Variational quantum time evolution without the quantum geometric tensor

Julien Gacon [1,2] Jannes Nys [2] Riccardo Rossi [3] Stefan Woerner [1] and Giuseppe Carleo [2]

[1]*IBM Quantum, IBM Research Europe – Zurich, CH-8803 Rüschlikon, Switzerland*

[2]*Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

[3]*Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75005 Paris, France*

Real- and imaginary-time quantum state evolutions are crucial in physics and chemistry for exploring quantum dynamics, preparing ground states, and computing thermodynamic observables. On near-term devices, variational quantum time evolution is a promising candidate for these tasks, as the required circuit model can be tailored to the available devices' capabilities. Due to the evaluation of the quantum geometric tensor (QGT), however, this approach quickly becomes infeasible for relevant system sizes. Here, we propose a dual formulation for variational time evolution, which replaces the calculation of the QGT by solving a fidelity-based optimization to compute updates to the dynamics in each time step. We demonstrate our algorithm for the time evolution of the Heisenberg Hamiltonian and show that it accurately reproduces the system dynamics at a fraction of the cost of standard variational quantum time evolution algorithms.

I. INTRODUCTION

Quantum time evolution is a central task in physics. Real-time evolution provides detailed insight into properties of quantum mechanical systems, such as phase transitions [1–3] or thermalization [4,5]. Imaginary-time evolution is an important tool that enables the preparation of ground states or thermal states [6–8]. These can, in turn, be used for the calculation of thermodynamic observables [8,9]. In particular, combining real- and imaginary-time evolution would allow the direct calculation of dynamical correlation functions at thermal equilibrium.

The range of applications of imaginary-time evolution extends beyond the field of physics. Ground-state preparation with imaginary-time evolution for gapped, nondegenerate Hamiltonians is guaranteed to converge in the generic case of nonzero overlap between the ground state and the initial trial state. This makes it a promising candidate in settings where a good initial state can be constructed, e.g., in chemistry applications [10] or in classical optimization problems [11]. In quantum machine learning, the preparation of Gibbs states with imaginary-time evolution is a subroutine for quantum Boltzmann machines, which can, for example, be used in distribution learning or classification [12].

Since performing quantum time evolution generally requires representing the exponentially large wave function of a quantum system, quantum computers are a promising platform for developing efficient algorithms [13]. In fact, in 1996, the Trotter algorithm for real-time evolution was among the first proposed use cases for a quantum computer [14]. However, the complexity of the quantum circuits required for the Trotter algorithm depends on the Hamiltonian, and the circuit depth scales with the simulation time and accuracy [15,16]. This renders the algorithm currently unsuitable for general time evolution on near-term devices, which are characterized by limited qubit connectivity and coherence times. The imaginary-time counterpart of Trotter suffers from the same restriction [8].

Variational algorithms for quantum time evolution, on the other hand, allow to choose a parameterized circuit as an ansatz to approximate the wave function that operates within the device's capabilities. Using a variational principle, variational quantum time evolution (VarQTE) maps the quantum state evolution to the evolution of parameters in the model [17], both for real-time evolution (VarQRTE) and imaginary-time evolution (VarQITE). The parameter update rules depend on the evaluation of the Quantum Geometric Tensor (QGT) and gradients of the current energy and state. For an ansatz with d variational parameters, the number of circuits required to evaluate the QGT and gradients scale as $\mathcal{O}(d^2)$ and $\mathcal{O}(d)$, respectively. While this does not pose a problem for small systems, the evaluation of the QGT quickly becomes a bottleneck once the system size, and therefore the number of variational parameters, increases.

Figure 1 shows a runtime estimate for VarQITE, assuming current superconducting processor specifications (see Appendix A for details on the derivation). For a few parameters the runtime is of the order of hours, however, increasing to a mere 200 parameters already renders the algorithm impractical as this would require approximately one week of computation time. In contrast, the algorithm proposed in this work holds the potential to reduce the runtime to only a single day. Such improvements of the quantum computational resource requirements become increasingly crucial to find
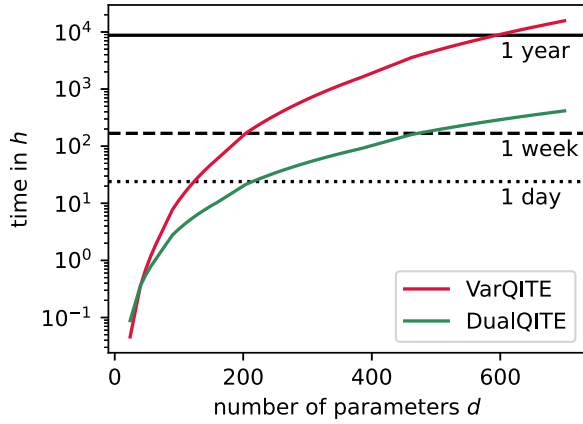
FIG. 1. Estimated runtimes of variational imaginary-time evolution (VarQITE) and our proposed dual method (DualQITE) as a function of the number of parameters d of the variational model, for an exemplary Heisenberg model and 200 time steps.

practically relevant applications for quantum computers as processor sizes keep increasing, recently achieving milestones beyond 1000 qubits [18].

Recently, focus has shifted to optimization-based algorithms, which implement partial steps or approximations of the full Suzuki-Trotter step [19–23]. In the case of real-time evolution, for example, the projected variational quantum dynamics (p-VQD) algorithm [19] provides a scalable alternative to VarQRTE on near term devices, if a single Trotter step can be efficiently implemented. However, the required quantum circuit gates in p-VQD reflect the couplings of the Hamiltonian. This means that, for Hamiltonians with long-distance interactions or numerous Pauli terms (e.g., in molecular dynamics), even a single step could involve global connections or deep circuits hindering the execution on near-term devices. Furthermore, the p-VQD algorithm is not directly applicable to imaginary-time evolution.

Other approaches concerned with real-time evolution are variational fast forwarding (VFF) methods [24–27] and classical pre-processing approaches [28,29]. VFF methods rely on diagonalizing the Hamiltonian or the Trotterized time evolution operator with a variational ansatz. However, finding the diagonalizing unitary remains challenging in practice, which limits demonstrations to very few qubits. Classical pre-processing techniques, on the other hand, impose additional restrictions on the simulated system, such as translational invariance [29] or Hamiltonians with low entanglement [28]. Within such systems, these techniques scale to large systems, but they do not allow for general quantum time evolution.

Another line of work directly focuses on the preparation of thermal states by minimizing the free energy of a variational ansatz [30]. This approach, however, also does not implement general quantum time evolution.

In this paper, we propose a novel variational algorithm for quantum time evolution based on a *dual* optimization problem, which allows to replace the QGT by evaluating the overlap of the variational ansatz for different parameter values. This formulation applies equally real- and imaginary-time evolution and does not require additional qubits or connections than already present in the ansatz. We show that

this new algorithm requires significantly fewer measurements and thereby drastically reduces the expected runtime compared to VarQTE. This is summarized in Fig. 1, where, under the same assumptions, our proposed method can reduce the expected runtimes from several weeks for VarQTE to only a few days. Following the naming conventions of VarQTE, we name the algorithm DualQTE with specifiers DualQITE for imaginary-time evolution and DualQRTE for real-time evolution.

The remainder of this paper is structured as follows. In Sec. II, we recap VarQTE based on variational principles, derive the proposed dual formulation, and discuss how to implement it on a quantum computer. Then, in Sec. III, we demonstrate our proposed algorithm for the imaginary-time evolution of the Heisenberg model and investigate the resource requirements. As a practical application, we use the quantum minimally entangled typical thermal states method (QMETTS) to calculate thermodynamic observables. Sec. IV demonstrates the dual formulation for real-time evolution, including the calculation of variational error bounds. Finally, Sec. V concludes the paper and gives an outlook on possible applications and further research directions.

II. DUAL FORMULATION OF VARIATIONAL TIME EVOLUTION

For a time-independent Hamiltonian H acting on n qubits, an initial quantum state $|\Psi_0\rangle$ and an evolution time t, the real-time evolved quantum state is

$$|\Psi(t)\rangle = e^{-itH} |\Psi_0\rangle. \tag{1}$$

For an imaginary-time evolution, the time evolution operator is nonunitary, and the normalized state reads

$$|\Psi(t)\rangle = \frac{1}{\sqrt{\langle\Psi_0|e^{-2tH}|\Psi_0\rangle}} e^{-tH} |\Psi_0\rangle. \tag{2}$$

Variational quantum time evolution maps the evolution of the quantum state $|\Psi(t)\rangle$ to the evolution of parameters $\boldsymbol{\theta}(t) \in \mathbb{R}^d$ of a parameterized quantum state $|\phi(\boldsymbol{\theta}(t))\rangle$. The parameters' dynamics can be derived with variational principles such as the Dirac-Frenkel, McLachlan, or time-dependent variational principle [17]. In McLachlan's formulation, the derivatives of the parameters are determined by the linear system of equations

$$g(\boldsymbol{\theta}(t)) \dot{\boldsymbol{\theta}}(t) = \boldsymbol{b}(\boldsymbol{\theta}(t)), \tag{3}$$

where the matrix $g = \text{Re}(G) \in \mathbb{R}^{d \times d}$ is the real part of the QGT, and we call $\boldsymbol{b} \in \mathbb{R}^d$ the evolution gradient.

The QGT is defined as

$$G_{ij}(\boldsymbol{\theta}) = \langle\partial_i\phi(\boldsymbol{\theta})|\partial_j\phi(\boldsymbol{\theta})\rangle - \langle\partial_i\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\langle\phi(\boldsymbol{\theta})|\partial_j\phi(\boldsymbol{\theta})\rangle, \tag{4}$$

where we use the notation $\partial_i := \partial/(\partial\theta_i)$ and do not explicitly state the time dependence of the parameters. The evolution gradient for VarQRTE is given by the expression

$$b_i^{\text{R}}(\boldsymbol{\theta}) = \text{Im}\big(\langle\partial_i\phi(\boldsymbol{\theta})|H|\phi(\boldsymbol{\theta})\rangle - \langle\partial_i\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle E(\boldsymbol{\theta})\big), \tag{5}$$

whereas a VarQITE evolution yields the following:

$$b_i^{\text{I}}(\boldsymbol{\theta}) = -\text{Re}\big(\langle\partial_i\phi(\boldsymbol{\theta})|H|\phi(\boldsymbol{\theta})\rangle\big) = -\frac{\partial_i E(\boldsymbol{\theta})}{2}, \tag{6}$$

with the energy $E(\boldsymbol{\theta}) = \langle \phi(\boldsymbol{\theta})|H|\phi(\boldsymbol{\theta})\rangle$. From hereon, we present general equations that apply to both real and imaginary-time evolution; thus, unless specified, we simply use $\boldsymbol{b}$ without a specific superscript.

Note that these equations are introduced for a time-independent Hamiltonian, but they can also be applied to the time-dependent case $H = H(t)$.

A. Dual formulation

Instead of solving the linear system defined in Eq. (3), we propose to solve the dual formulation of the problem [31,32] given by

$$\dot{\boldsymbol{\theta}} = \arg\min_{\dot{\boldsymbol{\theta}}} \frac{\dot{\boldsymbol{\theta}}^T g(\boldsymbol{\theta})\dot{\boldsymbol{\theta}}}{2} - \dot{\boldsymbol{\theta}}^T \boldsymbol{b}(\boldsymbol{\theta}). \tag{7}$$

The term $\|\dot{\boldsymbol{\theta}}\|^2_{g(\boldsymbol{\theta})} = \dot{\boldsymbol{\theta}}^T g(\boldsymbol{\theta})\dot{\boldsymbol{\theta}}$ is the squared norm of the parameter derivative in the metric of the QGT. This quantity describes the magnitude of the derivative from an information geometric point of view and is derived from the Fubini-Study metric. For infinitesimal displacements $\delta\boldsymbol{\theta}$, we have

$$\begin{aligned} \|\delta\boldsymbol{\theta}\|^2_{g(\boldsymbol{\theta})} &= \delta\boldsymbol{\theta}^T g(\boldsymbol{\theta})\delta\boldsymbol{\theta} \\ &= 1 - |\langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle|^2 + \mathcal{O}(\|\delta\boldsymbol{\theta}\|^3_2), \end{aligned} \tag{8}$$

where $\|\cdot\|_2$ is the ℓ_2 norm [32], see Appendix B for the derivation. By writing $\dot{\boldsymbol{\theta}} = \delta\boldsymbol{\theta}/\delta\tau$, for some time perturbation $\delta\tau > 0$, we can now reformulate the optimization in terms of the fidelity $F(\boldsymbol{\theta}, \boldsymbol{\theta}') = |\langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta}')\rangle|^2$ as

$$\begin{aligned} \delta\boldsymbol{\theta} &\approx \arg\min_{\delta\boldsymbol{\theta}} \frac{1 - F(\boldsymbol{\theta}, \boldsymbol{\theta} + \delta\boldsymbol{\theta})}{2(\delta\tau)^2} - \frac{\delta\boldsymbol{\theta}^T \boldsymbol{b}(\boldsymbol{\theta})}{\delta\tau} \\ &= \arg\min_{\delta\boldsymbol{\theta}} \frac{\mathcal{L}(\delta\boldsymbol{\theta})}{(\delta\tau)^2}, \end{aligned} \tag{9}$$

where we directly optimize for the parameter update $\delta\boldsymbol{\theta}$ and introduce the loss function

$$\mathcal{L}(\delta\boldsymbol{\theta}) = \frac{1 - F(\boldsymbol{\theta}, \boldsymbol{\theta} + \delta\boldsymbol{\theta})}{2} - \delta\tau \cdot \delta\boldsymbol{\theta}^T \boldsymbol{b}(\boldsymbol{\theta}). \tag{10}$$

In practice, the optimization problem can be solved without the factor $(\delta\tau)^{-2}$, which decouples the shape of the locally quadratic infidelity term from the time perturbation and improves the numerical stability of the optimization.

