

Parameterization and optimizability of pulse-level VQEs

Kyle M. Sherbert,^{1,2,3} Hisham Amer,^{2,3} Sophia E. Economou,^{2,3} Edwin Barnes,^{2,3} and Nicholas J. Mayhall^{1,3,*}

¹*Department of Chemistry, Virginia Tech, Blacksburg, VA 24061*

²*Department of Physics, Virginia Tech, Blacksburg, VA 24061*

³*Virginia Tech Center for Quantum Information Science and Engineering, Blacksburg, VA 24061, USA*

In conventional variational quantum eigensolvers (VQEs), trial states are prepared by applying series of parameterized gates to a reference state, with the gate parameters being varied to minimize the energy of the target system. Recognizing that the gates are intermediates which are ultimately compiled into a set of control pulses to be applied to each qubit in the lab, the recently proposed ctrl-VQE algorithm takes the amplitudes, frequencies, and phases of the pulse as the variational parameters used to minimize the molecular energy. In this work, we explore how all three degrees of freedom interrelate with one another. To this end, we consider several distinct strategies to parameterize the control pulses, assessing each one through numerical simulations of a transmon-like device. For each parameterization, we contrast the pulse duration required to prepare a good ansatz, and the difficulty to optimize that ansatz from a well-defined initial state. We deduce several guiding heuristics to implement practical ctrl-VQE in hardware, which we anticipate will generalize for generic device architectures.

I. INTRODUCTION

Variational quantum eigensolvers (VQEs) are among the most promising candidates for achieving useful computations in chemistry on near-term quantum computers [1–6]. At their core, they are predict-and-test methods, where a quantum state, determined by a set of classical parameters as specified by an *ansatz*, is prepared on the quantum computer, and its energy measured. Then a new quantum state is prepared with a new set of parameters, selected by any classical optimization algorithm, and the process is repeated until the energy is minimized. This minimal energy is an upper-bound to the ground-state energy of the system, and, if the optimization is carried out perfectly, the final state is the best approximation to the ground state attainable with the chosen ansatz.

Clearly, a key decision in any VQE is the choice of ansatz, typically expressed as a series of parameterized unitary *gates* in a quantum circuit. While the original formulation of VQE utilized an ansatz inspired by the underlying chemistry [7], many novel ansätze have been proposed which balance chemical intuition with information about which parametric and entangling operations are most easily implemented on the device [8–14]. In either case, for experimental realization, the quantum circuit must be compiled into a sequence of electromagnetic pulses that are applied to the device. For transmon-type devices (those considered in this work), the control pulses are microwave signals which drive the states of each physical qubit as prescribed by each unitary gate in the circuit. Unfortunately, if each gate is to be implemented with high fidelity in modern hardware, even the most compact ansätze for moderately-sized systems typically compile into control pulses with a duration far exceeding the coherence time of the device [15].

Three of the authors previously proposed the algorithm ctrl-VQE [15, 16], which takes the idea of a hardware-efficient ansatz to the extreme by parameterizing the actual physical control pulses used in the lab, bypassing the use of gates entirely. Designing the ansatz at the pulse level allows drastically shorter evolution times, even approaching the quantum speed limits imposed by the hardware [17], and hypothetically enabling the VQE to study much larger or more complex systems. The methods and mathematical theory behind ctrl-VQE are built upon the closely related field of quantum control, and we refer the reader to Refs. [18, 19] for introductions to this topic. Recently, several works have built on the original ctrl-VQE [20–27], including a handful with proof-of-concept hardware demonstrations on simple systems.

Most of these efforts have focused on parametric modifications of established one- and two-qubit gates, while our original work has emphasized *de novo* pulses which act more like a multi-qubit gate. In fact, ctrl-VQE represents the most flexible variational ansatz possible for any given hardware. While this approach is very powerful, it presents the challenge that there is an overwhelming number of accessible degrees of freedom. There is typically one or more control pulses simultaneously applied to each qubit, and each control pulse is parameterized by an amplitude (A), a phase (ϕ), and an (angular) frequency (ν), each of which may be varied in time, resulting in an ostensibly infinite number of parameters (before hardware bandwidth limits are considered). An important open question is which parameters are most important for quickly preparing a target state while avoiding optimization issues.

In this paper, we investigate the impact that each of the pulse parameters tends to have and establish useful heuristics for practical pulse parameterizations. For example, we present empirical evidence that representing the amplitude A and phase ϕ together as a complex amplitude with real and imaginary components enables notably more efficient optimizations than when varying

* nmayhall@vt.edu

each pulse parameter independently. Furthermore, we find that varying the phase ϕ is likely to be far more effective than varying drive frequency ν in a practical ctrl-VQE experiment, since varying the latter tends to significantly impede the efficiency of optimization, without providing any commensurate improvement in controllability or minimal evolution time. We mainly focus on transmon-type architectures in this work, but we anticipate our conclusions can guide pulse-level VQE experiments on a wide array of devices.

The paper is organized as follows: In Section II, we review the mathematical basis of the ctrl-VQE algorithm, and we introduce the various choices of parameterization for our ansatz. In Section III, we briefly describe the methods and parameters used to generate the data presented in this paper. In Section IV, we present a series of experiments contrasting the optimizability of each parameterization. Section V summarizes our results and condenses our findings into a set of guidelines for experimental realizations of ctrl-VQE.

II. THEORY

We take as our starting point a quantum observable \hat{O} (e.g., a molecular Hamiltonian), and a reference state $|\psi_0\rangle$ (e.g., the Hartree-Fock state), both mapped onto qubits via a suitable transformation (e.g., Jordan-Wigner). Our variational ansatz $|\psi(\boldsymbol{\theta})\rangle$ will take the form:

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_0\rangle, \quad (1)$$

where $U(\boldsymbol{\theta})$ is the unitary state-preparation operator. The expectation value of \hat{O} is the energy $E(\boldsymbol{\theta})$:

$$E(\boldsymbol{\theta}) \equiv \langle \psi(\boldsymbol{\theta}) | \hat{O} | \psi(\boldsymbol{\theta}) \rangle, \quad (2)$$

which is the function that the optimization routine in ctrl-VQE will attempt to minimize.

In ctrl-VQE, the unitary $U(\boldsymbol{\theta})$ is the physical evolution under a device Hamiltonian (as distinct from traditional VQEs, where it is typically a sequence of logical operations on the qubit space). The device Hamiltonian may be separated into a static component \hat{H}_0 and a drive component $\hat{V}(t)$. Variational parameters are typically associated with the time-dependent drive $\hat{V}(t)$. Frequently, $\hat{V}(t)$ is comprised of several independent drives; in this work, we consider an independent drive $\hat{V}_q(t)$ applied to each qubit q . As such, our choice of ansatz in ctrl-VQE is essentially the unitary time evolution operator:

$$U(\boldsymbol{\theta}) = \hat{\mathcal{T}} \exp \left(-i \int_0^T dt \left[\hat{H}_0 + \sum_q \hat{V}_q(\boldsymbol{\theta}, t) \right] \right), \quad (3)$$

where \mathcal{T} is the time-ordering operator and T is the total pulse duration. While the underlying physical mechanism varies for different quantum computing architectures, the drive term in a qubit Hamiltonian can typically

be modeled as an interaction between a dipole moment (the qubit) and a classical electric field (the *control field*):

$$\hat{V}_q(t) \propto D(t)(\hat{a}_q + \hat{a}_q^\dagger), \quad (4)$$

where \hat{a}_q is the bosonic lowering operator acting on the transmon qubit state. The electric field arises from an applied voltage that oscillates sinusoidally at microwave frequencies:

$$D(t) = A \cos(\nu t + \phi). \quad (5)$$

The parameters we have control over are therefore the amplitude A , the frequency ν , and the phase ϕ of the field, each of which may themselves be functions of time.

When the amplitude is relatively weak with respect to the drive frequency, and when the difference between the drive frequency and the qubit frequency is negligible compared to either of these quantities, the rotating wave approximation can be used to convert the sinusoidal form of Eq. (4) into a complex phase:

$$\hat{V}_q(t) \xrightarrow{\text{RWA}} \Omega e^{i\nu t} \hat{a}_q + \text{h.c.}, \quad (6)$$

where the complex amplitude Ω is defined in polar notation by the amplitude (i.e., modulus) and phase:

$$\Omega \equiv A e^{i\phi} / 2. \quad (7)$$

Alternatively, we could parameterize our control field equivalently in ‘‘Cartesian’’ notation, defining the real-valued amplitudes α and β as the real and imaginary parts of Ω :

$$\Omega \equiv \alpha + i\beta. \quad (8)$$

This Cartesian parameterization is often more convenient in software implementations.

