

## Efficient quantum computation of molecular forces and other energy gradients

Thomas E. O'Brien<sup>1,\*</sup> Michael Streif<sup>2,†</sup> Nicholas C. Rubin<sup>1,‡</sup> Raffaele Santagati<sup>2,§</sup> Yuan Su<sup>1</sup> William J. Huggins<sup>1</sup> Joshua J. Goings<sup>1</sup> Nikolaj Moll<sup>2</sup> Elica Kyoseva<sup>2</sup> Matthias Degroote<sup>2</sup> Christofer S. Tautermann<sup>3</sup> Joonho Lee<sup>1,4</sup> Dominic W. Berry<sup>5</sup> Nathan Wiebe<sup>6,7,||</sup> and Ryan Babbush<sup>1,¶</sup>

<sup>1</sup>Google Research, Venice, California 90291, USA

<sup>2</sup>Quantum Lab, Boehringer Ingelheim, 55218 Ingelheim am Rhein, Germany

<sup>3</sup>Boehringer Ingelheim Pharma GmbH & Co KG, Birkendorfer Strasse 65, 88397 Biberach, Germany

<sup>4</sup>Department of Chemistry, Columbia University, New York, USA

<sup>5</sup>School of Mathematical and Physical Sciences, Macquarie University, New South Wales 2109, Australia

<sup>6</sup>Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4

<sup>7</sup>Pacific Northwest National Laboratory, Richland, Washington 99354, USA



(Received 21 December 2021; accepted 13 October 2022; published 26 December 2022)

While most work on the quantum simulation of chemistry has focused on computing energy surfaces, a similarly important application requiring subtly different algorithms is the computation of energy derivatives. Almost all molecular properties can be expressed as an energy derivative, including molecular forces, which are essential for applications such as molecular dynamics simulations. Here, we introduce new quantum algorithms for computing molecular energy derivatives with significantly lower complexity than prior methods. Under cost models appropriate for noisy-intermediate scale quantum devices, we demonstrate how low-rank factorization and other tomography schemes can be optimized for energy derivative calculations. We numerically demonstrate that our techniques reduce the number of circuit repetitions required by many orders of magnitude for even modest systems, and that the cost of estimating an entire force vector may in some systems be lower than the cost of estimating the energy. In the context of fault-tolerant algorithms, we develop new methods of estimating energy derivatives with Heisenberg limited scaling, incorporating state-of-the-art techniques for block encoding fermionic operators. In contrast to our near-term results, we find that the cost of estimating forces with any of our Heisenberg-limited methods is bounded by the cost of estimating energies, due to inner loops requiring either energy estimation or reflections around the ground state. This implies that applications such as geometry optimization, coupling parameter estimation, and spectral prediction may be practical on fault-tolerant quantum devices, but tractable molecular dynamics simulations of large-scale systems requires further algorithmic advances.

DOI: [10.1103/PhysRevResearch.4.043210](https://doi.org/10.1103/PhysRevResearch.4.043210)

### I. INTRODUCTION

Quantum chemistry is widely regarded as one of the most promising areas of application for quantum computers. This is due to the relative ease of mapping the electronic structure problem onto a quantum device [1–3], its difficulty in simulating classically, and its high relevance to industry. Interest in such applications has been steadily increasing following

initial beyond-classical quantum computing demonstrations [4,5] and experimental [6–8] demonstrations of hardware near the fault-tolerant threshold for quantum error correction [9]. A significant body of work has emerged in recent years optimizing quantum algorithms for chemistry, both for fault-tolerant quantum computers [10–16] and current NISQ devices [17–22], including various experimental implementations [23–29]. Predominantly this work has focused on estimating energies of ground states of the electronic structure problem, perhaps the most natural property to extract from a quantum chemistry simulation. However, ground-state energies are not a quantity typically measured in the laboratory, and further processing of energy data is required to obtain properties of relevance to industry. Thus, quantum algorithms to estimate properties other than ground-state energies are of high interest as we progress towards larger NISQ or future fault-tolerant devices.