Note that this dual formulation can alternatively be obtained from the derivation of quantum natural gradients [30,32], which is detailed in Appendix C. For an intuitive understanding of the relationship of the infidelity and QGT the effect of approximating $\|\delta\boldsymbol{\theta}\|^2_{g(\boldsymbol{\theta})} \approx 1 - F(\boldsymbol{\theta}, \boldsymbol{\theta} + \delta\boldsymbol{\theta})$ in an illustrative example is demonstrated in Appendix D.

Instead of computing the QGT at each time step, which requires $\mathcal{O}(d^2)$ circuit evaluations, we now have to solve an optimization problem where the loss function requires only one fidelity evaluation. The required resources of DualQTE per time step are therefore $\mathcal{O}(d)$ for the computation of the evolution gradient $\boldsymbol{b}$, times the number of iterations in the optimization. Thus we improve upon the direct QGT approach if the number of iterations scales better than $\mathcal{O}(d)$, which, as we show in the following sections, is the case for the examples we investigate in this work.

B. Evaluating the loss function

The evaluation of the loss function $\mathcal{L}$, defined in Eq. (10), requires the calculation of the evolution gradient $\boldsymbol{b}$ and the fidelity of the ansatz $|\phi(\boldsymbol{\theta})\rangle$ for two different parameter sets. For imaginary-time evolution, the evolution gradient can, for example, be evaluated with analytic gradient rules, such as the parameter-shift rule or a linear combination of unitaries (LCU), or with finite difference methods [33]. In the case of real-time evolution, however, we are restricted to an LCU approach, as this is the only method that allows the calculation of the imaginary part of gradients [12].

The fidelity F can, for example, be estimated using the swap test [34] and its variants [35], where the states are prepared in separate qubit registers followed by entangling gates across these registers, or with the Hadamard test, which adds only a single auxiliary qubit, but requires controlling the state-preparing unitary [36]. A more near-term-friendly option is the compute-uncompute method [37], which does not introduce additional global operations. If the states are given by $|\phi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle$ for a parameterized unitary U and two different parameter values $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, the fidelity can be calculated by preparing $U^\dagger(\boldsymbol{\theta})U(\boldsymbol{\theta}')|0\rangle$ and measuring the probability of obtaining $|0\rangle$ on all qubits.

If the state $|\phi(\boldsymbol{\theta})\rangle$ has n qubits and the preparing unitary U has depth m, the swap test variants require a circuit width of $2n$ with depth of $m + \mathcal{O}(1)$, whereas the evaluated circuits for the compute-uncompute method are of only width n, but of depth $2m$. The Hadamard test for fidelities between the same circuit with different parameters can be evaluated by controlling the parameterized gates, resulting in depth of m and depth of $n + 1$, plus the overhead of controlling the gates. For sparse device connectivities, this can be a challenge. To avoid increasing the circuit complexity, the overlap can also be estimated via randomized measurements of two independent state preparations [38]. However, this technique requires an exponential number of measurements.

Evaluating the QGT for VarQTE, however, suffers from similar issues. The QGT can be evaluated as the Hessian of the infidelity [39] using a parameter-shift or finite difference technique, which comes with the restrictions for fidelity evaluations as described above. Alternatively, Eq. (4) can be directly computed with an LCU approach, which adds two auxiliary qubits and two entangling gates [12]. This method is less demanding than, e.g., a Hadamard test, but still comes with additional connectivity requirements. In practice, for both VarQTE and DualQTE, a suitable combination of parameterized quantum state $|\phi(\boldsymbol{\theta})\rangle$ and gradient method must be selected, such that the resulting circuits can be executed reliably.

Depending on the topology and coherence times of the available hardware and the structure and size of the unitary, either method for gradient and fidelity calculations can be advantageous. In this work, we focus on near-term friendly methods and use the parameter-shift rule for gradients (if possible) and the compute-uncompute method for the fidelity, as these do not require additional gate connections or an exponential number of measurements. Note that, for systems with a large number of qubits, this method might become unsuitable as it measures the global zero projector. Then, approaches using only local measurements, such as the Hadamard test, could be the better choice.

C. Solving for the update step

The infidelity-based loss function $\mathcal{L}$ is locally convex around $\delta\boldsymbol{\theta} = \mathbf{0}$, as its Hessian at this point is $\nabla\nabla^T\mathcal{L}(0) = g/2$, and g is positive semi-definite. To leverage this property, we use gradient descent as a local optimization routine, which also allows the use of analytic gradient formulas that have proven more stable in presence of shot noise. The gradient of $\mathcal{L}$ with respect to the parameter update $\delta\boldsymbol{\theta}$ is

$$\nabla_{\delta\boldsymbol{\theta}}\mathcal{L}(\delta\boldsymbol{\theta}) = -\frac{\nabla_{\delta\boldsymbol{\theta}}F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta})}{2} - \delta\tau \cdot \boldsymbol{b}(\boldsymbol{\theta}).$$

The gradient of the fidelity can be evaluated with a parameter-shift rule

$$\frac{\partial F}{\partial(\delta\theta)_i} = \frac{F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}+\boldsymbol{e}_i s) - F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}-\boldsymbol{e}_i s)}{2\sin(s)},$$

where $\boldsymbol{e}_i$ is the ith unit vector and s is the parameter shift, which can be chosen as, e.g., $\pi/2$ for single-qubit Pauli rotations [33].

At each time step, the gradient descent update for the update step $\delta\boldsymbol{\theta}$ is

$$\delta\boldsymbol{\theta}^{(k+1)} = \delta\boldsymbol{\theta}^{(k)} - \eta_k \nabla\mathcal{L}(\delta\boldsymbol{\theta}^{(k)}),$$

where $\eta_k > 0$ is the learning rate at step k. This iteration is continued until a maximum number of iterations or a convergence criterion is met. An example of the latter is a minimum tolerance in the change of the cost function or the norm of the gradient.

An intuitive choice for the initial guess $\delta\boldsymbol{\theta}^{(0)}$ is the zero vector, which corresponds to no change in the parameters. However, a more efficient choice can be to introduce momentum by warm starting the optimization with the update step from the previous time step. This heuristic is motivated by the idea that, especially for small time steps, we do not expect the parameter derivatives $\dot{\boldsymbol{\theta}}$ to change significantly.

Methods that approximate the gradient, such as finite difference or SPSA [40], may face challenges in the optimization. For small time steps, the fidelity is close to 1 and the noise in the readout, e.g., from finite sampling statistics or device noise, can easily mask changes in the cost function. Parameter-shift gradients suffer less from this problem, as they allow to evaluate the cost function over larger perturbations, and do not amplify the noise by dividing by a small constant.

The ideal choice of the time perturbation $\delta\tau$ is a trade-off: The error in approximating the QGT scales as $(\delta\tau)^3$, but a smaller perturbation amplifies any measurement noise in the loss function as the update step is obtained as $\delta\boldsymbol{\theta}/\delta\tau$. Appendix D displays this trade-off using an illustrative example.

D. Trainability

Recently, there has been a lot of research showing that, in certain settings, the loss function gradients of variational algorithms decay to zero exponentially and cannot be evaluated efficiently, as they would require an exponential number of measurements. These so-called barren plateaus can be induced, for example, if the loss function requires measuring a global observable [41], if the quantum circuit preparing the parameterized state is too deep or generates too much

entanglement [41–43], or if the measurements are too noisy [44].

Since variational quantum dynamics is driven by the evolution gradient defined Eqs. (5) and (6), it can be affected by a barren plateau and fail to track the true evolution of the quantum state. However, it is important to note that the gradients only vanish on average for a random initialization, whereas in time-evolution the initial quantum state is typically specifically chosen. Furthermore, Hamiltonians of physical systems are usually local, as they reflect the interactions of the quantum mechanical system, and exponentially vanishing gradients can be avoided by choosing a circuit depth scaling logarithmically in system size [41]. Alternatively, an application-specific ansatz with few variational parameters can help mitigate barren plateaus, such as circuits based on Hamiltonian evolutions [45,46].

In addition to the evolution gradient, the DualQTE loss function gradient $\nabla_{\delta\boldsymbol{\theta}}\mathcal{L}$ depends on the gradient of the fidelity, which relies on measuring a global observable. This can be seen by writing the fidelity of two n-qubit states prepared by unitaries $U(\boldsymbol{\theta})$ and $U(\boldsymbol{\theta}')$ as $|\langle 0|U^\dagger(\boldsymbol{\theta})U(\boldsymbol{\theta}')|0\rangle|^2 = \langle\lambda|P_0|\lambda\rangle$, where $|\lambda\rangle = U^\dagger(\boldsymbol{\theta}')U(\boldsymbol{\theta})|0\rangle$ and $P_0 = |0\rangle\langle 0|^{\otimes n}$ is the global projector on the all-zero state. Thus evaluating the fidelity gradient for two randomly selected parameter sets $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ would exhibit barren plateaus at any circuit depth [41]. However, the optimization in DualQTE starts at zero perturbations, $\boldsymbol{\theta} = \boldsymbol{\theta}'$ where the total state preparing unitary is the identity, $|\lambda\rangle = U^\dagger(\boldsymbol{\theta})U(\boldsymbol{\theta})|0\rangle = \mathbb{I}|0\rangle$, which is an initialization that is proven to not exhibit barren plateaus even for global cost functions [47]. Together with the fact that the DualQTE loss function is locally convex, the nonvanishing gradients at the initial point of the optimization is a strong motivation for the efficient trainability of DualQTE.

In Appendix F 4, we provide numerical evidence that for a local Hamiltonian and a logarithmic-depth circuit, neither the evolution gradient or the fidelity gradients decay exponentially with system size. We stress, however, that a careful selection of the circuit and initial point are crucial even if the gradients do not vanish exponentially. If the circuit is unable to track the dynamics, or the evolution gradient of the McLachlan variational principle is zero VarQTE and DualQTE cannot produce accurate results.

E. Sample complexity

The implementation of VarQTE on quantum hardware has several sources of errors: the model $|\phi(\boldsymbol{\theta})\rangle$ could lack expressivity to capture the dynamics, the time integration scheme introduces errors, the QGT and evolution gradient are subject to sampling error from a finite number of measurement, and each operation is affected by hardware noise. If we denote the ideal VarQTE parameters without sampling or hardware noise by $\boldsymbol{\theta}(t)$ and the noisy parameters by $\tilde{\boldsymbol{\theta}}(t)$, the error contributions can be split as

$$\begin{aligned}\varepsilon(t) &= D_B(\phi(\tilde{\boldsymbol{\theta}}(t)), \Psi(t)) \\ &\leqslant \varepsilon_M(t) + \varepsilon_S(t),\end{aligned} \tag{11}$$

where we measured the error in Bures distance

$$D_B(\psi, \phi) = \sqrt{2(1 - |\langle\psi|\phi\rangle|)}, \tag{12}$$

and distinguish in error due to lack of model expressivity plus integration error $\varepsilon_M(t) = D_B(\phi(\boldsymbol{\theta}(t)), \Psi(t))$ and error due to a noisy implementation of VarQTE $\varepsilon_S(t) = D_B(\phi(\tilde{\boldsymbol{\theta}}(t)), \phi(\boldsymbol{\theta}(t)))$ [48].

Since the proposed DualQTE algorithm promises a reduction in measurement cost of VarQTE, but is not concerned with the ansatz selection or hardware noise, we here focus on investigating the scaling of the sampling error. The model error ε_M can be bounded with *a posteriori* errorbounds [49], which we also investigate for the real-time evolution case in Sec. IV.

Deriving a concrete bound in terms of system quantities such as the energy or the number of parameters requires an assumption on the circuit structure. Here we assume a circuit with only Pauli rotations $R_P(\theta_i)$ where each parameter θ_i is unique and does not have coefficients. Note that the bounds can be adjusted for different circuit structures and parametrizations. In addition, we assume a cutoff $\delta_c > 0$ on the smallest eigenvalue of g due to a regularization of the linear system.

Then, we can state the following upper bound on the number of samples required to achieve a sampling error of ε_S,

$$N \leqslant \mathcal{O}\left(\frac{d^3 E_{\max}^2 \Delta_t^2}{\delta_c^4 \varepsilon_S^2}\right), \qquad (13)$$

where $E_{\max}$ is the maximal eigenvalue of the Hamiltonian.

In contrast to a similar approach described in Ref. [48], we state the upper bound in terms of the number of parameters in the model or the system's Hamiltonian, instead of the QGT and evolution gradients. Further, we are able to derive a tighter result by leveraging Latala's theorem from random matrix theory to upper bound the sampling error in g [50].

Since the DualQTE algorithm does not construct the QGT directly but only the evolution gradient, we expect a reduction of a factor d in the complexity, and an additional factor for the number of optimization steps K in each time step. Indeed, we can show that the upper bound for the number of samples is

$$N \leqslant \mathcal{O}\left(\frac{d^2 K^2 \Delta_t^2}{\varepsilon_S^2}\left(\frac{1}{\delta\tau} + E_{\max}\right)^2\right). \qquad (14)$$

The detailed derivation of both bounds is described in Appendix E.

While it is possible to construct circuits where each component of these bounds are tight Sec. III shows that in practice the actual number of required samples scales less than this upper bound, which is further discussed in Appendix.

III. IMAGINARY-TIME EVOLUTION

In this section, we show the results of DualQTE and investigate the circuit costs compared to VarQTE. As an application, we use our algorithm as a subroutine to prepare typical thermal states of the Heisenberg model, which are then used to calculate the energy per site as a thermodynamic observable. All circuits are constructed and simulated using QISKIT [51].

A. Heisenberg model

We simulate the imaginary-time evolution of the Heisenberg model with nearest-neighbor interaction on a 12-qubit circle in a transverse field:

$$H = J \sum_{\langle ij \rangle} \left(X_i X_j + Y_i Y_j + Z_i Z_j\right) + h \sum_i Z_i, \qquad (15)$$

with interaction strength $J = 1/4$ and transversal field strength $h = -1$. As a variational ansatz, we use a circuit with Pauli-Y and Pauli-Z single qubit rotation layers that alternate with pairwise CNOT entangling gates. The circuit structure is shown in Appendix F 1 and we use $r = 3$ repetitions. The initial state for the evolution is the equal superposition of all qubits, $|+\rangle^{\otimes n}$, which we prepare by setting the rotation angles of the last Pauli-Y layer to $\pi/2$ and the remaining angles to 0.