In the absence of any control field, a single qubit will evolve under its static Hamiltonian H_0 , which in transmons can be modeled as an anharmonic oscillator with a characteristic frequency ω describing the energy gap between the lowest two states $|0\rangle$ and $|1\rangle$. In such cases, it is common to make a coordinate transformation from the lab frame into the rotating frame of the device.

$$U_{\text{RF} \leftarrow \text{LAB}}^{(t)} = e^{iH_0 t} \quad (9)$$

The Hamiltonian in this frame is written as:

$$V_R(t) = \Omega e^{i\Delta t} \hat{a} + \text{h.c.}, \quad (10)$$

where the detuning Δ is simply the difference between the drive frequency ν and the qubit frequency ω :

$$\Delta \equiv \nu - \omega, \quad (11)$$

so that $\Delta = 0$ corresponds to driving on resonance. Because the drive frequency ν and the detuning Δ differ by a constant, they will behave identically under a gradient-based optimization. In what follows, we choose to use Δ

Label	Ω	ν	Figs
$\{\alpha\beta\}$	2 param. per window, Cartesian	Resonant.	1, 2, 4, 5
$\{\alpha\beta\}_2$	Same as $\{\alpha\beta\}$, twice as many windows.		2
$\{\alpha\beta\}_\infty$	Same as $\{\alpha\beta\}$, continuous windows.		2
$\{A\phi\}$	2/window, polar	Resonant.	4
$\{\alpha\}$	1/window, real only	Resonant.	5
$\{\alpha\Delta\}$	1/window, real only	1/pulse	5
$\{\alpha\beta\Delta\}$	2/window, Cartesian	1/pulse	5

TABLE I. Summary of pulse parameterizations considered in this paper. Except when otherwise noted, optimizations begin with all complex amplitudes Ω initialized to zero, and drive frequencies ν initialized on resonance (i.e., $\Delta = 0$).

as our variational parameter. In multi-qubit systems, static coupling between qubits leads to a dressing of the bare qubit parameters, but the detuning Δ of any given pulse from the frequency ω of the qubit it targets remains a useful parameter.

In our previous work [15, 16], we took the amplitude Ω to be real, essentially restricting the phase ϕ to either 0 (when $\Omega \geq 0$) or π (when $\Omega < 0$). In the present work, we instead consider various parameter sets, including complex amplitudes in some cases; all parameter sets are listed in Table I. Each parameter can be varied in an optimization loop, between control pulses. Furthermore, each parameter could in principle be varied as an arbitrary function in time, throughout the pulse duration. As a practical matter, drive frequencies are more difficult to experimentally vary smoothly, so in this paper as in our previous papers we will take Δ as time-independent for any given control pulse. For simplicity, we will limit ourselves to “windowed” pulses where each window has a constant amplitude and phase; arbitrary pulse shapes can be approximated by taking windows of arbitrarily short duration.

III. METHODS

The results presented in this paper focus on the lithium hydride molecule (LiH) mapped onto four qubits as a case study, usually with a nuclear separation of 3.0Å. We use the `pyscf` package [28] to calculate electronic integrals for each molecular geometry, and we use the `qiskit` package [29] to construct the second-quantized molecular Hamiltonian, map it onto qubits with the parity mapping, and apply two-qubit reduction. [30] We use the Hartree-Fock singlet state as our reference, and we represent the Hamiltonian as a matrix in the Hartree-Fock basis so that the reference state is a computational basis state.

We use our own Julia code [31, 32] to simulate unitary time evolution under the rotating wave approximation with arbitrary control pulses, using Trotterized time evolution with twenty time steps per nanosecond. Before the pulse, the qubits are prepared into the Hartree-Fock reference state, and after the pulse, the resulting statevec-

tor (in the rotating frame) is used to measure the energy, i.e., the expectation value of our molecular Hamiltonian. We measure in the rotating frame so that when all pulse amplitudes are set to zero, the cost function (molecular energy) does not change in time.

The static Hamiltonian H_0 is modeled as a set of coupled harmonic oscillators, with each qubit truncated to the first two energy levels:

$$\hat{H}_0 = \sum_q \omega_q \hat{a}_q^\dagger \hat{a}_q + \sum_{\langle p,q \rangle} g_{pq} (\hat{a}_p^\dagger \hat{a}_q + \hat{a}_q^\dagger \hat{a}_p). \quad (12)$$

Note that this is a common model for transmon devices, [33, 34] excepting an additional anharmonicity term $\hat{a}_q^\dagger \hat{a}_q^\dagger \hat{a}_q \hat{a}_q$ which is dropped in the two-level approximation. In this paper, coupling strengths are fixed to $g_{pq} = 0.02 \cdot 2\pi$ GHz if $q = p + 1$ (i.e., linear nearest-neighbor coupling), and qubit frequencies are equally spaced such that $\omega_q = (4.80 + 0.02q) \cdot 2\pi$ GHz. These parameters roughly match those found in real IBMQ devices, while being systematically scalable to larger systems.

We use the `Optim.jl` implementation of BFGS (a quasi-Newton optimization algorithm) [35] with analytical gradients (see Appendix A and [36]) to iteratively minimize the molecular energy with respect to a specified drive parameterization. Hardware bounds on the maximum amplitude are enforced by including a smooth penalty term in the cost-function on any time step for which the pulse amplitude exceeds $\Omega_{\max} = 0.02 \cdot 2\pi$ GHz. Optimization terminates when the maximum of the analytical gradient has converged to within 1e-6.

IV. RESULTS

Previous works on ctrl-VQE have emphasized the role of the amplitude A as a function of time. In this work, we will assess the relative significance of the phase ϕ and the detuning Δ , as measured by the pulse duration necessary to achieve high-accuracy estimates of a ground state energy. We will make extensive use of “error-vs-time” plots and “iterations-vs-time” plots, which show how the accuracy and iteration count of a ctrl-VQE optimization change as a function of pulse duration.

Each data point in these plots is obtained by simulating ctrl-VQE to optimize parameters in a pulse with fixed duration T . At the start of optimization, pulse parameters are initialized to zero, meaning the pulse shape starts flat and the drive frequencies start on resonance. The pulse is divided into windows with uniform spacing s , with longer pulses divided into more windows (details given in Section IV A). In the “error-vs-time” plots, we plot the difference between the final optimized energy and the value obtained from exact diagonalization of the Hamiltonian matrix. In the “iterations-vs-time” plots, we plot the number of iterations the BFGS optimizer requires to converge. This number is unique to our choice