The calculation of forces (the derivative of energies with respect to nuclear positions) is a subroutine in most modern computational approaches to navigate molecular potential energy surfaces. Beyond identifying minima and other

\*teobrien@google.com

†michael.streif@boehringer-ingelheim.com

‡nickrubin@google.com

§raffaele.santagati@boehringer-ingelheim.com

||nawiebe@cs.toronto.edu

¶babbush@google.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

stationary points, reaction path following is also based on the determination of gradients [30]. The knowledge of low-energy stationary points allows the generation of conformational Boltzmann ensembles, calculation of reaction rates, and prediction of tautomer equilibria. Forces are also an essential ingredient of molecular dynamics (MD) simulations, which are invaluable for studying macroscopic thermodynamic properties. This covers highly diverse applications such as the description of heterogeneous processes on surfaces including catalysis [31], observation of phase transitions, such as nucleation processes for water [32], and maybe of highest importance for pharmaceutical research, the interaction of drugs with their targets in the human body [33]. By free energy calculations based on MD simulations, these interactions can be quantified, allowing the prediction of compound affinities via classical [34] or proposed quantum [35,36] methods, which are eventually linked to therapeutic doses. Beyond that, MD simulations of drug-target systems enable the observation of conformational changes of the target. The corresponding drug-induced active and inactive states or ligand bias is a ligand-dependent selective signaling pattern [37], which is especially useful to avoid drug-induced side effects [38]. Other decisive parameters such as drug residence time can nowadays be determined through specialized MD simulations [39]. These powerful techniques render MD simulations one of the most broadly used and powerful tools in drug design, and hence make forces a clear target for quantum computing.

Although some research on quantum algorithms for force and gradient estimation has been performed previously, efforts to accurately cost algorithms in a fault-tolerant or NISQ setting have been limited. The suggestion to estimate nuclear forces on a quantum device was first suggested by [40], which studied estimation via the Hellman-Feynman theorem and via the quantum gradient estimation algorithm of [41]. This topic was then relatively untouched by the quantum community for a decade until it was revived by [42–44]. Reference [42] studied force estimation in both a NISQ and FT framework and performed the first experimental force calculation, but only found loose asymptotic bounds of  $N^7 - N^{15}$  to estimate a single force component. Reference [43] put the mathematical formulation of force estimation in NISQ on a significantly stronger footing, combined this with gradient estimation for the optimization of variational quantum eigensolvers, but only considered the cost of estimating all  $N^4$  terms in the fermionic 2-reduced density matrix to constant precision. Reference [44] firmed the theoretical chemistry behind force estimation on a quantum device, presenting a detailed derivation in a Lagrangian formalism focusing on an *ab initio* exciton model, and stressed the importance of including full response. The paper presented explicit formulas and circuits based on the parameter-shift rule [45,46] but did not provide asymptotic costs for the estimation of forces on a quantum device. These works were followed by small experimental demonstrations of molecular dynamics simulations for various applications [47–49], and theoretical studies extending gradient calculations to the derivatives of energies beyond the ground state [50–52]. However, many possible optimizations remain for both NISQ and fault-tolerant algorithms to estimate forces. Furthermore, little work has been done to estimate the magni-

tude of force operator quantities that are relevant for quantum algorithm resource requirements (e.g., induced 1-norms).

In this paper we perform an in-depth study of the cost of estimating forces on a quantum computer using either NISQ or future fault-tolerant devices. For NISQ devices, we optimize and cost tomography methods, introducing the notion that one needs to estimate the number of experiments to measure the gradient vector to a target 2-norm as the relevant cost model for this regime. We introduce a NISQ tomography technique where single-shot preparations and measurements are allocated to optimize the 2-norm, and estimate the cost of previous tomography methods [19,20,53,54] in terms of the 2-norm as well. We analyze the cost of all NISQ tomography methods to estimate the entire force vector on hydrogen chains of up to 20 atoms, and give empirical and asymptotic scaling with the system size. We find that all studied methods have similar or even slightly better asymptotic costs to estimate an entire force vector to a given accuracy compared to the cost of estimating energies to a similar accuracy.