The optimization problems in DualQTE are solved using gradient descent with a fixed learning rate of $\eta = 0.1$ and time perturbation $\delta\tau = 0.01$. The initial iteration performs 100 update steps and the subsequent, warm-started iterations, only 10. These values are motivated by the simulations shown in Appendix F 2 and are partially heuristic, as a termination criterion is challenging to define with access only to noisy loss function and gradient evaluations. The parameters are integrated with an explicit Euler scheme with time step $\Delta_t = 0.01$, i.e.,

$$\boldsymbol{\theta}(t + \Delta_t) = \boldsymbol{\theta}(t) + \Delta_t \dot{\boldsymbol{\theta}}(t) = \boldsymbol{\theta}(t) + \Delta_t \frac{\delta\boldsymbol{\theta}}{\delta\tau}.$$

Note that the integration time step Δ_t, which determines the accuracy of the time integration, can be chosen differently from the time perturbation $\delta\tau$, which affects the approximation error of the QGT metric with the infidelity.

We compare the performance to VarQITE with the same integration scheme and use an L-curve regularization [52] for a stable solution of the linear system. Among all regularization techniques we attempted, such as adding a diagonal shift, truncating small or negative singular values or solving on a stable subsystem, the L-curve regularization provided the most accurate and stable results.

In Fig. 2(a), we present the results for a varying number of shots along with the exact time evolution based on exact diagonalization. Already with as little as 100 measurements per circuit evaluation (shots) on the 12-qubit model, the dual time evolution is able to qualitatively follow the imaginary-time evolution and, up to time $t \approx 1$, even outperform VarQITE with 1024 shots. Increasing the number of measurements of DualQITE to 1024 shots allows the dual method to closely track the exact solution towards the ground state, with a higher accuracy than VarQITE with 8192 shots.

B. Resource requirements

In the above experiment, DualQITE requires fewer circuit evaluations to achieve the same accuracy as VarQITE. To investigate the total resource requirements, we perform both DualQITE and VarQITE with different resources and compute the achieved error. Since we are interested in following the imaginary-time dynamics as closely as possible at each time step, we define the error as the average integrated Bures
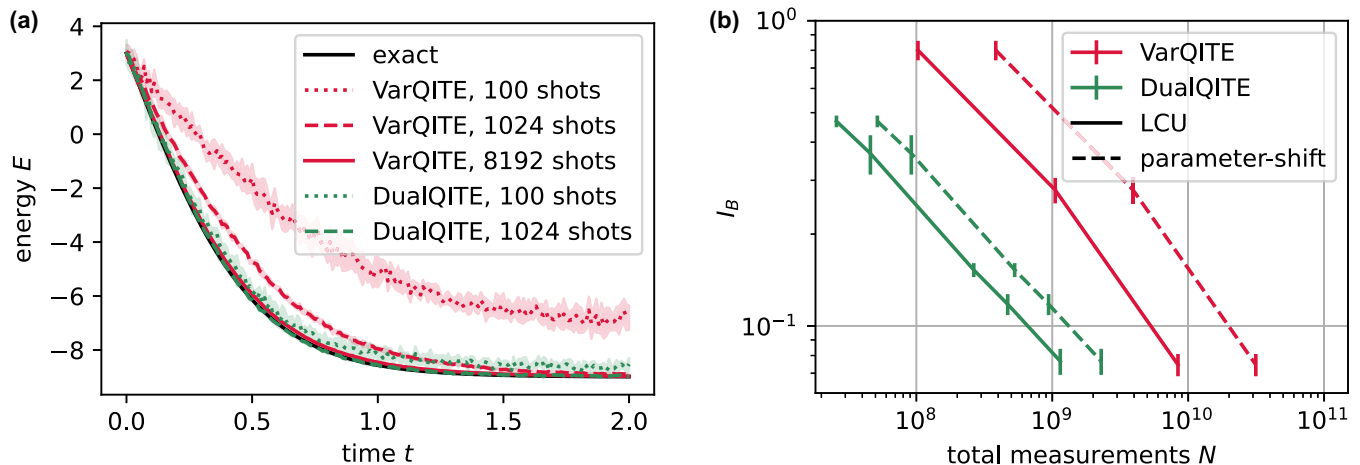
FIG. 2. (a) The mean and standard deviation of DualQITE and VarQITE, each averaged over five independent experiments for a varying number of shots. (b) The accuracy measured in integrated Bures distance I_B [see Eq. (16)] for DualQITE and VarQITE, with mean and standard deviation of five experiments. The resources are measured in total number of measurements and are shown for evaluation of gradients (and the QGT, in the case of VarQITE) using parameter-shift rules (dashed) or a LCU approach (solid lines).

distance to the exact solution over the time evolution,

$$I_B(T) = \frac{1}{T} \int_0^T D_B(\phi(\boldsymbol{\theta}(t)), \Psi(t)) dt. \tag{16}$$

The state fidelity is computed exactly, i.e., we compute the state vector of the model $|\phi\rangle$ at variational parameters $\boldsymbol{\theta}(t)$, and take the inner product with the exact time-evolved state $|\Psi(t)\rangle$.

The results for an integration time of $T = 2$ are shown in Fig. 2(b). We show the integrated Bures distance with respect to the total number of measurements recorded during the time evolution. In DualQITE, the resources can be split between using more optimization steps in each time step or more shots to evaluate the gradients. The algorithm settings are detailed in Appendix F 3. The figure shows the resource counts for gradient calculations via the parameter-shift rule (PSR) and linear combination of unitaries (LCU). The LCU technique requires additional auxiliary qubits and additional nonlocal operations, but less overall circuits than PSR. For P Pauli terms in the Hamiltonian, the total number of required circuits C per time step

$$C_{\text{LCU}}^{\text{VarQITE}} = \frac{d(d+5)}{2} + Pd,$$

$$C_{\text{PSR}}^{\text{VarQITE}} = 2d(d + P + 1). \tag{17}$$

For DualQITE, the number of circuits is

$$C_{\text{LCU}}^{\text{DualQITE}} = Pd + Kd, \tag{18}$$

and $C_{\text{PSR}}^{\text{DualQITE}} = 2C_{\text{LCU}}^{\text{DualQITE}}$, where K is the number of optimization steps per time step. The total number of measurements N is obtained by multiplying the number of circuits with the number of shots.

We see that, on average, DualQITE requires about one order of magnitude fewer measurements to achieve the same accuracy as VarQITE. With an increasing number of parameters, we expect this difference to grow, since VarQITE scales as $\mathcal{O}(d^2)$ whereas our algorithm, with warm starting, only computes small corrections at each time step.

C. Sample complexity

In addition to the fixed-size model with 12 qubits, we investigate how the resource requirements scale with system size. We compare VarQITE and DualQITE for the Heisenberg model from Eq. (15) with varying number of spins n and the same circuit structure as before, but with an adjusted number of repetitions of $r = \lceil \log_2(n) \rceil$ times, plus a final rotation layer. We then tune the settings of VarQITE and DualQITE to achieve a mean accuracy of $I_B \leqslant 0.1$ over five experiments and count the total number of required measurements N. This threshold corresponds to a per-time step fidelity of 0.995.

The results are presented in Fig. 3, which show the improved scaling of DualQITE compared to VarQITE. For small system sizes and few parameters, the overhead of solving the optimization problem in DualQITE is larger than evaluating the QGT. However, as we increase the problem size, the quadratic scaling of VarQITE takes over and our algorithm becomes more efficient.

This experiment allows to validate the upper bound on the number of measurements of Sec. II E. As shown in Fig. 3, the model error ε_M is negligible in comparison to the sampling error ε_S and we approximately have $\varepsilon_S \approx I_B \approx 0.1$. The maximal energy of the Heisenberg model on a periodic chain scales with the number of spins, and can be bounded by $E_{\max} = \mathcal{O}(n) \leqslant \mathcal{O}(n \log_2(n)) = \mathcal{O}(d)$. Inserting these values in Eqs. (13) and (14), we expect the scaling to be upper bounded by $\mathcal{O}(d^5)$ for VarQITE and $\mathcal{O}(d^4 K)$ for DualQITE. In practice, we observe approximately a scaling of $d^{3.77}$ for VarQITE and $d^{2.37}$ for DualQITE, which shows the expected improved scaling for our algorithm. The measured scaling also suggests that the bounds are not yet tight, which we discuss further in Appendix E.

D. Calculating thermodynamic observables

As an application of imaginary-time evolution, we calculate thermodynamic observables using the quantum minimally entangled thermal states algorithm (QMETTS) [8,53]. While the classical METTS algorithm has been specifically devel-
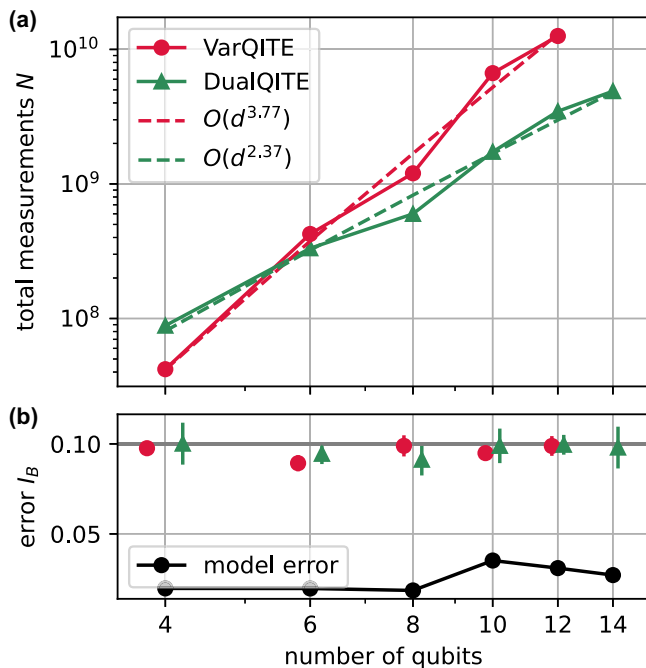
FIG. 3. (a) Total number of measurements required to achieve a mean accuracy of $I_B \leqslant 0.1$ over an average of five experiments. See Table II for the exact algorithm settings. Dotted and dashed lines show fits for the number of measurements. The bumps in the fits are due to the discontinuity of the circuit depth, which depends on $\lceil \log_2(n) \rceil$. VarQITE is not evaluated for $n = 14$ qubits as it requires too many measurements. (b) Mean accuracy and standard deviation of each point of the top panel. The grey line indicates the infidelity threshold of $I_B = 0.1$.

oped for Matrix Product State simulations, the thermal state preparation is still costly, and classical simulations fail if the system produces macroscopic entanglement during the imaginary time evolution (e.g., for low-temperature, 2D systems). Due to these restrictions, QMETTS is a promising application for quantum imaginary-time evolution algorithms.

For an observable A and inverse temperature β, the QMETTS algorithm generates samples $\{A_m\}_m$ using a Markov chain whose average approximate the ensemble average:

$$\langle A \rangle_{\text{ens}} = \frac{\text{Tr}(e^{-\beta H} A)}{\text{Tr}(e^{-\beta H})} \approx \frac{1}{M} \sum_{m=1}^{M} A_m.$$

The sampling process to obtain the sample A_m is the following.

(1) Start from a product state $|\phi_m(t = 0)\rangle$.
(2) Evolve up to imaginary time $t = \beta/2$

$$|\phi_m(\beta/2)\rangle \propto e^{-\beta H/2} |\phi_m(0)\rangle .$$

(3) Evaluate the observable to obtain the sample

$$A_m = \langle \phi_m(\beta/2) | A | \phi_m(\beta/2) \rangle .$$

(4) Measure $|\phi_m(\beta/2)\rangle$ in some basis to obtain the next random product state $|\phi_{m+1}(0)\rangle$.

We investigate the Heisenberg model from Eq. (15) on a chain with $n = 6$ spins with parameters $J = 1/4$ and $h = -1$. As a thermodynamic observable we compute the energy
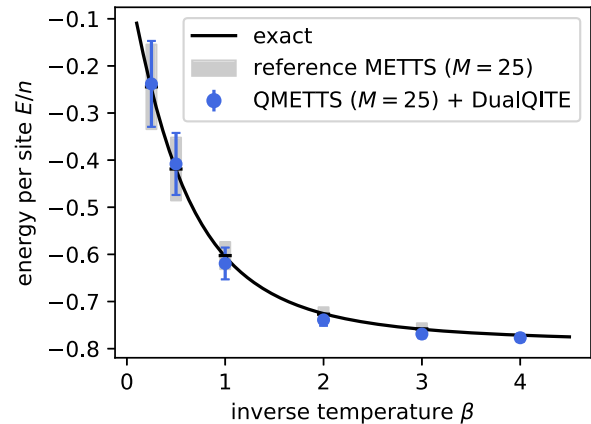

FIG. 4. Energy per site for the Heisenberg model on a six-spin chain, comparing mean and standard deviation of QMETTS with DualQITE (blue circle and errorbars) with a reference METTS implementation (black line and grey shade).

per site, $\langle H \rangle /n$. To reduce the auto-correlation length in the QMETTS Markov chain, and for faster convergence to the ensemble average, it is favorable to measure in different bases in each step. Since the Heisenberg Hamiltonian conserves the number of qubits in the $|1\rangle$ state, avoiding the Z basis greatly reduces the standard deviation of the Markov chain. Thus we here alternate between the X and Y basis for each sample.