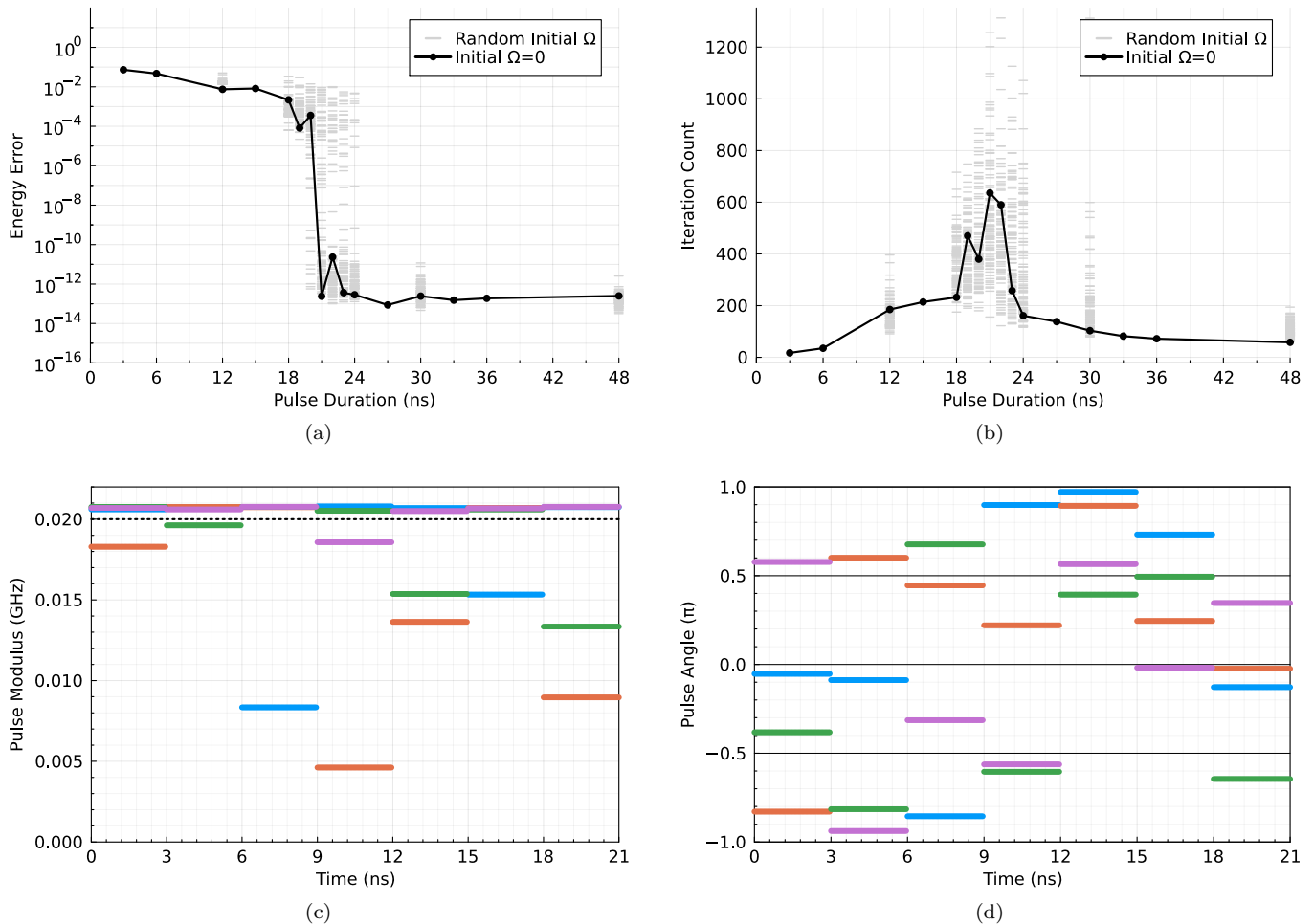


FIG. 1. (1a) Ground state energy error vs. pulse duration for a ctrl-VQE simulation of LiH with nuclear separation 3.0\AA , using the $\{\alpha\beta\}$ parameterization. Each pulse is divided into as many uniformly spaced windows as possible such that window length satisfies $s \geq 3.0$ ns. For the black curve, optimization is carried out with all parameters starting at zero. Gray dashes mark optimized runs when parameters are initialized with random values. (1b) The number of BFGS iterations needed to obtain the energies in (1a). (1c) Optimized pulse amplitudes $A(t)$ for the ctrl-VQE run at $T = 21$ ns. Each color corresponds to the drive on a different qubit. The dashed line marks the bound above which penalties are imposed, ostensibly the maximum drive allowed by the device. (1d) Optimized pulse phases $\phi(t)$ for the ctrl-VQE run at $T = 21$ ns. The dashed lines occur at intervals of $\pi/2$, separating quadrants of the unit circle.

of optimizer, but it allows a heuristic comparison of the expected number of circuit evaluations required to obtain the accuracies presented in the corresponding “error-vs-time” plot.

Typical behavior in an “error-vs-time” plot is for the error to start off large, as seen in Fig. 1a: When the pulse duration is very small, the time evolution is insufficient to drive the system far from the reference state. As pulse duration is increased, the optimized energy error tends to decrease exponentially (appearing linear on the semi-log plot), but slowly. However, at some critical time, there is a sharp transition to highly accurate energies, where the optimized energy error is on the order of numerical precision; increasing pulse duration beyond this point has negligible impact on the accuracy. The pulse duration

for which this transition occurs is closely related to the *minimal evolution time* (MET) from optimal control theory, which is the absolute shortest time required to prepare one state from another, subject to a given set of pulse constraints [17, 37]. According to optimal control theory, amplitudes will tend to saturate their bounds as the pulse duration approaches the MET. For real-valued amplitudes (the $\{\alpha\}$ and $\{\alpha\Delta\}$ parameterizations), the MET is achieved with *bang-bang* pulse shapes which are characterized as pulses where the amplitude switches between $\pm\Omega_{\max}$, as observed in our previous work [16] and in a control-theoretic analysis of the quantum approximate optimization algorithm [38]. For complex-valued amplitudes, optimal control theory imposes no *a priori* constraints on the phases ϕ , but the amplitude param-

ters A will tend to saturate the bounds $\pm\Omega_{\max}$. One can see this effect in Figs. 1c and 1d, which show the amplitudes and phases optimized from zero at $T = 21$ ns, the shortest pulse duration in Fig. 1a which results in a ctrl-VQE run obtaining energy errors less than 10^{-8} Ha when all parameters are initialized to zero.

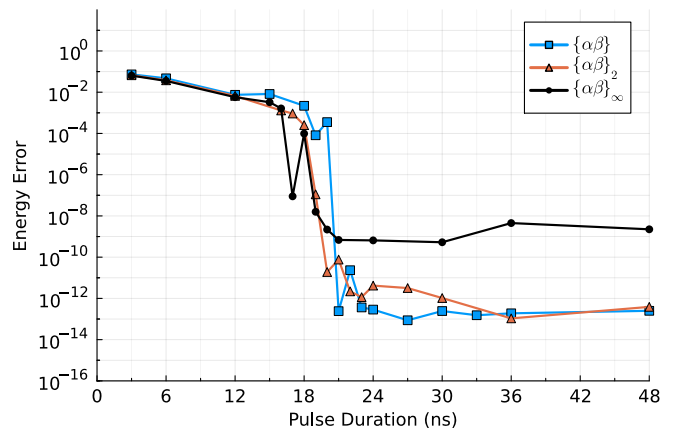
Converging to the true MET is extremely computationally expensive, as it requires, in principle, an exhaustive search over all parameter space. The approach we have taken previously [16] simply samples over multiple optimizations using randomly initialized pulses. For example, the gray ticks we plot in Fig. 1a are the optimized energy errors from 100 runs when the initial pulse parameters are uniformly sampled from the allowed range. Several runs with $T = 20$ ns are successful, and it may well be that a perfect global optimization could locate successful pulses with even lower duration. However, we choose to use the more computationally tractable zero-pulse initialization as a heuristic for the remainder of this work, and we will refer to the evolution time in an “error-vs-time” plot as the “*effective*” minimal evolution time (eMET) T_0 .

Interestingly, an “iterations-vs-time” plot (Fig. 1b) exhibits an iteration count that is relatively low for most pulse durations, *except* for a broad and large peak near the eMET T_0 . This is due to the saturation of the amplitude bounds observed above, straining the optimizer as it balances reductions to the molecular energy and to the penalty term enforcing the amplitude bounds. Therefore, we anticipate real ctrl-VQE experiments will want to target pulse durations sufficiently beyond the eMET, to achieve the best trade-off between the number of circuit evaluations and decoherence error.

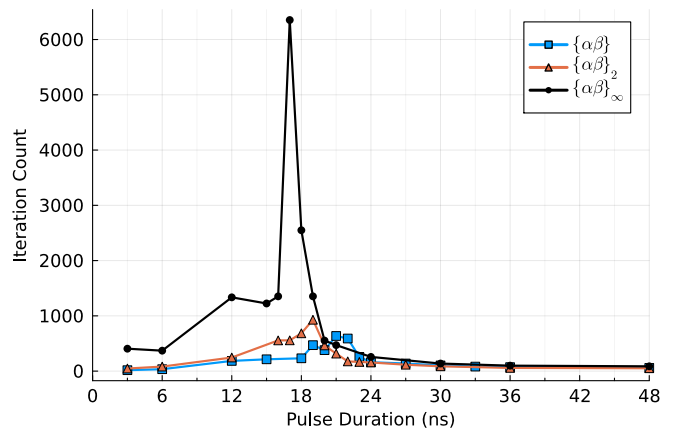
A. Characterizing eMET for different window lengths

Throughout this paper, pulses are divided into a number of windows chosen such that the duration of each window is as close as possible to, but never below, some minimum duration s_{\min} . For example, if $s_{\min} = 3$ ns, a $T = 24$ ns pulse would be divided into eight windows of $s = 3$ ns each, while a $T = 23.8$ ns pulse would be divided into seven windows of $s = 3.4$ ns each. This choice ensures that longer pulses have more degrees of freedom while simultaneously dampening high-frequency components in the control signals that are difficult to implement experimentally. By choosing $s_{\min} = 3$ ns, we roughly approximate an effective bandwidth of roughly $1/2s_{\min} \sim 167$ MHz in the microwave pulse generator.

