 URL https://api.semanticscholar.org/CorpusID:28889897

[31] Cetin B, Barhen J and Burdick J 1993 *Journal of optimization theory and applications* **77** 97–126

[32] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 *New Journal of Physics* **18** 023023 URL http://dx.doi.org/10.1088/1367-2630/18/2/023023

[33] Cerezo M, Arrasmith A, Babbush R, Benjamin S C, Endo S, Fujii K, McClean J R, Mitarai K, Yuan X, Cincio L *et al.* 2021 *Nature Reviews Physics* **3** 625–644 URL https://doi.org/10.1038/s42254-021-00348-9

[34] Choi M S 2022 *A Quantum Computation Workbook* (Springer)

[35] Flammia S T and Liu Y K 2011 *Physical review letters* **106** 230501 URL https://doi.org/10.1103/PhysRevLett.106.230501

[36] Pfeuty P 1970 *ANNALS of Physics* **57** 79–90

[37] Rieger H and Young A P 1994 *Physical review letters* **72** 4141 URL https://doi.org/10.1103/PhysRevLett.72.4141

[38] Fisher D S 1995 *Physical review b* **51** 6411 URL https://doi.org/10.1103/PhysRevB.51.6411

[39] Kuzmak A 2021 *Quantum Information Processing* **20** 269 URL https://doi.org/10.1007/s11128-021-03196-9

[40] Lee J, Kim M S and Brukner i c v 2003 *Phys. Rev. Lett.* **91**(8) 087902 URL https://link.aps.org/doi/10.1103/PhysRevLett.91.087902

[41] Wächter A and Biegler L T 2006 *Mathematical programming* **106** 25–57

[42] Beck A and Teboulle M 2009 *SIAM journal on imaging sciences* **2** 183–202