As ansatz for DualQITE, we use problem-inspired circuits with pairwise CNOT couplings and $r = 2$ repetitions of rotation and entanglement layers, plus a final rotation layer, see Appendix F 1 for a circuit diagram. For evolutions of product states $|\pm\rangle$ in the X basis, the rotation layers are single qubit $R_Y R_Z$ gates, and for the states $|\pm i\rangle$ in the Y basis, the layers implement $R_X R_Z$ gates. The initial product states $|\phi_m(0)\rangle$ are prepared by setting the parameters in the final layer rotation layer of the ansatz as follows:

$$|\pm\rangle \rightarrow R_Y\left(\frac{\pm\pi}{2}\right) R_Z(0),$$

$$|+i\rangle \rightarrow R_X\left(\frac{\pi}{2}\right) R_Z(\pi),$$

$$|-i\rangle \rightarrow R_X\left(\frac{\pi}{2}\right) R_Z(0).$$

Each energy sample is evaluated with 1024 measurements per basis. The optimization problem in DualQITE is solved with a time perturbation $\delta\tau = 0.01$ and gradient descent with a learning rate of $\eta = 0.1$ and 100 iterations in the first time step, followed by 10 iterations in the following, warm-started time steps. We integrate with a fixed time step of $\Delta_t = 0.01$.

Figure 4 shows the estimated energy per site, along with the standard deviation of the samples, for different inverse temperatures β. For the alternating X-Y basis, the Markov chain converges quickly and $M = 25$ samples suffice for an accurate estimate of the observable. For the imaginary-time evolution, we compare DualQITE with the same settings as in the previous sections to an exact evolution performed with matrix exponentials. It shows that using the dual method allows to reliably reproduce the mean and standard deviation of the Markov chain samples compared to the exact reference METTS.

IV. REAL-TIME EVOLUTION

The focus of this paper is on imaginary-time evolution as, to date, no other QGT-free time evolution algorithms exist in this setting. For real-time evolution, p-VQD has a similar structure as our algorithm and solves an optimization problem rather than evaluating the QGT. However, there are key differences to the dual algorithm applied to real-time evolution.

The p-VQD algorithm [19] projects a single Suzuki-Trotter step onto the circuit model by solving the following optimization problem:

$$\boldsymbol{\theta}(t + \Delta_t) = \underset{\boldsymbol{\theta}'}{\text{argmax}} \left| \langle \phi(\boldsymbol{\theta}')|e^{-iH\Delta_t}|\phi(\boldsymbol{\theta}(t))\rangle \right|^2.$$

For Hamiltonians with many Pauli terms or long-range interactions, such as those arising in molecular dynamics, the single step might already lead to large circuits with nonlocal gates. While DualQRTE requires an LCU method to evaluate the imaginary part of the energy and state gradients, see Eq. (5), this only adds a single entangling gate, compared to a full Suzuki-Trotter step is required. Furthermore, our dual time evolution allows the evaluation of error bounds at almost no additional cost, which is not possible in p-VQD. Due to these differences, this section presents DualQRTE: the dual time evolution for real-time evolution.

A. Heisenberg model

We present the real-time evolution under the Heisenberg Hamiltonian of Eq. (15) on a linear chain with $n = 4$ spins with parameters $J = 1/4$, $h = -1$. As variational model, we use a circuit with alternating Pauli-X and Pauli-Y rotation layers, and Pauli-ZZ entangling gates that reflect the connectivity of the spins. The circuit structure is visualized in Appendix G and, in this experiment, all algorithms use $r = 3$ repetitions of the rotation as well as entangling gates. To prepare the initial state, $|+\rangle^{\otimes 4}$, we set the parameters of the final Pauli-Y rotations to $\pi/2$ and the rest to 0.

During the evolution, we track the average magnetization in the X and Z direction,

$$\langle X \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle X_i \rangle, \quad \langle Z \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle Z_i \rangle.$$

Since this Heisenberg Hamiltonian preserves the qubit excitations, and the initial state is the equal superposition, the $\langle Z \rangle$ expectation value should remain 0 throughout the evolution.

The results of the different time evolution algorithms for an integration time of $T = 2$ and time step $\Delta_t = T/100$ are presented in Fig. 5. Both DualQRTE and p-VQD accurately track the observables using only 200 shots per circuit. With the same resources, VarQRTE, on the other hand, has lower accuracy and we need to use 1024 shots per circuit to match the result of the optimization-based algorithms.

B. Error bounds

In variational real-time evolution, the model error in terms of Bures distance D_B due to restriction to the variational
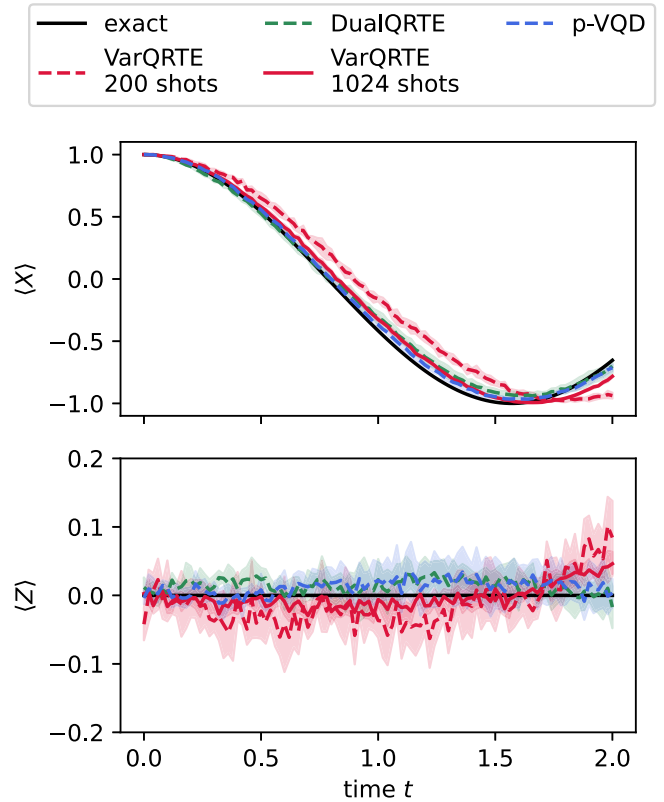

FIG. 5. Average magnetization in X and Z directions as tracked by different variational algorithms.

manifold can be expressed as [49]

$$\dot{\varepsilon}_M := \left\| \sum_{k=1}^{d} \dot{\theta}_k |\partial_k \phi(\boldsymbol{\theta})\rangle + iH |\phi(\boldsymbol{\theta})\rangle \right\|_2^2 \qquad (19)$$
$$= \text{Var}(H|\phi(\boldsymbol{\theta})) + \dot{\boldsymbol{\theta}}^T g(\boldsymbol{\theta})\dot{\boldsymbol{\theta}} - 2\dot{\boldsymbol{\theta}}^T \boldsymbol{b}(\boldsymbol{\theta}),$$

where we set $\hbar \equiv 1$. Integrating this error rate provides an upper bound on the Bures distance, that is

$$D_B(\phi(\boldsymbol{\theta}(T)), \Psi(T)) \leqslant \int_0^T \dot{\varepsilon}_M(t)dt,$$

where $|\Psi(t)\rangle$ is the exact time-evolved state and the time dependence of ε_M is due to the time-dependence of the parameters $\boldsymbol{\theta} = \boldsymbol{\theta}(t)$.

Up to the variance $\text{Var}(H|\phi(\boldsymbol{\theta})) = \langle\phi(\boldsymbol{\theta})|H^2|\phi(\boldsymbol{\theta})\rangle - (\langle\phi(\boldsymbol{\theta})|H|\phi(\boldsymbol{\theta})\rangle)^2$, this error is proportional to the loss function used in DualQRTE. By using the same expansion $\dot{\boldsymbol{\theta}} = \delta\boldsymbol{\theta}/\delta\tau$ and using the infidelity to approximate the inner product with respect to the geometric tensor, we can rewrite the error as

$$\dot{\varepsilon}_M = \text{Var}(H|\phi(\boldsymbol{\theta})) + \frac{1 - F(\boldsymbol{\theta}, \boldsymbol{\theta} + \delta\boldsymbol{\theta})}{(\delta\tau)^2} - \frac{2\delta\boldsymbol{\theta}^T \boldsymbol{b}(\boldsymbol{\theta})}{\delta\tau} + \mathcal{O}(\delta\tau)$$

$$= \text{Var}(H|\phi(\boldsymbol{\theta})) + \frac{2\mathcal{L}(\delta\boldsymbol{\theta})}{(\delta\tau)^2} + \mathcal{O}(\delta\tau).$$

Note that the error scales linearly in time perturbation $\delta\tau$ as the infidelity approximation has a cubic error term [32], which is divided by the square of the perturbation. If we, for exam-
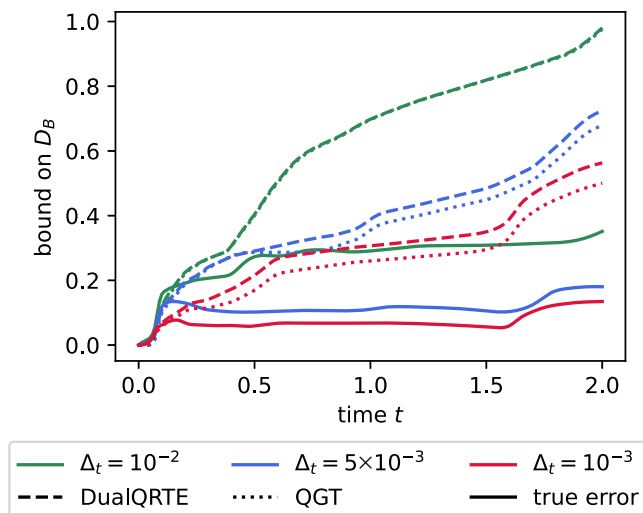
FIG. 6. Error of the real-time evolution in Bures distance, plus error bounds obtained with VarQRTE and DualQRTE.

ple, use a forward Euler rule with time step Δ_t to integrate the variational error, the integration error scales as $\mathcal{O}(\Delta_t + T\delta\tau)$. This highlights the importance of differentiating between the time step Δ_t for the integration and the time-perturbation $\delta\tau$ to approximate the derivative.

In Fig. 6, we show the error bounds along with the true error for the time evolution of the Heisenberg model. The bounds are computed for different time steps Δ_t for VarQRTE, and for DualQRTE for a fixed time perturbations $\delta\tau = 10^{-3}$ in exact simulations. Firstly, we can verify that the error bounds hold. Secondly, the larger the time step relative to the time-perturbation, the more accurate the approximation of the dual time evolution, as the error $\mathcal{O}(\Delta_t + T\delta\tau)$ is dominated by the integration error.

V. CONCLUSION

In this paper, we present a novel algorithm for variational quantum time evolution that does not require the evaluation of the QGT, but instead solves a dual optimization problem in each time step. The proposed dual time evolution algorithm, DualQTE, is particularly interesting for imaginary-time evolution, as there is currently no alternative variational algorithm able to circumvent the $\mathcal{O}(d^2)$ cost of VarQITE. For real-time evolution, p-VQD also offers an optimization-based approach by projecting a single Trotter step onto the variational form. In comparison, the dual time evolution has the advantage that no Suzuki-Trotter step has to be implemented, which could require deep circuits or nonlocal operations, depending on the Hamiltonian. Furthermore, our algorithm allows to evaluate variational error bounds [49], although how accurately they can be evaluated in the presence of shot noise remains an open question.

We demonstrated DualQTE for the imaginary-time evolution of a Heisenberg Hamiltonian on 12 qubits, and found that, in this setting, it requires about one order of magnitude less measurements to achieve the same accuracy as VarQITE. As a practical application of imaginary-time evolution, we calculated thermodynamic observables with the QMETTS algorithm and showed that the DualQITE is suitable to re-

produce the sampling distributions. Finally, we applied our algorithm to an illustrative example for real-time evolution, where it produced comparable results to p-VQD for the same amount of resources, while both algorithms outperformed VarQRTE.

In the presented experiments, we used standard gradient descent algorithms with a fixed learning rate. We expect that the performance could be further improved by using more advanced optimization schemes, or methods that also take into account information from previous iterations. Another possible improvement would be a suitable termination criterion for noisy evaluations of the loss function. As for other optimization-based time evolution algorithms, such as p-VQD, it remains challenging to accurately measure the fidelity in the presence of hardware noise.

In conclusion, the proposed DualQTE is an efficient variational algorithm for quantum time evolution that does not suffer from the quadratic complexity of evaluating the QGT. This cost reduction enables scaling imaginary-time evolution to larger, practically relevant system sizes and allows the simulation and demonstration of a wide variety of important tasks such as Gibbs state preparation, mixed time evolution, or the evaluation of thermodynamic observables. Improving the resource requirements for near-term algorithms is an important step for scaling demonstrations to the full size of today's quantum computers and work towards practical applications.

APPENDIX A: RUNTIME ESTIMATES OF VARIATIONAL TIME EVOLUTION

The benchmark in Sec. III C provides a scaling for the total number of measurements N required by VarQITE and DualQITE, which allows a runtime estimation on the algorithms on quantum hardware. In this estimation we neglect the overhead of classical processors and assume a superconducting quantum computer with a basis gate set including $\sqrt{X}$, R_Z and CX gates, as reported by several IBM Quantum backends, for example. This gate set allows to compile any sequence of single qubit gates into two $\sqrt{X}$ gates and three virtual R_Z gates. For an n-qubit simulation of the Heisenberg model and the considered circuit model (see Fig. 10) with r repetitions, the time for a single measurement can then be approximated as

$$t_{\text{shot}} = 2rt_{\text{CX}} + 2(r+1)t_{\sqrt{X}} + t_{\text{meas}} + t_{\text{reset}}, \quad \text{(A1)}$$

where t_{CX} is the duration of a CX gate, $t_{\sqrt{X}}$ the duration of a $\sqrt{X}$ gate, t_{meas} the time of a measurement and t_{reset} the time to reset the qubits for the next execution. Since the R_Z gates are virtual they do not contribute to the runtime. The total runtime is then estimated by Nt_{shot}.