Figure 2 shows the “error-vs-time” and “iterations-vs-time” plots comparing different choices of s_{\min} for a LiH molecule with a nuclear separation of 3.0\AA , using the parameter set $\{\alpha\beta\}$ (complex amplitudes parameterized with Cartesian notation, and the drive frequency for each qubit is fixed on resonance). The “continuous” curve $\{\alpha\beta\}_{\infty}$ assigns independent α and β parameters for each



(a)



(b)

FIG. 2. (2a) Ground state energy error vs. pulse duration of ctrl-VQE applied to LiH with nuclear separation 3.0\AA , using the $\{\alpha\beta\}$ parameterization with three different numbers of time windows. Each pulse is divided into as many uniformly spaced windows as possible such that the window length s is constrained as labeled. The $\{\alpha\beta\}_{\infty}$ simulations use as many windows as there are time-steps in the evolution, such that $s = 0.05$ ns. All parameters are initialized to zero in the optimization steps. (2b) The number of BFGS iterations needed to obtain the energies in (2a).

point in the Trotterized time evolution, corresponding to a s on the order of the Trotter step, ~ 0.05 ns. Despite the vastly larger number of parameters in this case compared to the other two cases shown in the figure, the eMET T_0 changes by only about 3 ns or so, emphasizing the importance of evolution time over degrees of freedom. As can be seen in the “iterations-vs-time” plot, the larger number of parameters does have a significant effect on computational cost, so we will adopt the reasonable value of $s_{\min} = 3.0$ ns for most results in this paper.

We note that the somewhat higher asymptote observed for energy errors in Fig. 2a is an artifact of our optimization routine using the L_{∞} norm rather than, say, the L_2 norm as the convergence criterion. Thus, gradient vec-

tors with a very large number of small but non-zero values will be somewhat prematurely flagged as converged. We anticipate that a correspondingly stricter choice of convergence criterion in the runs with more parameters would result in a taller, broader, and more symmetric “iterations-vs-time” curve than the one observed in Fig. 2b, though we have not confirmed this numerically.

We can gain an intuitive understanding for why the number of parameters does not have a significant impact on the eMET T_0 by studying the cartoon in Fig. 3a: When the complex amplitude Ω varies continuously in time, the optimized pulses generate the shortest possible path through Hilbert space between the reference state and the ground state. The time it takes to traverse this path is determined by the *Quantum Speed Limit*, which dictates that the rate at which a system can move through Hilbert space is bounded by the norm of the drive Hamiltonian, determined in our case by the drive amplitudes A [17]. Thus, hardware constraints on the maximal drive amplitude A for each pulse impose a *fundamental* MET T_0 required to prepare our ground state. When we impose finite windows on the pulse shape, the optimal path traced through Hilbert space is perturbed from the optimum. Longer windows result in larger perturbations, and therefore longer evolution times. However, as long as the window duration s is relatively small with respect to the pulse duration T , the dominant contribution to the evolution time remains the original fundamental MET T_0 .

Fig. 3b provides a particularly striking visualization of this perturbation. First, we take two optimized pulses which successfully prepare the target state, one with $s = 0.05$ ns (“continuous”) and $T = 17$ ns, and one with $s = 1.5$ ns and $T = 18$ ns. We note that optimization of the $s = 1.5$ ns case required randomly initializing pulse parameters to find a successful pulse with $T = 18.0$ ns, as opposed to the data presented in Fig. 2, which reports results when initializing all pulses from zero. Next, we calculate the trajectories $|\psi(t)\rangle, |\phi(t)\rangle$ in the rotating frame, when applying each pulse. Finally, for every pair of times t, t' , we evaluate the fidelity $|\langle\psi(t)|\phi(t')\rangle|^2$. The white spot in the bottom left of Fig. 3b indicates that both pulses are applied to the same initial state, while the white spot in the upper right indicates that both pulses prepare the same final state after the full pulse duration. The bright band roughly along the diagonal of the majority of the plot indicates that both trajectories largely overlapped (except in the beginning), corroborating the qualitative picture in Fig. 3a.

B. Polar vs Cartesian Parameterization

In this section, we contrast ctrl-VQE using Cartesian notation (Eq. 8) and polar notation (Eq. 7). Either choice leads to the exact same dynamics for any single pulse, but a finite step in the different parameter sets will have different resulting pulses, meaning a gradient-based op-

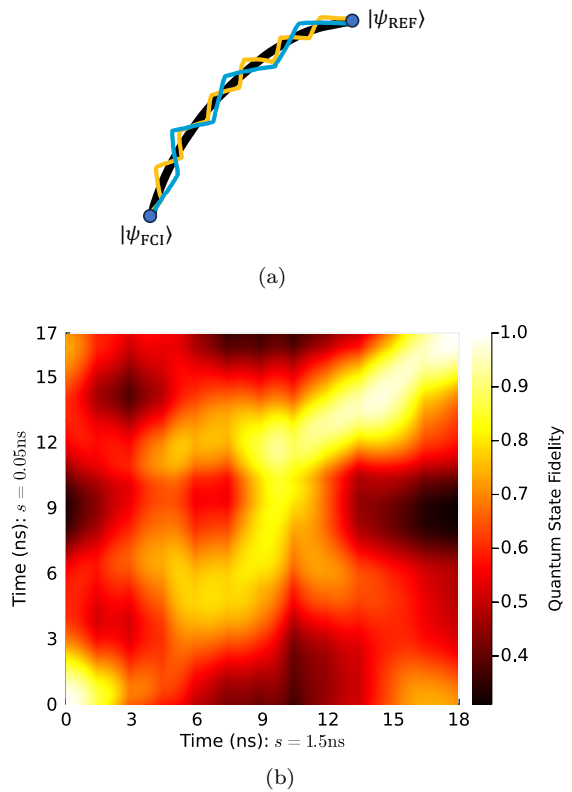
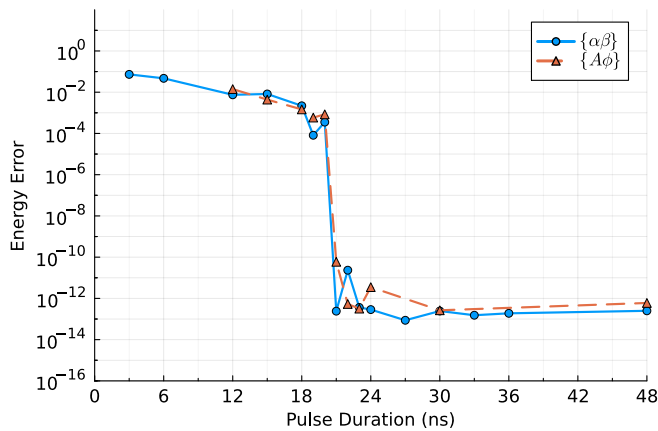
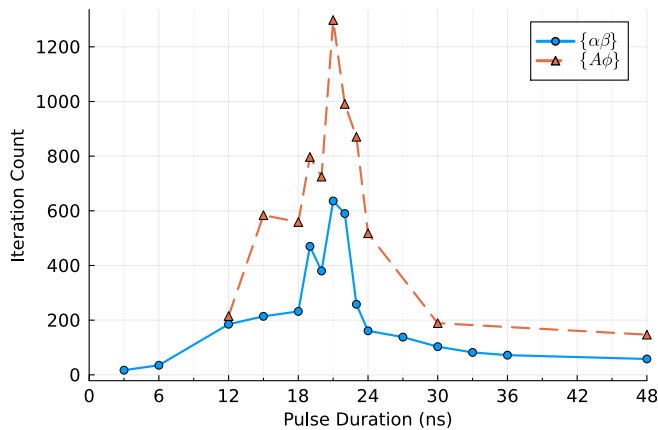


FIG. 3. (3a) A cartoon depiction of how finite window lengths affect evolution time. When pulse parameters are allowed to vary continuously in time, denoted by the black curve, ctrl-VQE can in principle find the shortest possible path between the reference state and the ground state. When parameters are constrained to be constant over finite windows, ctrl-VQE can still approximate this ideal path. The total path length, and therefore the total evolution time, will be marginally longer, but the main contribution to evolution time is the length of the ideal path. (3b) The fidelity between quantum states as a function of time, following trajectories solving LiH with nuclear separation 3.0\AA , from two optimized pulses: a 17.0 ns pulse where parameters were varied quasi-continuously, and an 18.0 ns pulse where parameters were restricted to window lengths of $s = 1.5$ ns each. Bright colors indicate high overlap at the corresponding points in the trajectory, while black indicates orthogonal states.

imizer like BFGS may well proceed along different optimization trajectories. In Fig. 4, we show “error-vs-time” and “iterations-vs-time” plots comparing the two different notations applied to LiH with a bond separation of 3.0\AA , holding frequencies fixed on resonance. The energy plots and the eMET T_0 are essentially identical, indicating that both runs identified pulses preparing essentially the same estimate of the ground state for each pulse duration T . Note that the actual pulse shapes do vary, but more than one pulse generates the same state. However, the optimization trajectories for the polar notation are notably more expensive, so we recommend adopting Cartesian notation. We assume this arises due to larger off-diagonal matrix elements in the pulse Hessian of the



(a)