In a fault-tolerant setting, we construct and optimize three independent methods for force estimation: a finite difference algorithm, a direct expectation value estimation algorithm (based on the overlap estimation algorithm of [55]), and an extension of the new gradient-based expectation value estimation algorithm of [56]. As part of this, we design several quantum subroutines: to block-encode force operators in first quantization; to recycle system registers between separate force optimizations (with unit fidelity); to optimize allocation of resources in the classical outer loop of a higher-order finite difference estimation routine; and to replace state preparation with a reflection around the ground state in the gradient-based expectation value estimation algorithm. We study the cost of block-encoding force operators in first and second quantization under different encodings, proving that in a plane-wave basis all force operators are simultaneously diagonalized by the fermionic fast Fourier transform, and that in an atomic orbital basis the rank of the tensor-hypercontracted derivative operator can be bounded by a constant factor times the rank of the tensor-hypercontracted Hamiltonian. Numerically, we find that for state-of-the-art techniques block encodings of derivatives are at most constant or polylog factors more costly than block-encodings of the corresponding Hamiltonian, and that in practice they may be significantly cheaper. We determine asymptotic costings for our three fault-tolerant force estimation algorithms on hydrogen chains and water clusters, and for plane-wave systems in first quantization (as studied in [57,58]). Surprisingly, due to difficulties to parallelize the overlap estimation algorithm and the need to perform Hamiltonian simulation as part of the reflection subroutine, we find that in some cases the finite difference method will be preferable (or at worst competitive) compared to the gradient-based expectation value estimation algorithm, which strictly asymptotically dominates the overlap estimation algorithm.

Finally, we consider the cost of applying our force-estimation algorithms to real applications in quantum chemistry. We study the required tolerance on the error in a force estimation for molecular dynamics and geometry optimization, finding a relevant figure for accurate estimation of the pair correlation function of a moderate-sized water simulation being a root mean square (RMS) error of no more than

6.4 mHa/Å per single derivative component. For applications such as geometry optimization, coupling parameter estimation or spectral prediction, which do not require such a high number of repeat derivative estimations, our methods appear feasible for early fault-tolerant devices to solve problems beyond the reach of classical computers. However, we do not see a useful beyond-classical molecular dynamics simulation to be tractable in a NISQ or FT quantum computing setting using our methods, due to the large number of repeated derivative estimations required.

### A. Outline

We begin this paper in Sec. II with a review of *ab initio* electronic structure theory and how energy derivatives may be estimated as the expectation value of a derivative operator through the Hellman-Feynman theorem. Although we focus on atomic forces (i.e., derivatives with respect to nuclei positions) for the majority of this paper, we briefly detail here how these methods may be immediately extended to other first-order properties of a molecular system. In Sec. II A we present a simple, calculable derivation of the force operator in second quantization for an atomic-centered basis orbital set, based on the orbital connection theory of Helgaker and Almlöf [59]. Then, in Sec. II B we derive the exact form of the force operator in a plane wave basis in first and second quantization, and demonstrate that this operator is diagonalized by the quantum Fourier transform with aliased frequencies and the fermionic fast Fourier transform respectively. In Sec. II C, we estimate the error tolerance on a force vector required for geometry optimization and molecular dynamics simulations. Based on radial distribution function calculations for a system of 216 water molecules, we estimate that it is relevant for molecular dynamics and geometry optimization to target a RMS of the error in one force component below 0.6 mHa/Å.

We then turn in Sec. III to the optimization and costing of state tomography for the estimation of force vectors in molecular systems. We overview a general scheme for low-cost NISQ tomography methods of arbitrary operators, and in Sec. III A we review previous work on choices of basis rotation to implement this scheme. In Sec. III B, we extend previous work on importance sampling to directly target the 2-norm error in a force vector, and demonstrate the importance of parallelization of measurements where possible. In Sec. III C, we review the fermionic shadow tomography scheme of Refs. [54,60], and calculate the relevant bound on the cost of the number of measurements to estimate a constant 2-norm error here as well. Then, in Sec. III D, we find bounds on the costs of the different methods above, both analytically and numerically for hydrogen chains, and estimate the asymptotic costing of each.

A critical piece of fault-tolerant quantum computation is block encoding, so before giving fault-tolerant algorithms for force estimation we study the cost of block-encoding a force operator. We review general block encodings in Sec. IV A. We give explicit methods to block-encode Hamiltonians and force operators in second quantization for factorized methods (Sec. IV B), and in first quantization for plane waves (Sec. IV C). In molecular systems the cost of simulating

block-encoded derivative operators is found to be at most a constant factor worse (as one may differentiate the Hamiltonian in its factorized form), while in plane-wave systems the rescaling factor of the block encodings is found to be identical and the circuit cost only poly( $\log(N)$ ) worse. In Sec. IV D we study the rescaling factors for block encoding in atomic orbital bases, finding that when using sparse simulation methods the cost of simulating forces is similar to the cost of simulating Hamiltonians, but for factorized methods the cost is clearly asymptotically lower.