Depending on the architecture and the gate decomposition the duration and fidelity of single- and two-qubit operation, as well as measurements, varies on superconducting qubit chips [54]. Here we use gate times of $t_{CX} = 451\ ns$, $t_{\sqrt{X}} = 36\ ns$ and $t_{\mathrm{meas}} = 860\ ns$ as reported by `ibm_peekskill` (v2.6.5), which is an IBM Quantum Falcon processors [55]. For shallow circuits in particular, the time to reset qubits for the following execution is a crucial bottleneck. The reset operation can, for example, be implemented by waiting 5-10 T1 times and let the qubits decay to the computational ground-state, but with T1 times of the order of 400 μs the reset via relaxation is orders of magnitude slower than the other circuit operations. Active resets instead measure the qubit state and apply an X operation conditionally if the state $|1\rangle$ is measured. This technique allows to reduce the reset times to typically 50 to 250 μs on IBM hardware [56], which, however, still dominates the overall runtime for the considered circuits. By

using a second excited state it is possible to implement reset schemes with 500 ns to 2 μs [57,58] and we therefore use $t_{\mathrm{reset}} = 2$ μs in our estimation.

APPENDIX B: RELATION OF THE INFIDELITY AND THE QUANTUM GEOMETRIC TENSOR

Based on Appendix A of Ref. [32], we explicitly derive Eq. (8). We consider the Taylor expansion of the fidelity, defined by

$$f(\delta\boldsymbol{\theta}) := F(\boldsymbol{\theta}, \boldsymbol{\theta} + \delta\boldsymbol{\theta}) = |\langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle|^2, \quad (B1)$$

where we introduce the function f for brevity. Up to second order, the Taylor expansion of f around $\delta\boldsymbol{\theta} = \mathbf{0}$ is given by

$$f(\delta\boldsymbol{\theta}) = f(\mathbf{0}) + \delta\boldsymbol{\theta}^T \nabla f(\mathbf{0}) + \frac{1}{2}\delta\boldsymbol{\theta}^T \nabla\nabla^T f(\mathbf{0})\delta\boldsymbol{\theta} + \mathcal{O}(\|\delta\boldsymbol{\theta}\|_2^3),$$
$$(B2)$$

where $\nabla f(\mathbf{0})$ first differentiates f and then evaluates the derivative at $\delta\boldsymbol{\theta} = \mathbf{0}$. Since $f(\mathbf{0}) = F(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1$ is the maximum of the fidelity, the second term in the expansion vanishes. The elements of the Hessian of f are

$$\partial_{\delta\theta_i}\partial_{\delta\theta_j}f(\mathbf{0}) = \partial_{\delta\theta_i}\partial_{\delta\theta_j}\langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle\langle\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\big|_{\delta\boldsymbol{\theta}=\mathbf{0}} = 2\partial_{\delta\theta_i}\mathrm{Re}\big(\langle\phi(\boldsymbol{\theta})|\partial_{\delta\theta_j}\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle\langle\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\big)\big|_{\delta\boldsymbol{\theta}=\mathbf{0}}$$

$$= 2\mathrm{Re}\big(\langle\phi(\boldsymbol{\theta})|\partial_{\delta\theta_i}\partial_{\delta\theta_j}\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle\langle\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle + \langle\phi(\boldsymbol{\theta})|\partial_{\delta\theta_j}\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle\langle\partial_{\delta\theta_i}\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\big)\big|_{\delta\boldsymbol{\theta}=\mathbf{0}}$$

$$= 2\mathrm{Re}\big(\langle\phi(\boldsymbol{\theta})|\partial_i\partial_j\phi(\boldsymbol{\theta})\rangle + \langle\phi(\boldsymbol{\theta})|\partial_j\phi(\boldsymbol{\theta})\rangle\langle\partial_i\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\big), \quad (B3)$$

where we explicitly state when we differentiate with respect to elements of $\delta\boldsymbol{\theta}$, using $\partial_{\delta\theta_i}$, or elements of $\boldsymbol{\theta}$, using ∂_i. The first term of Eq. (B3) can be further rewritten by considering the second derivative of $1 = \langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle$, which is

$$0 = \partial_i\partial_j\langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle = 2\partial_i\mathrm{Re}\big(\langle\partial_j\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\big) = 2\mathrm{Re}\big(\langle\partial_i\partial_j\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle + \langle\partial_j\phi(\boldsymbol{\theta})|\partial_i\phi(\boldsymbol{\theta})\rangle\big) \quad (B4)$$

and implies

$$\mathrm{Re}\big(\langle\partial_i\partial_j\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\big) = -\mathrm{Re}\big(\langle\partial_j\phi(\boldsymbol{\theta})|\partial_i\phi(\boldsymbol{\theta})\rangle\big). \quad (B5)$$

Plugging this into the expression for the Hessian of f, we obtain

$$\partial_i\partial_j f(\mathbf{0}) = 2\mathrm{Re}\big(-\langle\partial_i\phi(\boldsymbol{\theta})|\partial_j\phi(\boldsymbol{\theta})\rangle + \langle\phi(\boldsymbol{\theta})|\partial_j\phi(\boldsymbol{\theta})\rangle\langle\partial_i\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta})\rangle\big) = -2g_{ij}(\boldsymbol{\theta}), \quad (B6)$$

for the real part of the QGT, $g(\boldsymbol{\theta})$. Hence, the Taylor expansion of the fidelity is

$$f(\delta\boldsymbol{\theta}) = 1 - \delta\boldsymbol{\theta}^T g(\boldsymbol{\theta})\delta\boldsymbol{\theta} + \mathcal{O}(\|\delta\boldsymbol{\theta}\|_2^3), \quad (B7)$$

which, after rearranging, yields Eq. (8).

APPENDIX C: DERIVATION VIA QUANTUM NATURAL GRADIENT DESCENT

VarQITE is inherently connected to the quantum natural gradient (QNG) algorithm [32]. In fact, this connection is a motivation for the convergence of the QNG as imaginary-time evolution is guaranteed to converge to the ground state, if there is sufficient initial overlap with it.

With a forward Euler integration the VarQITE update rule is

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \Delta_t g^{-1}(\boldsymbol{\theta}^{(t)})\left(-\frac{\nabla E(\boldsymbol{\theta}^{(t)})}{2}\right).$$

This coincides with the QNG update step for the loss function $\ell(\boldsymbol{\theta}) = E(\boldsymbol{\theta})/2$ and a learning rate of $\eta = \Delta_t$,

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta g^{-1}(\boldsymbol{\theta}^{(t)})\nabla\ell(\boldsymbol{\theta}^{(t)}). \quad (C1)$$

The natural gradients step can be expressed in a dual formulation as

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}\ \langle\nabla\ell(\boldsymbol{\theta}^{(t)}), \boldsymbol{\theta} - \boldsymbol{\theta}^{(t)}\rangle + \frac{1}{2\eta}d^2(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)}),$$

with a distance metric d. In this equation, we see that the update step is going into the opposite direction of the gradient $\nabla\ell$, while the magnitude is limited by the distance metric and the learning rate.

Standard gradient descent uses the model-agnostic ℓ_2 norm as distance metric. Natural gradients on the other hand limit
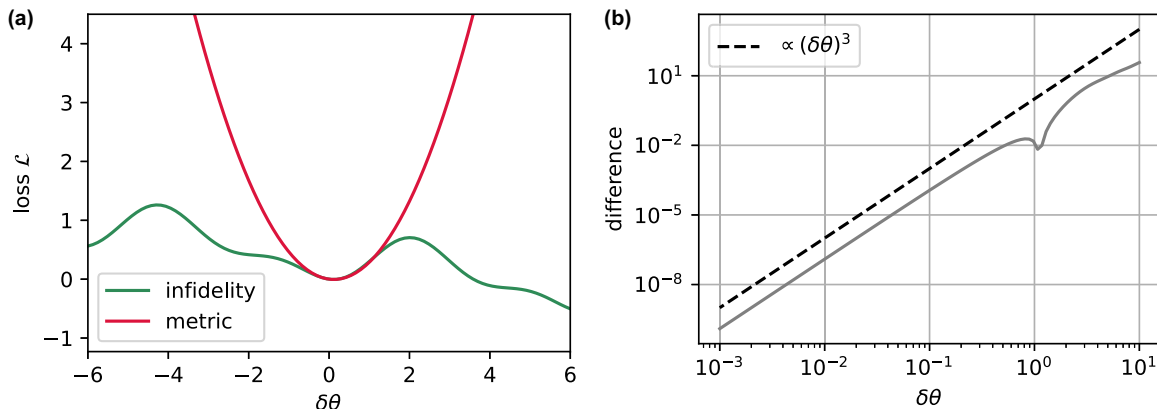
FIG. 7. (a) Values of the loss function $\mathcal{L}$ for evaluation with the QGT metric, and with introduced infidelity approximation. (b) Difference of the QGT metric and infidelity as function of the perturbation $\delta\theta$.

the update step by the amount of change it induces in the model. To measure the induced change the metric d is chosen to be the Fubini-Study metric, which, as shown in Ref. [32], if locally approximated, yields the QGT:

$$d^2(\phi(\boldsymbol{\theta}), \phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})) = \arccos^2 |\langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle|$$
$$= 1 - |\langle\phi(\boldsymbol{\theta})|\phi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle|^2 + \mathcal{O}(||\delta\boldsymbol{\theta}||_2^4)$$
$$= \langle\delta\boldsymbol{\theta}, g(\boldsymbol{\theta})\delta\boldsymbol{\theta}\rangle + \mathcal{O}(||\delta\boldsymbol{\theta}||_2^3).$$

The formulation in Eq. (C1) is then obtained by solving the minimization problem.

To circumvent the explicit evaluation of the QGT the natural gradient update can instead be calculated without the quadratic local approximation, and instead solve the optimization problem directly. If we use the infidelity as distance metric and replace the loss function gradient by the evolution gradient $\nabla\ell(\boldsymbol{\theta}) = \nabla E(\boldsymbol{\theta})/2 = -\boldsymbol{b}(\boldsymbol{\theta})$, we obtain the same update rule as the main text

$$\boldsymbol{\theta}^{(t+1)} = \underset{\boldsymbol{\theta}}{\arg\min} - \langle\boldsymbol{b}(\boldsymbol{\theta}^{(t)}), \boldsymbol{\theta} - \boldsymbol{\theta}^{(t)}\rangle + \frac{1 - F(\boldsymbol{\theta}, \boldsymbol{\theta}^{(t)})}{2\Delta_t}.$$

APPENDIX D: ILLUSTRATIVE EXAMPLE

For an intuitive understanding of the approximations of the QGT norm, we investigate an illustrative example with the variational model $|\phi(\theta)\rangle = R_Z(\theta)R_Y(\theta)|0\rangle$, the Hamiltonian $H = Z$ and a time step of $\delta\tau = 1/2$. In Fig. 7(a), we compare the exact values of the loss function $\mathcal{L}$ for imaginary-time evolution around $\theta = \pi/4$ obtained by using the metric $\langle\delta\theta, g(\theta)\delta\theta\rangle$ or the infidelity $1 - F(\theta, \theta + \delta\theta)$ as norm.

At $\delta\theta = 0$ the approximation is exact and in the vicinity the difference scales as $(\delta\theta)^3$, see also Fig. 7(b). Note, that the infidelity is periodic and bounded in $[0, 1]$ but the linear term $b^T\delta\theta$ is unbounded, which leads to the fact that the minimum of the infidelity-based loss function close to $\delta\theta = 0$ is not the global minimum. This is well visible in Fig. 7(a) where the infidelity-based loss function achieves lower values for large $\delta\theta$ than the minimum of the QGT close to $\delta\theta = 0$. Since we aim to find the same minimum as the QGT-based loss function using a local optimization routine, such as gradient descent, is crucial for the dual time evolution.

1. Impact of the time perturbation

The approximation error scales with the norm of $\delta\boldsymbol{\theta}$ and, therefore, solving for the update step $\dot{\boldsymbol{\theta}} = \delta\boldsymbol{\theta}/\delta\tau$ with a smaller time perturbation $\delta\tau$ should result in a smaller error in the update step. Remembering the definition of the loss function

$$\mathcal{L}(\delta\boldsymbol{\theta}) = \frac{1 - F(\boldsymbol{\theta}, \boldsymbol{\theta} + \delta\boldsymbol{\theta})}{2} - \delta\tau \cdot \boldsymbol{b}^T(\boldsymbol{\theta})\delta\boldsymbol{\theta},$$

we see that a smaller $\delta\tau$ moves the minimum closer to the minimum of the infidelity at $\delta\boldsymbol{\theta} = \boldsymbol{0}$, leading to a smaller approximation error.

Since the fidelity is bounded but the linear part $\boldsymbol{b}^T(\boldsymbol{\theta})\delta\boldsymbol{\theta}$ is not, there is a maximum feasible range for the value of $\delta\tau$. A necessary condition for the existence of the minimum is that the gradient of the loss function vanishes, $\nabla\mathcal{L} = \boldsymbol{0}$, which requires

$$\forall i \in \{1, \dots, d\} : \frac{1}{2}\frac{\partial}{\partial(\delta\theta)_i}F(\boldsymbol{\theta}, \boldsymbol{\theta} + \delta\boldsymbol{\theta}) = -\delta\tau \cdot b_i(\boldsymbol{\theta}).$$
(D1)

For a circuit with unique parameters and only Pauli rotations gates, the gradient of the fidelity can be bounded via the parameter-shift rule to be in $[-1/2, 1/2]$ (see also Appendix E). Thus a necessary condition for the time step perturbation is

$$\forall i \in \{1, \dots, d\} : \delta\tau \in \left[\frac{-1}{4|b_i(\boldsymbol{\theta})|}, \frac{1}{4|b_i(\boldsymbol{\theta})|}\right],$$
(D2)

which can be generalized to circuits with repeated parameters or other than Pauli gates. Note that this is only a necessary and not a sufficient condition for the existence of a minimum since, depending on the circuit structure, the fidelity gradient may not support the full range $[-1/2, 1/2]$.