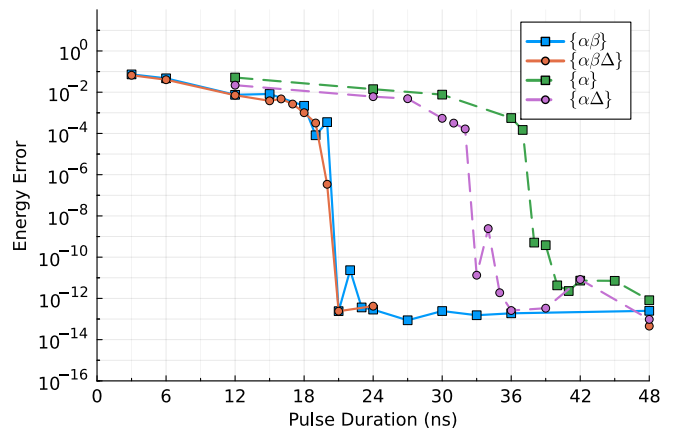
(b)

FIG. 4. (4a) Ground state energy error vs. pulse duration for ctrl-VQE applied to LiH with nuclear separation 3.0\AA , parameterizing each pulse’s complex amplitude with polar $\{A\phi\}$ or Cartesian $\{\alpha\beta\}$ notations. Each pulse is fixed on resonance and divided into as many uniformly spaced windows as possible such that the window length $s \geq 3.0$ ns. All parameters are initialized to zero in the optimization. (4b) The number of BFGS iterations needed to obtain the energies in (4a).

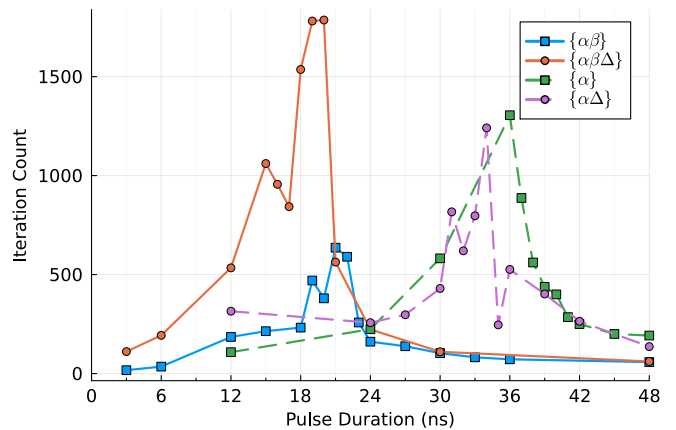
polar representation, but this is a topic we will explore in future work.

C. Importance of ϕ and Δ

Previous works on ctrl-VQE [15, 16] used the $\{\alpha\Delta\}$ parameterization, restricting drive phases to $\phi \in \{0, \pi\}$ and allowing drive frequencies to vary off resonance. In this section, we will explore what happens when the phase degree of freedom is included ($\{\alpha\beta\Delta\}$, in Cartesian notation), when the frequency degree of freedom is removed ($\{\alpha\}$), and when both changes are effected ($\{\alpha\beta\}$). Fig. 5 gives “error-vs-time” and “iterations-vs-time” plots for all four parameterizations, applied to LiH with a bond separation of 3.0\AA . For the sake of fair com-



(a)



(b)

FIG. 5. (5a) Ground state energy error vs. pulse duration for ctrl-VQE applied to LiH with nuclear separation 3.0\AA , for various parameterizations. Each pulse for the $\{\alpha\beta\}$ and $\{\alpha\beta\Delta\}$ curves is divided into as many uniformly spaced windows as possible such that the window length $s \geq 3.0$ ns, while those for $\{\alpha\}$ and $\{\alpha\Delta\}$ use $s \geq 1.5$ ns, so that the total number of parameters scale similarly. All parameters are initialized to zero in the optimization. (5b) The number of BFGS iterations needed to obtain the energies in (5a).

parison, pulses with real-valued amplitudes are divided into twice as many windows, so that the total number of parameters is similar for each value of T (parameterizations varying the frequency include one more parameter for each pulse, for a total of four more parameters).

Unsurprisingly, the longest eMET T_0 is observed with the most restrictive parameterization $\{\alpha\}$, and the second longest is the second-most restrictive, $\{\alpha\Delta\}$. (Recall that varying Δ is more restrictive than varying β , since each pulse has only one Δ parameter, irrespective of the number of windows.) Perhaps surprisingly, the remaining two parameterizations $\{\alpha\beta\}$ and $\{\alpha\beta\Delta\}$ are practically indistinguishable, except that including the frequency degree of freedom puts a heavy strain on the optimizer (this effect was also observed in Ref. [25]). Apparently, once

the pulse amplitude is allowed to become complex, there is no further variational flexibility afforded by the frequency detuning.

We must emphasize that this is *not* related to the dimensionality of the optimization problem. The number of degrees of freedom is similar (within a constant) for all four parameterizations. We verify this numerically by calculating the effective quantum dimension [39] of our trial state at several steps of the optimization. This quantity is defined as the rank of the quantum Fisher information [40], and it describes the number of directions in which infinitesimal variations in our parameters can drive the system. When the pulse has a sufficient number of parameters, we anticipate the effective quantum dimension to saturate at the number of independent degrees of freedom in the Hilbert space (30, for a four-qubit system). Computational basis states are *singular*, isolated points with reduced dimension. Thus, we expect (and find) initial pulses, with $\Omega = 0$, to have an effective quantum dimension of 16, irrespective of the number of parameters in the pulse. However, random perturbations from a singular point are almost sure to result in a state with full effective quantum dimension [39, 41], and indeed we find that the effective quantum dimension is saturated at 30 for all subsequent trial states when $T \geq 12$ ns.

We now address why the frequency degree of freedom has no apparent impact on the eMET, so long as we include the phase degree of freedom (i.e., contrasting the $\{\alpha\beta\}$ and $\{\alpha\beta\Delta\}$ parameterizations). This can be *partly* explained by noting that the drive phase ϕ and drive frequency ν are intrinsically related: the argument to the trigonometric function in Eq. (5) is the sum $\nu t + \phi$. In the limit where the phase ϕ is allowed to vary continuously, one could implement perturbations to the drive frequency $\nu \rightarrow \nu + \Delta$ without actually changing the drive frequency directly, by selecting $\phi \rightarrow \phi + \Delta t$.

However, we observe essentially identical eMETs with and without a frequency degree of freedom even for finite window lengths, where the phase ϕ cannot have a linear dependence on time. To obtain a better intuition for this behavior, we will consider the analytical solution for a single qubit driven by a pulse with a constant amplitude A , detuning Δ , and phase ϕ . We will describe our quantum state using the Bloch sphere parameterization (θ, φ) :

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle. \quad (13)$$

The objective in ctrl-VQE is to prepare a particular state from a reference state; in this example, let us take the north pole $(0, 0)$ as our reference state ($|\psi_0\rangle = |0\rangle$), and the target state corresponds to the point (θ_0, φ_0) .

Under the rotating wave approximation, the system evolves under the Hamiltonian in Eq. (10) according to the time-dependent Schrödinger equation, which can be solved exactly. This is the Rabi problem, which can be found in standard quantum mechanics textbooks, e.g. see Ref. [42]. The Cartesian coordinates of the state on a

Bloch sphere in the rotating frame are:

$$\langle X \rangle(t) = x(t) \cdot \cos\phi + y(t) \cdot \sin\phi, \quad (14)$$

$$\langle Y \rangle(t) = y(t) \cdot \cos\phi - x(t) \cdot \sin\phi, \quad (15)$$

$$\langle Z \rangle(t) = \cos^2(\lambda t/2) + \sin^2(\lambda t/2) \cdot \cos\theta_L, \quad (16)$$

where

$$\theta_L = 2 \arctan(A/\Delta) \quad (17)$$

is the largest attainable θ on the Bloch sphere, using the following definitions:

$$x(t) \equiv -\sin(\theta_L/2) \cdot \sin(\Delta t) \cdot \sin(\lambda t) + \sin(\theta_L) \cdot \cos(\Delta t) \cdot \sin^2(\lambda t/2), \quad (18)$$

$$y(t) \equiv -\sin(\theta_L/2) \cdot \cos(\Delta t) \cdot \sin(\lambda t) - \sin(\theta_L) \cdot \sin(\Delta t) \cdot \sin^2(\lambda t/2), \quad (19)$$

$$\lambda \equiv \sqrt{A^2 + \Delta^2}. \quad (20)$$

The trajectory of the qubit state rotates around an axis oriented along a polar angle $\theta_L/2$ and an azimuthal angle that begins along ϕ but precesses slowly in time at a rate Δ . In addition to setting θ_L , the amplitude A also determines the speed at which this path is traversed, so we will always want it as large as our hardware allows in order to minimize the necessary evolution time.

When the detuning vanishes, $\Delta = 0$, the axis is in the x - y plane; since we start from the north pole, the trajectory is a meridian, with the longitude fixed by the pulse phase ϕ . Therefore, the azimuthal coordinate of our state φ is constant (up to $\pm\pi$). The most direct state preparation strategy is to choose $\phi = \varphi_0$ and drive on resonance ($\Delta = 0$). Since the shortest path (i.e., the geodesic) between a pole and any point on a sphere is along the meridian, this choice guarantees the shortest possible evolution time, subject to constraints on the amplitude A .

If we restrict ourselves to real-valued pulses as in the previous section (i.e., $\phi \in \{0, \pi\}$), the only azimuthal coordinates we can access with resonant pulses are $\varphi \in \{0, \pi\}$ (see Fig. 6a). If our target state has any other azimuthal coordinate φ_0 , we *must* use an off-resonant pulse (i.e., $\Delta \neq 0$). Doing so tilts the angle of the axis about which our trajectory rotates; the trajectory is no longer a geodesic. Thus, not only are large polar angles $\theta > \theta_L$ entirely inaccessible, but target states with small $\theta < \theta_L$ will require a longer path and therefore a longer pulse duration.

To summarize, reaching any arbitrary state requires varying either ϕ or Δ , but varying ϕ allows us to select an initial direction to move in Hilbert space, while varying Δ allows us to introduce a curve in the trajectory. In multi-qubit systems, the ideal trajectory is harder to characterize, since the optimal trajectory using the control fields accessible to us experimentally do not necessarily trace a geodesic in the full Hilbert space. However, the freedom to adjust phases will tend to allow us to select a more direct path, while adjusting frequency will tend to incur a longer path length which takes longer to traverse.

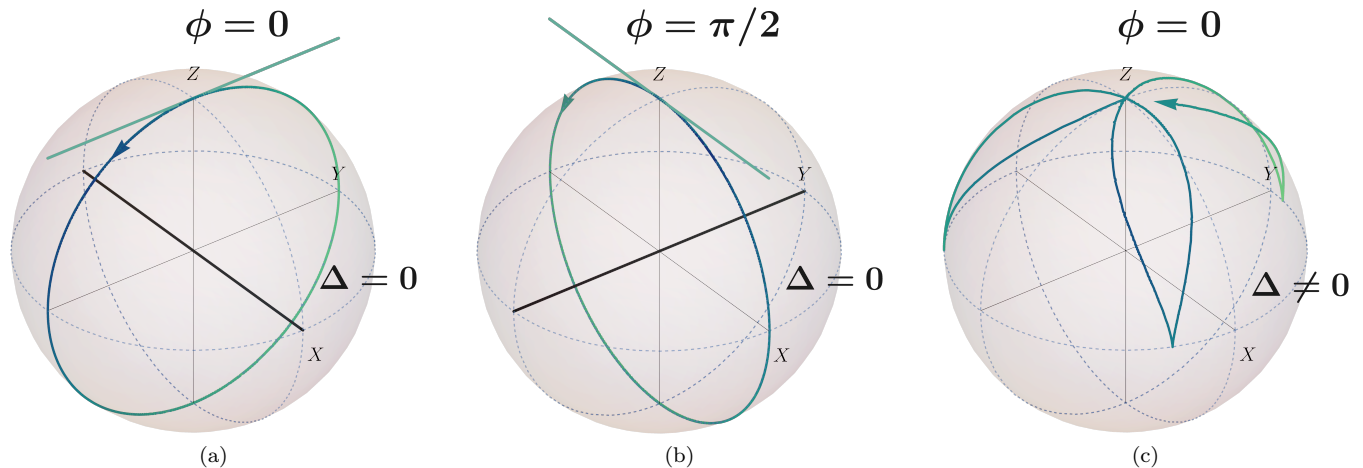


FIG. 6. Cartoon depictions of the path on a Bloch sphere traced by a single-qubit pulse with constant A , ϕ , and Δ , when starting from the north pole ($|0\rangle$). The path rotates around an axis with azimuthal orientation starting at ϕ and precessing at a rate Δ , and a vertical tilt given by the ratio A/Δ . When $\Delta = 0$ (6a and 6b), the axis is in the x - y plane, so the path is a meridian with longitude given by ϕ .

V. CONCLUSIONS

In this paper, we have explored the consequences of several different methods to parameterize a pulse-level ansatz for VQE experiments. For each parameterization, we analyzed how energy accuracy and optimization difficulty vary over pulse duration. Here, we review our most salient observations and summarize our main conclusions.

In every case, it is clear that ctrl-VQE requires a minimal pulse duration to have any chance of successfully preparing the desired eigenstate. While operating at this minimal time would be ideal for minimizing qubit decoherence, and it would seem to minimize runtime, we find that the optimization loop takes much longer to converge at and near this regime. Furthermore, even locating the MET (or eMET) would be intractable in practice. Consequently, a challenge arises for realistic simulations: how does one know *a priori* that the specified evolution time in fact exceeds the MET, and thus provides the possibility of preparing the target state? While answering this definitively would lie beyond the scope of this paper, we note that the quality of the final solution can be estimated by metrics such as variance [43] or Hamiltonian-reconstruction distance [44] to determine whether the optimization has located an eigenstate.

In Section IV A, we explored the consequences of restricting pulse parameters from continuous functions in time to discrete windows. So long as individual windows are not too long, we found that ctrl-VQE provides essentially consistent results, indicating that digitization of the analogue signals should not have a significant impact on the performance. The minimal evolution time required to obtain high-accuracy ansätze increases marginally as the number of free parameters is reduced, but the far more significant difference is that optimization is harder

when the parameterization better approximates continuously variable pulses. In some sense, this is a trivial statement: optimization with fewer parameters is easier. Nevertheless, it has important practical significance, since the runtime of ctrl-VQE is ultimately bounded by the number of optimization iterations.

Similarly, in Section IV B, we tested parameterizations for a complex amplitude represented in both polar ($\Omega = Ae^{i\phi}$) and Cartesian ($\Omega = \alpha + i\beta$) coordinates. These two choices are physically identical; the only difference between them is the trajectory through parameter space which the optimizer selects to find optimal pulses. This is analogous to the situation in classical molecular geometry optimization, where different coordinate representations (z -matrix, redundant internals, Cartesians, etc.) all specify the same molecule, but lead to widely varying iteration counts for optimization. [45] We find that the polar parameterization turns out to be significantly more expensive to converge, especially in the region near the minimal evolution time. Therefore, we recommend using Cartesian coordinates for experimental implementations of ctrl-VQE.

Our most important results are found in Section IV C, where we contrast parameterizations which allow or disallow the phase and the frequency to vary in the optimization. We find that the minimal evolution time is longest when only amplitudes are varied, that it is improved when amplitudes and frequencies are varied, and that it is improved even more when amplitudes and phases are varied. Counter-intuitively, we find *no* measurable improvement to the minimal evolution time when all three parameters are varied, although the optimization is disproportionately harder. These results suggest that, in a transmon-based ctrl-VQE, it is preferable to fix the drive frequencies of each pulse (perhaps at the associated resonance frequency) and optimize over amplitudes

and phases.

Since single-qubit gates are often implemented with resonant pulses, it may be surprising that resonant pulses are capable of building the entanglement needed to represent the ground state of a molecular system. However, in an architecture with static couplings, pulses applied locally in the “qubit” basis will nevertheless have a global effect in the eigenbasis of the device. In some sense, entanglement builds spontaneously, irrespective of pulse frequency; the pulses serve only to direct it in the appropriate direction and drive it along faster. Our findings imply that the phase of control fields is the more effective parameter to provide that direction. Modulating this parameter in time is also more natural in fixed-frequency, fixed-coupling devices, such as those presently used by IBMQ [46]. An interesting direction for future research is to investigate whether similar or complementary heuristics can be found for alternate architectures, such as those with tunable couplers, or those where the

drive field is applied globally to all qubits.

ACKNOWLEDGMENTS

The authors thank Chenxu Liu, Ayush Asthana, and Christopher Long for edifying conversations. This research was supported by the U.S. Department of Energy (DOE) Award No. **DE-SC0019199**, by the National Science Foundation (Grant No. 2137776). S.E.E. acknowledges support from the DOE Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under contract number **DE-SC0012704**.

VI. SUPPLEMENTARY INFORMATION

Our latest implementation of ctrl-VQE can be found at [32].