In Fig. 8(a), we visualize the impact of $\delta\tau$ on the loss landscape. For small time perturbations the QGT-based and dual loss landscapes almost coincide, but if $\delta\tau$ is chosen too large the dual loss function has no minimum. If the loss function can be evaluated exactly, choosing $\delta\tau$ as small as possible therefore minimizes the approximation error. In Fig. 8(b), we find the the error in the parameter derivative $\dot{\boldsymbol{\theta}} = \delta\boldsymbol{\theta}/\delta\tau$ scales approximately as $\mathcal{O}(\delta\tau)$. In practice, however, the loss
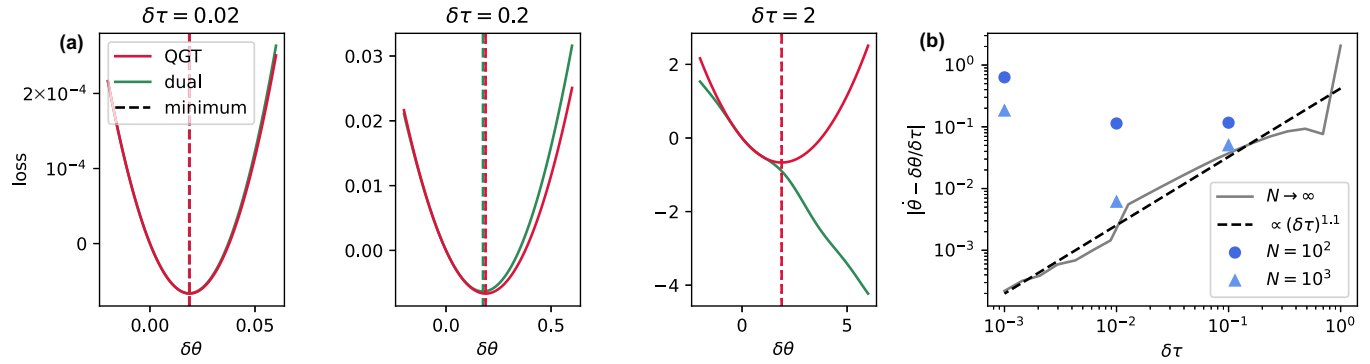
FIG. 8. (a) Loss landscapes and optimal solutions of the original, QGT-based loss function and the dual loss function for different $\delta\tau$. (b) Error in calculating the parameter derivative $\dot{\theta}$ depending on $\delta\tau$ and the number of measurements N.

function is subject to measurement noise and errors in the solution $\delta\boldsymbol{\theta}$ are amplified by $1/\delta\tau$. Hence, for a finite number of measurements there is a trade-off between QGT approximation error and controlling the noise amplification.

APPENDIX E: BOUND THE SAMPLE COMPLEXITY OF VarQTE

In this section, we present the derivation on the upper bound of the sample complexity of VarQTE and DualQTE. The target error is measured in integrated Bures distance,

$$\varepsilon_S = \frac{1}{T}\int_0^T \sqrt{2(1 - |\langle\phi(\boldsymbol{\theta})|\phi(\tilde{\boldsymbol{\theta}})\rangle|)}dt. \quad (E1)$$

Assuming a forward Euler integration, the Bures distance can be formulated in terms of the QGT as

$$\varepsilon_S = \frac{1}{T}\int_0^T \sqrt{2\left(1 - \sqrt{1 - \Delta_t^2\Delta\dot{\boldsymbol{\theta}}^T g(\boldsymbol{\theta})\Delta\dot{\boldsymbol{\theta}}}\right)}dt$$

$$= \frac{1}{T}\int_0^T \Delta_t\sqrt{\Delta\dot{\boldsymbol{\theta}}g(\boldsymbol{\theta})\Delta\dot{\boldsymbol{\theta}}}dt$$

$$\leqslant \frac{1}{T}\int_0^T \Delta_t\|g(\boldsymbol{\theta})\|_2\|\Delta\dot{\boldsymbol{\theta}}\|_2 dt$$

$$\leqslant \Delta_t\sqrt{\lambda_{\max}}\|\Delta\dot{\boldsymbol{\theta}}_{\max}\|_2, \quad (E2)$$

where we introduced $\Delta\dot{\boldsymbol{\theta}} = \tilde{\dot{\boldsymbol{\theta}}} - \dot{\boldsymbol{\theta}}$, $\lambda_{\max} \geqslant 0$ is a bound on the largest eigenvalue of g for any parameter value $\boldsymbol{\theta}$ and, similarly, $\|\Delta\dot{\boldsymbol{\theta}}_{\max}\|_2$ an upper bound on the norm $\Delta\dot{\boldsymbol{\theta}}$. In the first line, we dropped $\mathcal{O}(\Delta_t^3)$ error terms, in the second line we used a first order Taylor-expansion and in the last line we use the definition of the operator norm to bound the inner product of $\delta\boldsymbol{\theta}$ in the metric of g.

1. VarQTE

In each VarQTE step, we solve a linear system for the update step, where the measurements of the QGT and evolution gradient are subject to sampling error. We define the noisy quantities as $\tilde{g}(\boldsymbol{\theta}) = g(\boldsymbol{\theta}) + \Delta g(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{b}}(\boldsymbol{\theta}) = \boldsymbol{b}(\boldsymbol{\theta}) + \Delta\boldsymbol{b}(\boldsymbol{\theta})$

and, then, solve the noisy linear system

$$\tilde{g}(\boldsymbol{\theta})\tilde{\dot{\boldsymbol{\theta}}} = \tilde{\boldsymbol{b}}(\boldsymbol{\theta}), \quad (E3)$$

with the noisy update $\tilde{\dot{\boldsymbol{\theta}}} = \dot{\boldsymbol{\theta}} + \Delta\dot{\boldsymbol{\theta}}$. To stabilize the linear system and ensure the QGT and it's estimate are invertible, we assume a regularization of g and $\tilde{g}$ in form of a diagonal shift δ_c. This shift is a trade-off of stability and bias, which is also discussed in Ref. [39], Appendix D.

We can write the error in the update step using the difference of the noisy and exact linear system solutions, as

$$\|\Delta\dot{\boldsymbol{\theta}}\|_2 = \|(g + \Delta g)^{-1}(\boldsymbol{b} + \Delta\boldsymbol{b}) - g^{-1}\boldsymbol{b}\|_2$$

$$\approx \|(g^{-1} - g^{-1}\Delta g g^{-1})(\boldsymbol{b} + \Delta\boldsymbol{b}) - g^{-1}\boldsymbol{b}\|_2$$

$$= \|g^{-1}\Delta\boldsymbol{b} - g^{-1}\Delta g g^{-1}\boldsymbol{b} - g^{-1}\Delta g^{-1}g^{-1}\Delta\boldsymbol{b}\|_2$$

$$\approx \|g^{-1}\Delta\boldsymbol{b} - g^{-1}\Delta g\dot{\boldsymbol{\theta}}\|_2$$

$$\leqslant \|g^{-1}\|_2(\|\Delta\boldsymbol{b}\|_2 + \|\Delta g\|_2\|\dot{\boldsymbol{\theta}}\|_2), \quad (E4)$$

where we dropped the explicit parameter dependence for legibility. In the second line we used the Neumann series to approximate $(g + \Delta g)^{-1} = g^{-1} - g^{-1}\Delta g g^{-1} + \mathcal{O}(\|\Delta g\|_2^2\|g^{-1}\|_2^3)$ and dropped quadratic error terms on the fourth line. In the following, we derive upper bounds on the maximal value of the individual contributions in the error bound, such that we finally obtain a bound $\|\Delta\dot{\boldsymbol{\theta}}_{\max}\|$ on the error in the update step.

Spectrum of g. Each QGT entry can be computed as [39]

$$g_{ij}(\boldsymbol{\theta}) = -\frac{1}{2}\partial_i\partial_j F(\boldsymbol{\theta}', \boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}'=\boldsymbol{\theta}}. \quad (E5)$$

For a circuit with unique, noninteracting parameter and only plain Pauli rotation gates $R_P(\theta)$, we can use the parameter-shift rule [33] to write the entry as

$$g_{ij}(\boldsymbol{\theta}) = -\frac{1}{2}\frac{F_{ij}^{(++)} - F_{ij}^{(+-)} - F_{ij}^{(-+)} + F_{ij}^{(--)}}{4}, \quad (E6)$$

where we abbreviated $F_{ij}^{(\pm\pm)} = F(\boldsymbol{\theta}, \boldsymbol{\theta} \pm \boldsymbol{e}_i\pi/2 \pm \boldsymbol{e}_j\pi/2)$ for the ith unit vector e_i (and jth unit vector e_j). Since the fidelity is in $[0, 1]$ we can bound each entry by

$$-\frac{1}{4} \leqslant g_{ij}(\boldsymbol{\theta}) \leqslant \frac{1}{4}, \quad (E7)$$

for any value of $\boldsymbol{\theta}$. Gershgorin's circle theorem tells us that the maximal eigenvalue of g is bounded from above by the maximal sum over the columns or rows, which in this case is achieved by setting all elements of a column/row to 1/4. This gives the bound

$$\lambda_{\max} \leqslant \sum_{i=1}^{d} \frac{1}{4} = \frac{d}{4}. \tag{E8}$$

This bound can be generalized to circuits with coefficients or repeated parameters by applying the chain and product rules. For example, for a coefficient-free circuit where parameters can be repeated up to m times, the bound becomes $md/4$.

Norm of the update step. The update step can be bounded as $\|\dot{\boldsymbol{\theta}}\|_2 \leqslant \|g^{-1}\|_2 \|\boldsymbol{b}\|_2$, where $\|g^{-1}\|_2 \leqslant \delta_c^{-1}$. The evolution gradient can be bounded using the parameter-shift rule, under the circuit structure assumptions as the previous section. Each element in the gradient is bounded by

$$|b_i| = \frac{|E_i^{(+)} - E_i^{(-)}|}{2} \leqslant \frac{|E_i^{(+)}| + |E_i^{(-)}|}{2} \leqslant E_{\max}, \tag{E9}$$

where $E_i^{(\pm)} = E(\boldsymbol{\theta} \pm \boldsymbol{e}_i \pi/2)$ and $E_{\max}$ is the absolute maximum system energy. The norm over all elements is then $\|\boldsymbol{b}\|_2 \leqslant \sqrt{d} E_{\max}$, leading to an overall bound of

$$\|\dot{\boldsymbol{\theta}}\|_2 \leqslant \frac{\sqrt{d} E_{\max}}{\delta_c}, \tag{E10}$$

for any parameter value $\boldsymbol{\theta}$.

Sampling errors. Since the measurement noise is unbiased, the random variable $\Delta g = \tilde{g} - g$ has zero mean with i.i.d. entries. This allows to apply Latała's theorem [50], which states that

$$\mathbb{E}[\|\Delta g\|_2] \leqslant C \left(\max_i \sqrt{\sum_{j=1}^{d} \mathbb{E}[(\Delta g)_{ij}^2]} \right.$$
$$\left. + \max_j \sqrt{\sum_{i=1}^{d} \mathbb{E}[(\Delta g)_{ij}^2]} + \sqrt[4]{\sum_{i,j=1}^{d} \mathbb{E}[(\Delta g)_{ij}^4]} \right), \tag{E11}$$

for some constant $C \in \mathbb{R}$.

Reference [59] is concerned with the similar case of sampling the matrix $[F(x_i, x_j)]_{i,j=1}^{d}$ for a set of parameters $\{x_i\}_{i=1}^{d}$. There, the matrix entries are Bernoulli distributed with probability $F(x_i, x_j)$. Using QGT representation as Hessian and applying the parameter-shift rule, we can see that the entries of g_{ij} are Poisson binomial distributed [60] with probabilities $[F_{ij}^{(++)}, 1 - F_{ij}^{(+-)}, 1 - F_{ij}^{(-+)}, F_{ij}^{(--)}]$ over a shifted support $[0, 1, 2, 3, 4] \rightarrow [-2, -1, 0, 1, 2]$. Since the means of this distribution are independent of the number of circuit parameters, it can be shown analogous to Ref. [59] that

$$\mathbb{E}[|(\Delta g)_{ij}|^2] = \mathcal{O}\left(\frac{1}{N}\right) \text{ and } \mathbb{E}[|(\Delta g)_{ij}|^4] = \mathcal{O}\left(\frac{1}{N^2}\right), \tag{E12}$$

which leads to a total bound of

$$\mathbb{E}[\|\Delta g\|_2] = \mathcal{O}\left(\sqrt{\frac{d}{N}}\right). \tag{E13}$$

The bound on $\|\Delta \boldsymbol{b}\|_2$ does not need to be tighter than $\|\Delta g\|_2 \|\dot{\boldsymbol{\theta}}\|_2$, which is straightforward to achieve via the sampling error. Using the parameter-shift rule, we have

$$|\Delta b_i| = |\tilde{b}_i - b_i| = \frac{|\tilde{E}_i^{(+)} - \tilde{E}_i^{(-)} - E_i^{(+)} + E_i^{(-)}|}{2}$$
$$\leqslant \frac{|\tilde{E}_i^{(+)} - E_i^{(+)}| + |\tilde{E}_i^{(-)} - E_i^{(-)}|}{2}$$
$$= \mathcal{O}\left(\frac{\sqrt{\text{Var}(E_i^{(+)})} + \sqrt{\text{Var}(E_i^{(-)})}}{2\sqrt{N}}\right). \tag{E14}$$

The variance of any state $|\psi\rangle$ can be upper bounded by

$$\text{Var}(E) = \langle\psi|H^2|\psi\rangle - E^2 \leqslant \langle\psi|H^2|\psi\rangle \leqslant E_{\max}^2. \tag{E15}$$

Summing over all gradient elements we obtain

$$\|\Delta \boldsymbol{b}_{\max}\|_2 = \mathcal{O}\left(\frac{\sqrt{d} E_{\max}}{\sqrt{N}}\right). \tag{E16}$$

Final bound. Plugging the bounds in the previous paragraphs into Eq. (E4) and then into Eq. (E2), we obtain the final bound of

$$\varepsilon_S \leqslant \mathcal{O}\left(\frac{d^{3/2} E_{\max} \Delta_t}{\delta_c^2 \sqrt{N}}\right). \tag{E17}$$

The same asymptotic bound can be derived by performing a moment expansion on the expectation $\mathbb{E}[\dot{\boldsymbol{\theta}} - \tilde{\dot{\boldsymbol{\theta}}}]$.