-
- [1] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, Quantum Algorithms for Quantum Chemistry and Quantum Materials Science, *Chemical Reviews* **120**, 12685 (2020), publisher: American Chemical Society.
- [2] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, *Reviews of Modern Physics* **92**, 015003 (2020).
- [3] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya, *et al.*, Quantum chemistry in the age of quantum computing, *Chemical reviews* **119**, 10856 (2019).
- [4] D. A. Fedorov, B. Peng, N. Govind, and Y. Alexeev, Vqe method: A short survey and recent developments, *Materials Theory* **6**, 1 (2022).
- [5] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, The Variational Quantum Eigensolver: A review of methods and best practices, *Physics Reports* **986**, 1 (2022).
- [6] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Variational quantum algorithms, *Nature Reviews Physics* , 1 (2021).
- [7] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum processor, *Nature communications* **5**, 1 (2014).
- [8] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, *Nature* **549**, 242 (2017).
- [9] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, Generalized unitary coupled cluster wave functions for quantum computation, *Journal of chemical theory and computation* **15**, 311 (2018).
- [10] I. G. Ryabinkin, T.-C. Yen, S. N. Genin, and A. F. Izmaylov, Qubit coupled cluster method: a systematic approach to quantum chemistry on a quantum computer, *Journal of chemical theory and computation* **14**, 6317 (2018).
- [11] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes, Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm, *npj Quantum Information* **6**, 1 (2020).
- [12] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, *Nature communications* **10**, 1 (2019).
- [13] Y. S. Yordanov, V. Armaos, C. H. W. Barnes, and D. R. M. Arvidsson-Shukur, Qubit-excitation-based adaptive variational quantum eigensolver, *Communications Physics* **4**, 1 (2021), publisher: Nature Publishing Group.
- [14] H. L. Tang, V. Shkolnikov, G. S. Barron, H. R. Grimsley, N. J. Mayhall, E. Barnes, and S. E. Economou, qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor, *PRX Quantum* **2**, 020310 (2021).
- [15] O. R. Meitei, B. T. Gard, G. S. Barron, D. P. Pappas, S. E. Economou, E. Barnes, and N. J. Mayhall, Gate-free state preparation for fast variational quantum eigensolver simulations, *npj Quantum Information* **7**, 155 (2021).
- [16] A. Asthana, C. Liu, O. R. Meitei, S. E. Economou, E. Barnes, and N. J. Mayhall, Leakage Reduces Device Coherence Demands for Pulse-Level Molecular Simulations, *Physical Review Applied* **19**, 064071 (2023), publisher: American Physical Society.
- [17] S. Deffner and S. Campbell, Quantum speed limits: from heisenberg’s uncertainty principle to optimal quantum control, *Journal of Physics A: Mathematical and The-*

- oretical **50**, 453001 (2017).
- [18] D. Liberzon, *Calculus of variations and optimal control theory* (Princeton university press, 2011).
- [19] Q. Ansel, E. Dionis, F. Arrouas, B. Peaudecerf, S. Guérin, D. Guéry-Odelin, and D. Sugny, Introduction to theoretical and experimental aspects of quantum optimal control (2024), arXiv:2403.00532 [quant-ph].
- [20] A. B. Magann, C. Arenz, M. D. Grace, T.-S. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms, *PRX Quantum* **2**, 010101 (2021).
- [21] J. Leng, Y. Peng, Y.-L. Qiao, M. Lin, and X. Wu, Differentiable Analog Quantum Computing for Optimization and Control, in *Advances in Neural Information Processing Systems*, Vol. 35, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., 2022) pp. 4707–4721, arXiv: 2210.15812.
- [22] D. Meirom and S. H. Frankel, PANSATZ: pulse-based ansatz for variational quantum algorithms, *Frontiers in Quantum Science and Technology* **2**, 10.3389/frqst.2023.1273581 (2023), publisher: Frontiers.
- [23] Z. Liang, J. Cheng, Z. Song, H. Ren, R. Yang, H. Wang, K. Liu, P. Kogge, T. Li, Y. Ding, and Y. Shi, Towards Advantages of Parametrized Quantum Pulses (2023), arXiv:2304.09253 [quant-ph].
- [24] D. J. Egger, C. Capecci, B. Pokharel, P. K. Barkoutsos, L. E. Fischer, L. Guidoni, and I. Tavernelli, Pulse variational quantum eigensolver on cross-resonance-based hardware, *Physical Review Research* **5**, 033159 (2023), publisher: American Physical Society.
- [25] K. Kottmann and N. Killoran, Evaluating analytic gradients of pulse programs on quantum computers (2023), arXiv:2309.16756 [quant-ph].
- [26] R. d. Keijzer, O. Tse, and S. Kokkelmans, Pulse based Variational Quantum Optimal Control for hybrid quantum computing, *Quantum* **7**, 908 (2023), publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.
- [27] N. Entin, M. M. Roses, R. Cohen, N. Katz, and A. Makmal, Molecular groundstate determination via short pulses on superconducting qubits (2024), arXiv:2403.17789 [quant-ph].
- [28] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, Pyscf: the python-based simulations of chemistry framework, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **8**, e1340 (2018).
- [29] Qiskit contributors, Qiskit: An open-source framework for quantum computing (2023).
- [30] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, Tapering off qubits to simulate fermionic hamiltonians (2017), arXiv:1701.08213 [quant-ph].
- [31] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A Fresh Approach to Numerical Computing, *SIAM Review* **59**, 65 (2017), publisher: Society for Industrial and Applied Mathematics.
- [32] K. M. Sherbert and N. J. Mayhall, Ctrlvqe.jl, <https://github.com/mayhallgroup/CtrlVQE.jl> (2023).
- [33] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, A quantum engineer’s guide to superconducting qubits, *Applied Physics Reviews* **6**, 10.1063/1.5089550 (2019).
- [34] T. E. Roth, R. Ma, and W. C. Chew, The transmon qubit for electromagnetics engineers: An introduction, *IEEE Antennas and Propagation Magazine* **65**, 8–20 (2023), conference Name: IEEE Antennas and Propagation Magazine.
- [35] P. K. Mogensen and A. N. Riseth, Optim: A mathematical optimization package for Julia, *Journal of Open Source Software* **3**, 615 (2018).
- [36] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms, *Journal of Magnetic Resonance* **172**, 296 (2005).
- [37] F. K. Wilhelm, S. Kirchhoff, S. Machnes, N. Wittler, and D. Sugny, An introduction into optimal control for quantum technologies (2020), arXiv:2003.10132 [quant-ph].
- [38] Z.-C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon, Optimizing variational quantum algorithms using pontryagin’s minimum principle, *Physical Review X* **7**, 021027 (2017).
- [39] T. Haug, K. Bharti, and M. Kim, Capacity and Quantum Geometry of Parametrized Quantum Circuits, *PRX Quantum* **2**, 040309 (2021), publisher: American Physical Society.
- [40] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum Natural Gradient, *Quantum* **4**, 269 (2020), publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.
- [41] L. Funcke, T. Hartung, K. Jansen, S. Kühn, and P. Stornati, Dimensional expressivity analysis of parametric quantum circuits, *Quantum* **5**, 422 (2021), publisher: Verein zur Förderung des Open Access Publizierens in den Quantenwissenschaften.
- [42] J. Townsend, *A Modern Approach to Quantum Mechanics*, International series in pure and applied physics (University Science Books, 2000).
- [43] D.-B. Zhang, B.-L. Chen, Z.-H. Yuan, and T. Yin, Variational quantum eigensolvers by variance minimization, *Chinese Physics B* **31**, 120301 (2022), publisher: Chinese Physical Society and IOP Publishing Ltd.
- [44] L. J. I. Moon, M. M. Sohoni, M. A. Shimizu, P. Viswanathan, K. Zhang, E.-A. Kim, and P. L. McMahon, Hamiltonian-reconstruction distance as a success metric for the Variational Quantum Eigensolver (2024), arXiv:2403.11995 [quant-ph].
- [45] J. Baker, A. Kessi, and B. Delley, The generation and use of delocalized internal coordinates in geometry optimization, *The Journal of Chemical Physics* **105**, 192 (1996), https://pubs.aip.org/aip/jcp/article-pdf/105/1/192/19076673/192_1_online.pdf.
- [46] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient ZZ gates for quantum computing, *Physical Review A* **96**, 022330 (2017), publisher: American Physical Society.

Appendix A: Evaluating Analytical Gradients with Complex Amplitudes

In this appendix, we derive an expression for the analytic gradient of the expectation value $\langle \psi | \hat{O} | \psi \rangle$ with respect to our control parameters θ , where $|\psi\rangle = |\psi(\theta)\rangle$ is

our fully-evolved state and \hat{O} is a Hermitian observable. This treatment is an adaptation of GRAPE [36], tailored to our particular problem.

Formally, the objective function takes the form:

$$E(\boldsymbol{\theta}) = \langle \psi_R(\boldsymbol{\theta}) | \hat{O} | \psi_R(\boldsymbol{\theta}) \rangle \quad (\text{A1})$$

By using a subscript R in $|\psi_R(\boldsymbol{\theta})\rangle$, we denote that \hat{O} is measured in the rotating frame.

The variational ansatz takes the form:

$$|\psi_R(\boldsymbol{\theta})\rangle = U_R(\boldsymbol{\theta}) |\psi_0\rangle \quad (\text{A2})$$

where $U_R(\boldsymbol{\theta})$ is the unitary operator describing evolution under the device Hamiltonian:

$$U_R(\boldsymbol{\theta}) = \hat{\mathcal{T}} \exp\left(-i \int_0^T dt \hat{V}_R(\boldsymbol{\theta}, t)\right) \quad (\text{A3})$$

The device Hamiltonian itself (in the rotating frame) takes the following form:

$$\hat{V}_R(\boldsymbol{\theta}, t) = e^{it\hat{H}_0} \hat{V}(\boldsymbol{\theta}, t) e^{-it\hat{H}_0} \quad (\text{A4})$$

The static component \hat{H}_0 does not depend on our control parameters, so it need not be specified further in this derivation. The drive component \hat{V} has the form:

$$\hat{V}(\boldsymbol{\theta}, t) = \sum_q \hat{V}_q(\boldsymbol{\theta}, t) = \sum_q \left(\Omega_q(\boldsymbol{\theta}, t) e^{i\nu_q(\boldsymbol{\theta}, t)t} \hat{a}_q + \text{h.c.} \right) \quad (\text{A5})$$

Recall that Ω_q is a complex number, encapsulating both the amplitude and phase of the drive pulse. It will be helpful to recast the drive component into a sum of Hermitian time-independent operators with real scalar time-dependent coefficients:

$$\hat{V}(\boldsymbol{\theta}, t) = \sum_q \left(x_q(\boldsymbol{\theta}, t) \hat{Q}_q + y_q(\boldsymbol{\theta}, t) \hat{P}_q \right) \quad (\text{A6})$$

Here $\hat{Q}_q \equiv \hat{a}_q + \hat{a}_q^\dagger$ and $\hat{P}_q \equiv i(\hat{a}_q - \hat{a}_q^\dagger)$ are the canonical coordinate and momentum operators, equivalent to Pauli- X and Y spin operators in the two-level truncation (except that our definition of \hat{P} absorbs an extra negative sign for the sake of notational convenience). The scalar functions x_q and y_q are the real and imaginary parts of $\Omega_q \exp(i\nu_q t)$.

The simplest approach to characterizing time-dependent functions is to discretize time. Therefore, let us define a time-interval $\tau \equiv T/r$, where r should be understood as arbitrarily large. We will abbreviate all parametric time-dependent $f(\boldsymbol{\theta}, t)$ with the shorthand:

$$f_i \equiv f(\boldsymbol{\theta}, i\tau) \quad (\text{A7})$$

Now we choose to rewrite the integral in the exponential of U_R as a trapezoidal sum.

$$\int_0^T dt \hat{V}_R(\boldsymbol{\theta}, t) \approx \tau \sum_{j=1}^r \frac{\hat{V}_{R,j} + \hat{V}_{R,j-1}}{2} \quad (\text{A8})$$

Furthermore, since V_R is a conjugation of the drive component \hat{V} with respect to a unitary frame rotation $\exp(-it\hat{H}_0)$, this factor can be brought *outside* the exponentiation $\exp(-it\hat{V}_R)$:

$$e^{-i\tau\hat{V}_{R,i}} = e^{it\hat{H}_0} e^{-i\tau\hat{V}_i} e^{-it\hat{H}_0} \quad (\text{A9})$$

We may now expand U_R as a time-ordered product. At the same time, we will introduce left and right unitaries $U_{Ri}^{(l)}, U_{Ri}^{(r)}$ symmetrically about a single time point $t = i\tau$:

$$U_R \equiv U_{Ri}^{(l)} U_{Ri}^{(r)} \quad \text{for any } i \in 0..r \quad (\text{A10})$$

$$U_{Ri}^{(l)} \approx e^{iT\hat{H}_0} \prod_{j=i+1}^r \left(e^{-i\tau\hat{V}_j/2} e^{-i\tau\hat{H}_0} e^{-i\tau\hat{V}_{j-1}/2} \right) \quad (\text{A11})$$

$$U_{Ri}^{(r)} \approx \prod_{j=1}^i \left(e^{-i\tau\hat{V}_j/2} e^{-i\tau\hat{H}_0} e^{-i\tau\hat{V}_{j-1}/2} \right) \quad (\text{A12})$$

In order to take the gradient with respect to control parameters $\boldsymbol{\theta}$, let us first consider partial derivatives with respect to x_{qi} and y_{qi} . We will begin by writing $\hat{V}_i = \hat{\Sigma} + x_{qi} \hat{Q}_q$, where $\hat{\Sigma}$ represents all the terms in \hat{V}_i except $x_{qi} \hat{Q}_q$. The analytical expression for the derivative of the exponential of \hat{V}_i is [36]:

$$\partial_{x_{qi}} e^{-i\tau(\hat{\Sigma} + x_{qi} \hat{Q}_q)} = e^{-i\tau\hat{\Sigma}} \int_0^1 ds e^{-i\tau s \hat{\Sigma}} \left(-i\tau \hat{Q}_q \right) e^{i\tau s \hat{\Sigma}} \quad (\text{A13})$$

Because \hat{Q}_q and \hat{P}_q do not commute (in the two-level truncation, the commutator is not even scalar), this has no closed-form solution. However, so long as τ is very small, corrections due to the commutator are negligible. Therefore, we choose to adopt the following much simpler, approximate expressions:

$$\partial_{x_{qi}} U_{Ri}^{(l)} \approx U_{Ri}^{(l)} \left(-i\tau \hat{Q}_q / 2 \right) \quad (\text{A14})$$

$$\partial_{x_{qi}} U_{Ri}^{(r)} \approx \left(-i\tau \hat{Q}_q / 2 \right) U_{Ri}^{(r)} \quad (\text{A15})$$

$$\partial_{x_{qi}} U_R \approx -i\tau U_{Ri}^{(l)} \hat{Q}_q U_{Ri}^{(r)} \quad (\text{A16})$$

Similarly,

$$\partial_{y_{qi}} U_R \approx -i\tau U_{Ri}^{(l)} \hat{P}_q U_{Ri}^{(r)} \quad (\text{A17})$$

Note that the symmetric form of our trapezoidal time integration serves to further enhance the accuracy of the approximation.

We now define the ‘‘gradient signals’’ $\phi_{qi}^{(x)}, \phi_{qi}^{(y)}$:

$$\phi_{qi}^{(x)} = \text{Im} \langle \psi_0 | U_R^\dagger \hat{O} U_{Ri}^{(l)} \hat{Q}_q U_{Ri}^{(r)} | \psi_0 \rangle \quad (\text{A18})$$

$$\phi_{qi}^{(y)} = \text{Im} \langle \psi_0 | U_R^\dagger \hat{O} U_{Ri}^{(l)} \hat{P}_q U_{Ri}^{(r)} | \psi_0 \rangle \quad (\text{A19})$$

These can be computed numerically in a time comparable to that of simulating time-evolution itself. It is

straightforward to show that the gradient of the objective function E with respect to x and y is:

$$\partial_{x_{qi}} E = 2\tau \phi_{qi}^{(x)} \quad (\text{A20})$$

$$\partial_{y_{qi}} E = 2\tau \phi_{qi}^{(y)} \quad (\text{A21})$$

Finally, a simple chain rule gives us the expression for the energy gradient with respect to our control parameters $\boldsymbol{\theta}$:

$$\partial_{\theta_k} E = 2\tau \sum_{q,i} \left[(\partial_{\theta_k} x_{qi}) \cdot \phi_{qi}^{(x)} + (\partial_{\theta_k} y_{qi}) \cdot \phi_{qi}^{(y)} \right] \quad (\text{A22})$$

We note that the above formula is in effect a discretized

integral over time, and accuracy can be tuned by adopting any quadrature rule (e.g., the trapezoidal rule used in Eq. (A8)). We conclude by explicitly writing the integral formulation for continuous time:

$$\partial_{\theta_k} E = 2 \sum_q \int_0^T dt \left[\frac{\partial x_q(\boldsymbol{\theta}, t)}{\partial \theta_k} \cdot \phi_q^{(x)}(t) + \frac{\partial y_q(\boldsymbol{\theta}, t)}{\partial \theta_k} \cdot \phi_q^{(y)}(t) \right] \quad (\text{A23})$$