As an example we investigate a simple product-state model, which allows to show the tightness of several of the above bounds. We look a the first time step of the n-qubit Hamiltonian $H = \sum_{i=1}^{n} Z_i$ with an ansatz that consists of a single layer of Pauli-Y rotations, each with an individual parameter. The initial state is $|+\rangle^{\otimes n}$ which is prepared by setting each of the parameters to $\pi/2$. Each expectation value is computed with $N = 1000$ measurements and we use a regularization of $\delta_c = 10^{-2}$. We then vary the number of qubits from $n = 2$ to 10 and measure the error term contributions over 10 averages, since Δg and $\Delta \boldsymbol{b}$ are random variables.

The QGT is measure of the correlation between the parameter derivatives and as there is no light-cone connecting any two parameterized gates in the product state ansatz, the QGT is a diagonal matrix. Its norm is therefore $\|g\|_2 = 1/4$ for any system size. With this restriction, we observe in Fig. 9(a) that the bound on the Bures metric in Eq. (E2) is tight as $\varepsilon_S \propto \|\Delta \dot{\boldsymbol{\theta}}\|_2$. While all bounds are obeyed, we observe that in particular the bound on $\|\dot{\boldsymbol{\theta}}\|_2$ is loose, since the bound in Eq. (E10) scales with $\sqrt{d} E_{\max} \propto d^{1.5}$, but we only observe a $d^{0.5}$ scaling. This bound could potentially be further improved by taking into account that the magnitude of the update step is bounded by the change induced of the evolution operator $\exp(-\Delta_t H)$, which is independent of the number of parameters d.
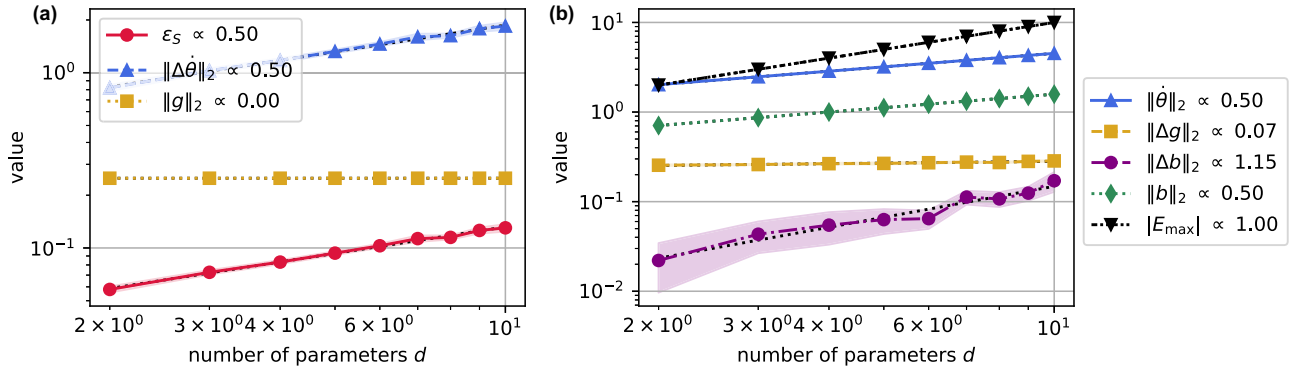
FIG. 9. Scaling of different error bound contributions for a product state setting. The labels include the scaling with number of parameters, i.e., $\propto \alpha$ indicates a scaling with d^α, where d is the number of parameters.

2. DualQTE

Assume we require K steps to converge. Then the error in the update step $\delta\boldsymbol{\theta}$ is

$$\|\Delta(\delta\boldsymbol{\theta})\|_2 = \|\Delta(\delta\boldsymbol{\theta}^{(K)})\|_2 = \|\widetilde{\delta\boldsymbol{\theta}}^{(K-1)} - \eta\widetilde{\nabla\mathcal{L}}(\widetilde{\delta\boldsymbol{\theta}}^{(K-1)}) - \delta\boldsymbol{\theta}^{(K-1)} + \eta\nabla\mathcal{L}(\delta\boldsymbol{\theta}^{(K-1)})\|_2$$
$$\leqslant \|\Delta(\delta\boldsymbol{\theta}^{(K-1)})\|_2 + \eta\|\widetilde{\nabla\mathcal{L}}(\widetilde{\delta\boldsymbol{\theta}}^{(K-1)}) - \nabla\mathcal{L}(\delta\boldsymbol{\theta}^{(K-1)})\|_2 \qquad (E18)$$
$$\leqslant \|\Delta(\delta\boldsymbol{\theta}^{(K-1)})\|_2 + \eta\|\Delta(\nabla\mathcal{L})_{\max}\|_2 \leqslant \eta K\|\Delta(\nabla\mathcal{L})_{\max}\|_2,$$

where we used that the error at the initial point is zero, $\|\Delta(\delta\boldsymbol{\theta}^{(0)})\|_2 = 0$. The error in the loss function gradient can then be written as

$$\|\Delta(\nabla\mathcal{L}(\delta\boldsymbol{\theta}))\|_2 = \left\|\frac{\Delta(\nabla F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}))}{2} + \delta\tau\Delta\boldsymbol{b}(\boldsymbol{\theta})\right\|_2$$
$$\leqslant \frac{\|\Delta(\nabla F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}))\|}{2} + \delta\tau\|\Delta\boldsymbol{b}(\boldsymbol{\theta})\|_2$$
$$\leqslant \frac{\|\Delta(\nabla F)_{\max}\|}{2} + \delta\tau\|\Delta\boldsymbol{b}_{\max}\|_2, \qquad (E19)$$

where $\Delta(\nabla F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta})) = \widetilde{\nabla F}(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta}) - \nabla F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta})$ and $\|\Delta(\nabla F)_{\max}\|_2$ is an upper bound on the maximum fidelity gradient error for any parameter.

The error in the gradient of F can be derived via the parameter-shift rule, as

$$|\Delta\partial_i F| = \frac{\Delta F_i^{(+)} - \Delta F_i^{(-)}}{2}$$
$$= \mathcal{O}\left(\sqrt{\frac{\text{Var}(F_i^{(+)})}{N}} + \sqrt{\frac{\text{Var}(F_i^{(-)})}{N}}\right) \qquad (E20)$$
$$= \mathcal{O}\left(\frac{1}{\sqrt{N}}\right),$$

where we used that the variance of the fidelity can be bounded for any state $|\psi\rangle$ as

$$\text{Var}(F) = \langle\psi|P_0^2|\psi\rangle - \langle\psi|P_0|\psi\rangle^2$$
$$= \langle\psi|P_0|\psi\rangle - \langle\psi|P_0|\psi\rangle^2 = F(1-F) \leqslant \tfrac{1}{4}. \quad (E21)$$

Hence the total error of the fidelity gradient in ℓ_2 norm is

$$\|\Delta(\nabla F)_{\max}\|_2 = \mathcal{O}\left(\sqrt{\frac{d}{N}}\right). \qquad (E22)$$

The bound on $\|\Delta\boldsymbol{b}_{\max}\|_2$ is already derived in the previous subsection, which gives then a total of

$$\|\Delta(\delta\boldsymbol{\theta})\|_2 = \mathcal{O}\left(\frac{\sqrt{d}K(1+\delta\tau E_{\max})}{\sqrt{N}}\right). \qquad (E23)$$

Using the definition $\dot{\boldsymbol{\theta}} = \delta\boldsymbol{\theta}/\delta\tau$, we then obtain

$$\varepsilon_S \leqslant \Delta_t\sqrt{\lambda_{\max}}\frac{\|\Delta(\delta\boldsymbol{\theta})\|_2}{\delta\tau} = \mathcal{O}\left(\sqrt{\frac{\lambda_{\max}d}{N}}\frac{\Delta_t K(1+\delta\tau E_{\max})}{\delta\tau}\right). \qquad (E24)$$
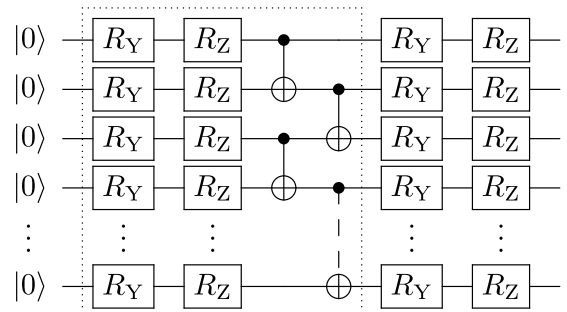

FIG. 10. The hardware efficient ansatz for the imaginary-time evolution experiments. In the QMETTS experiments the Pauli-Y rotations is replaced by Pauli-X rotations, if the evolution starts in the Y basis states $|\pm i\rangle$.
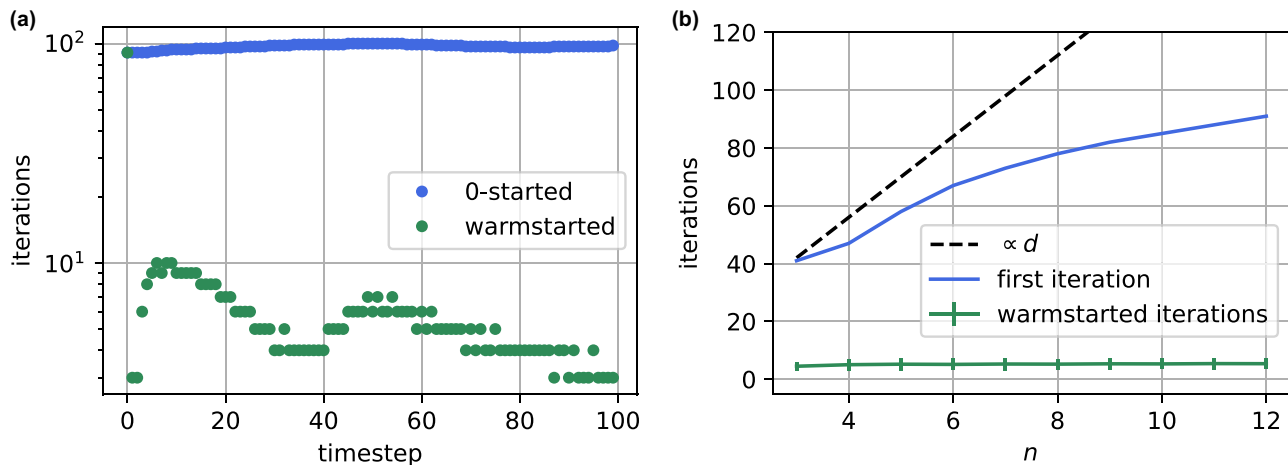
FIG. 11. (a) The number of iterations required per time step until convergence is reached with different initialization techniques. (b) The number of iterations for the first and warm-started iterations for different numbers of qubits. The warm-started points show mean and standard deviation of the number of iteration of all steps after the first.

APPENDIX F: IMAGINARY-TIME EVOLUTION OF THE HEISENBERG MODEL

1. Circuit diagram

The circuit used as variational model is schematically presented in Fig. 10. Each Pauli rotation gate has an independent parameter and the dotted box is repeated several times. For r repetitions and n qubits, the total number of tunable parameters is thus $2n(r + 1)$. The CNOT entangling gates are arranged in a pairwise manner to minimize the total depth to 2 per entangling layer.

2. Termination and warm starting

Termination criteria for gradient descent algorithms are typically defined as achieving a minimal threshold in the difference in the loss function between update steps or in the gradient norm. However, if only noisy readout of the loss function is available these criteria become unreliable as the noise in the evaluation might prevent the termination criterion to be fulfilled even though the algorithm converged.

One possible resolution would be to consider a moving average over a past batch of iterations. However, depending on the level of noise, this could require a large batchsize and therefore many iterations until the termination can be checked. Since the dual time evolution only has to compute small corrections, if small time steps are performed and the optimization are warm-started, we only expect a few iterations and a moving average is not a resource-efficient solution. Therefore we use a heuristic where the first optimization uses a large number of steps and the subsequent ones perform a fixed number of few iterations.

To demonstrate the effectiveness of warm starting and to calibrate the number of required steps for noisy evaluations we investigate the dual time evolution in an ideal setting with exact statevector simulations and no finite-sampling statistics. First, we perform the time evolution for a Heisenberg Hamiltonian with periodic boundary conditions with $n = 12$ sites, $J = 1/4$, $g = -1$ and the initial state $|+\rangle^{\otimes n}$. As circuit

model we use the hardware efficient circuit from Fig. 10 with $r = 6$ repetitions and optimize the update step with a gradient descent routine with a fixed learning rate of $\eta = 0.1$. In each time step we iterate until the change in loss function $\Delta\mathcal{L}$ is below the threshold of $10^{-4}\Delta_t = 10^{-6}$. The results are presented in Fig. 11(a), and we observe that warm starting drastically reduces the number of required iterations until the convergence criterion is reached.

In a second experiment we analyze how the required number of optimization steps scales with the system size. In Fig. 11(b), we repeat the above experiment for $n = 3$ to 12 spins and track the number of steps in the first iteration and the mean and standard deviation of the warm-started iterations. We see that the number of steps scales sublinearly in the number of parameters d and is almost constant for the warm-started iterations.

3. Resource requirements for the dual time evolution

This section shows the detailed VarQITE and DualQITE settings for the resource estimations in Sec. III. For VarQITE,

TABLE I. Detailed settings for the resource comparison of Var-QITE and DualQITE at fixed number of qubits $n = 12$: the achieved Bures distance D_B, the number of shots per circuit and the total number of measurements N. The dual method additionally shows the number of iterations K_0 in the first optimization and $K_{>0}$ in the subsequent, warm-started optimizations. Each optimization used gradient descent with a learning rate of $\eta = 0.1$.

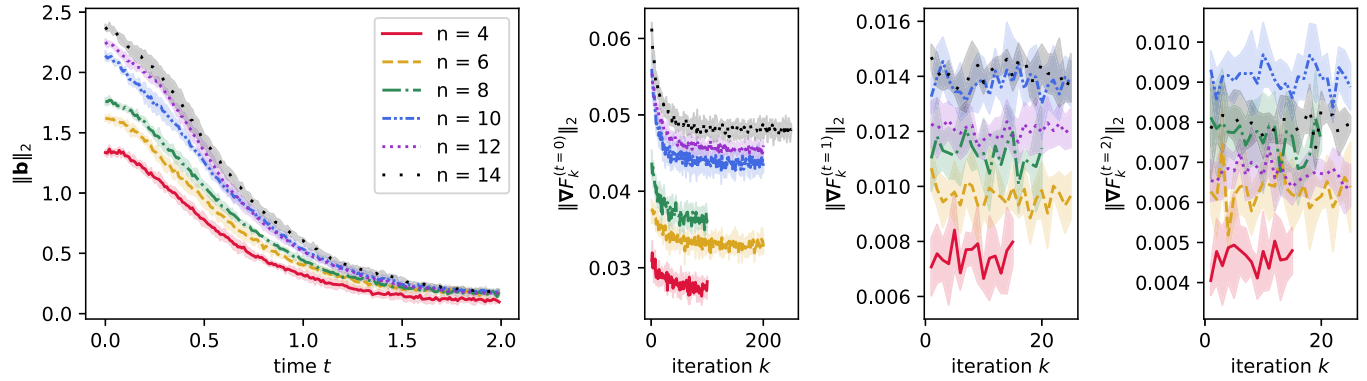| (a) Settings for VarQITE			(b) Settings for DualQITE				
I_B	shots	N	I_B	shots	K_0	$K_{>0}$	N
1.601	100	$\approx 10^8$	0.937	100	100	10	$\approx 2.5 \times 10^7$
0.558	1024	$\approx 10^9$	0.735	100	100	20	$\approx 5 \times 10^7$
0.149	8192	$\approx 8 \times 10^9$	0.305	1024	100	10	$\approx 2.5 \times 10^8$
			0.236	1024	200	20	$\approx 5 \times 10^8$
			0.153	2048	250	25	$\approx 10^9$

FIG. 12. The ℓ_2 norms for increasing number of qubits n of the imaginary evolution gradient, $\|\boldsymbol{b}\|_2$, during the entire evolution and the norm of the fidelity gradients $\|\nabla F_k^{(t)}\|_2$ at selected times, where t indicated the time step and k the iteration in the optimization within the time step. Lengths for the fidelity gradients differ as the optimization were performed using a different number of steps, see Table II.

we only varied the number of shots and in the dual method we additionally allowed to vary the number of iterations in the optimization in each time step. Especially DualQITE has a lot of additional degrees of freedom that could be optimized, such as the kind of optimizer, in addition to settings shared with VarQITE, such as time step size.

Table I shows the settings for VarQITE and DualQITE for the resource estimation in Fig. 2(b) and Table II the settings for the scaling with system size in Fig. 3.

4. Gradient benchmark

In this section, we measure how the norm of the loss function gradient, defined as

$$\nabla_{\delta\boldsymbol{\theta}}\mathcal{L}(\delta\boldsymbol{\theta}) = -\frac{\nabla_{\delta\boldsymbol{\theta}}F(\boldsymbol{\theta}, \boldsymbol{\theta}+\delta\boldsymbol{\theta})}{2} - \delta\tau \cdot \boldsymbol{b}(\boldsymbol{\theta})$$

scales with the number of qubits n in the Heisenberg Hamiltonian. This Hamiltonian is 2-local and since the ansatz depth grows logarithmic with the number of qubits we do not expect exponentially vanishing gradients for the evolution gradient b [41]. Further, as discussed in the main text, the initial point of each time step $\delta\boldsymbol{\theta}$ is close to **0**, which ensures the initial circuit is close to the identity and we do not expect to encounter a barren plateau [47].

Since both parts of the loss functions are not in a barren plateau setting, we expect the loss function to be measurable efficiently. In Fig. 12, we measure the ℓ_2 norm of both the evolution gradient and fidelity gradients and we find that neither gradient vanishes exponentially. Instead, the evolution gradient increases with system size, which reflects the extensiveness of energy in the Heisenberg model. Since we perform imaginary time evolution the system converges towards the ground state and we expect the energy gradients to vanish, once converged. In this case, we do not infer a barren plateau as the gradient norm does not systematically decrease faster for larger systems. Similarly, the fidelity gradients are expected to decay as the optimal parameter update $\delta\boldsymbol{\theta}$ is found. Note that for system sizes of 12 qubits the exponential decay of the gradients is typically clearly visible [42].

APPENDIX G: REAL-TIME EVOLUTION OF THE HEISENBERG MODEL

1. Circuit diagram

In the real-time evolution of the Heisenberg model we use the circuit sketched in Fig. 13, as used in Ref. [19].

TABLE II. Algorithm settings for the size scaling experiments of VarQITE and DualQITE.

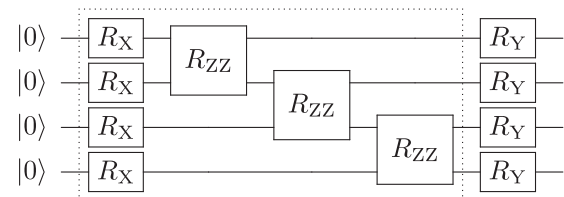| (a) Settings for VarQITE			(b) Settings for DualQITE					
n	shots	N	n	shots	K_0	$K_{>0}$	η	N
4	500	4.2×10^7	4	500	100	15	0.07	8.8×10^7
6	1500	4.2×10^8	6	600	200	25	0.07	3.3×10^8
8	2500	1.2×10^9	8	1000	100	20	0.1	6×10^8
10	6000	6.7×10^9	10	1500	200	25	0.12	1.7×10^9
12	8000	1.3×10^{10}	12	2500	200	25	0.1	3.5×10^9
			14	3000	250	25	0.12	4.9×10^9

FIG. 13. The real time evolution circuit. The dotted box is repeated three times and the rotation layer alternates between Pauli-X rotations (starting from the first layer) and Pauli-Y rotations.

[1] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe, Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator, Nature (London) **551**, 601 (2017).

[2] J. Dborin, V. Wimalaweera, F. Barratt, E. Ostby, T. E. O'Brien, and A. G. Green, Simulating groundstate and dynamical quantum phase transitions on a superconducting quantum computer, Nat. Commun. **13**, 5977 (2022).

[3] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin, Quantum phases of matter on a 256-atom programmable quantum simulator, Nature (London) **595**, 227 (2021).

[4] E. Altman, Many-body localization and quantum thermalization, Nat. Phys. **14**, 979 (2018).

[5] W. A. de Jong, K. Lee, J. Mulligan, M. Płoskoń, F. Ringer, and X. Yao, Quantum simulation of nonequilibrium dynamics and thermalization in the schwinger model, Phys. Rev. D **106**, 054508 (2022).

[6] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, npj Quantum Inf. **5**, 75 (2019).

[7] T. Jones, S. Endo, S. McArdle, X. Yuan, and S. C. Benjamin, Variational quantum algorithms for discovering hamiltonian spectra, Phys. Rev. A **99**, 062304 (2019).

[8] M. Motta, C. Sun, A. T. K. Tan, M. J. O. Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandao, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, Nat. Phys. **16**, 205 (2020).

[9] J. C. Getelina, N. Gomes, T. Iadecola, P. P. Orth, and Y.-X. Yao, Adaptive variational quantum minimally entangled typical thermal states for finite temperature simulations, SciPost Phys. **15**, 102 (2023).

[10] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions, Phys. Rev. A **98**, 022322 (2018).

[11] C. Zoufal, R. V. Mishmash, N. Sharma, N. Kumar, A. Sheshadri, A. Deshmukh, N. Ibrahim, J. Gacon, and S. Woerner, Variational quantum algorithm for unconstrained black box binary optimization: Application to feature selection, Quantum **7**, 909 (2023).

[12] C. Zoufal, A. Lucchi, and S. Woerner, Variational quantum Boltzmann machines, Quantum Mach. Intell. **3**, 7 (2021).

[13] A. Miessen, P. J. Ollitrault, F. Tacchino, and I. Tavernelli, Quantum algorithms for quantum dynamics, Nat. Comput. Sci. **3**, 25 (2023).

[14] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[15] H. Zhao, M. Bukov, M. Heyl, and R. Moessner, Making trotterization adaptive for NISQ devices and beyond, PRX Quantum **4**, 030319 (2023).

[16] P. Mukhopadhyay, N. Wiebe, and H. T. Zhang, Synthesizing efficient circuits for Hamiltonian simulation, npj Quantum Inf. **9**, 31 (2023).

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

[18] IBM Quantum, IBM releases first-ever 1,000-qubit quantum chip (2023).

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

[20] F. Barratt, J. Dborin, M. Bal, V. Stojevic, F. Pollmann, and A. G. Green, Parallel quantum simulation of large systems on small NISQ computers, npj Quantum Inf. **7**, 79(2021).

[21] 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).

[22] M. Benedetti, M. Fiorentini, and M. Lubasch, Hardware-efficient variational quantum algorithms for time evolution, Phys. Rev. Res. **3**, 033083 (2021).

[23] L. Slattery, B. Villalonga, and B. K. Clark, Unitary block optimization for variational quantum algorithms, Phys. Rev. Res. **4**, 023072 (2022).

[24] B. Commeau, M. Cerezo, Z. Holmes, L. Cincio, P. J. Coles, and A. Sornborger, Variational Hamiltonian diagonalization for dynamical quantum simulation, arXiv:2009.02559.

[25] C. Cîrstoiu, Z. Holmes, J. Iosue, L. Cincio, P. J. Coles, and A. Sornborger, Variational fast forwarding for quantum simulation beyond the coherence time, npj Quantum Inf. **6**, 82 (2020).

[26] 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, arXiv:2102.04313.

[27] K. H. Lim, T. Haug, L. C. Kwek, and K. Bharti, Fast-forwarding with nisq processors without feedback loop, Quantum Sci. Technol. **7**, 015001 (2022).

[28] C. Mc Keever and M. Lubasch, Classically optimized Hamiltonian simulation, Phys. Rev. Res. **5**, 023146 (2023).

[29] R. Mansuroglu, T. Eckstein, L. Nützel, S. A. Wilkinson, and M. J. Hartmann, Variational Hamiltonian simulation for translational invariant systems via classical pre-processing, Quantum Sci. Technol. **8**, 025006 (2023).

[30] F. M. Sbahi, A. J. Martinez, S. Patel, D. Saberi, J. H. Yoo, G. Roeder, and G. Verdon, Provably efficient variational generative modeling of quantum many-body systems via quantum-probabilistic information geometry, arXiv:2206.04663.

[31] S.-i. Amari, Natural gradient works efficiently in learning, Neural Comput. **10**, 251 (1998).

[32] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4**, 269 (2020).

[33] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).

[34] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, Quantum fingerprinting, Phys. Rev. Lett. **87**, 167902 (2001).

[35] L. Cincio, Y. Subaşı, A. T. Sornborger, and P. J. Coles, Learning the quantum algorithm for state overlap, New J. Phys. **20**, 113022 (2018).

[36] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, Quantum algorithms revisited, Proc. R. Soc. London A **454**, 339 (1998).

[37] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature (London) **567**, 209 (2019).

[38] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, Phys. Rev. A **99**, 052323 (2019).

[39] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Simultaneous Perturbation stochastic approximation of the quantum fisher information, Quantum **5**, 567 (2021).

[40] J. C. Spall, An overview of the simultaneous perturbation method for efficient optimization, Johns Hopkins APL Tech. Dig. **19**, 482 (1998).

[41] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum Circuits, Nat. Commun. **12**, 1791 (2021).

[42] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[43] C. Ortiz Marrero, M. Kieferová, and N. Wiebe, Entanglement-induced barren plateaus, PRX Quantum **2**, 040316 (2021).

[44] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[45] P. J. Ollitrault, A. Baiardi, M. Reiher, and I. Tavernelli, Hardware efficient quantum algorithms for vibrational structure calculations, Chem. Sci. **11**, 6842 (2020).

[46] C.-Y. Park and N. Killoran, Hamiltonian variational ansatz without barren plateaus, arXiv:2302.08529.

[47] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).

[48] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, Variational quantum simulation of general processes, Phys. Rev. Lett. **125**, 010501 (2020).

[49] C. Zoufal, D. Sutter, and S. Woerner, Error bounds for variational quantum time evolution, Phys. Rev. Appl. **20**, 044059 (2023).

[50] R. Latała, Some estimates of norms of random matrices, Proc. Am. Math. Soc. **133**, 1273 (2005).

[51] Qiskit contributors, Qiskit: An open-source framework for quantum computing, doi:10.5281/zenodo.2562110.

[52] A. Cultrera and L. Callegaro, A simple algorithm to find the L-curve corner in the regularisation of ill-posed inverse problems, IOP SciNotes **1**, 025004 (2020).

[53] E. M. Stoudenmire and S. R. White, Minimally entangled typical thermal state algorithms, New J. Phys. **12**, 055026 (2010).

[54] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting Qubits: Current state of play, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[55] IBM Quantum, https://quantum-computing.ibm.com/ (2023), accessed on 07/26/2023.

[56] C. Tornow, N. Kanazawa, W. E. Shanks, and D. J. Egger, Minimum quantum run-time characterization and calibration via restless measurements with dynamic repetition rates, Phys. Rev. Appl. **17**, 064061 (2022).

[57] D. J. Egger, M. Werninghaus, M. Ganzhorn, G. Salis, A. Fuhrer, P. Müller, and S. Filipp, Pulsed reset protocol for fixed-frequency superconducting qubits, Phys. Rev. Appl. **10**, 044030 (2018).

[58] P. Magnard, P. Kurpiers, B. Royer, T. Walter, J.-C. Besse, S. Gasparinetti, M. Pechal, J. Heinsoo, S. Storz, A. Blais, and A. Wallraff, Fast and unconditional all-microwave reset of a superconducting qubit, Phys. Rev. Lett. **121**, 060502 (2018).

[59] G. Gentinetta, A. Thomsen, D. Sutter, and S. Woerner, The complexity of quantum support vector machines, Quantum **8**, 1225 (2024).

[60] Y. H. Wang, On the number of successes in independent trials, Statistica Sinica **3**, 295 (1993).