We finish this paper in Sec. V by designing three new algorithms for force estimation on fault-tolerant quantum computers, and estimating their asymptotic costs on various chemical systems using results from previous sections. In Sec. V A we use higher-order difference formulas to estimate gradients with a fault-tolerant quantum computer as a subroutine to estimate the energy at different atomic configurations. We optimize the importance sampling, choice of finite difference order and step size, and consider the efficiency of reusing the state register on the quantum device between different calls to the subroutine. In Sec. V B, we use the overlap estimation algorithm of [55] to estimate gradients via the Hellman-Feynman theorem at the Heisenberg limit. We optimize this algorithm for general block-encoded Hermitian operators by a factor 4, and optimize importance sampling over the gradient terms. We further optimize the choice of reflection operator using techniques from [61], and demonstrate the ability to perfectly recycle the state register on the quantum device between calls to the amplitude estimation subroutine. However, due to the need to reflect about the ground state (which requires Hamiltonian simulation), we find that the overlap estimation algorithm can only outperform finite difference estimation when state preparation is the dominant cost of estimation in both routines. Finally, we implement force estimation using the new gradient estimation technique of [56], and compare it to both previous methods. We find it achieves a strict asymptotic improvement over the overlap estimation algorithm, which implies in turn that it may often be better than a semiclassical finite difference method. We conclude in Sec. VI, where we summarize our results, discuss the implications for the field of molecular dynamics, and suggest paths for further improvement. A summary of the results found in this section can be found in Table I.

## II. ENERGY DERIVATIVE CALCULATION IN ELECTRONIC STRUCTURE

The goal of *ab initio* electronic structure theory is to solve the time-independent Schrödinger equation,

$$H_{\text{tot}}|\Psi\rangle = E_{\text{tot}}|\Psi\rangle, \quad (1)$$

for a given molecular system. In most applications, it is sufficient to consider the nonrelativistic, time-independent molecular Hamiltonian  $H_{\text{tot}}$ , as we do here. Moreover, within the context of the Born-Oppenheimer approximation, one only needs to solve for the electronic Hamiltonian  $H$ , given

TABLE I. Summary of scalings of fault-tolerant algorithms for gradient vector estimation, taken from Tables IV, V, and IV. Here,  $N_e$  is the number of electrons,  $N$  the number of orbitals,  $\epsilon$  the error as measured by the 2-norm of the gradient vector, and  $\mathbb{E}(|Z_i|)$  is the expectation value of the nuclear charge. The ranges given for molecular system summarize the systems studied in this paper.

Method	System	Scaling
Finite differences	First quantized plane waves	$\tilde{\mathcal{O}}(N^{2/3}N_a^{17/6}(\mathbb{E}( Z_i ))^{1/3}\epsilon^{-1})$
Expectation-value estimation algorithm	First quantized plane waves	$\tilde{\mathcal{O}}(N^{4/3}N_a^{23/6}(\mathbb{E}( Z_i ))^{4/3}\epsilon^{-1})$
Gradient-based expectation value estimation algorithm	First quantized plane waves	$\tilde{\mathcal{O}}(N^{4/3}N_a^{10/3}(\mathbb{E}( Z_i ))^{4/3}\epsilon^{-1})$
Finite differences	Molecular systems	$\tilde{\mathcal{O}}(N_a^{3.83}\epsilon^{-1}) - \tilde{\mathcal{O}}(N_a^{4.42}\epsilon^{-1})$
Expectation-value estimation algorithm	Molecular systems	$\tilde{\mathcal{O}}(N_a^{4.11}\epsilon^{-1}) - \tilde{\mathcal{O}}(N_a^{4.74}\epsilon^{-1})$
Gradient-based expectation value estimation algorithm	Molecular systems	$\tilde{\mathcal{O}}(N_a^{3.61}\epsilon^{-1}) - \tilde{\mathcal{O}}(N_a^{4.24}\epsilon^{-1})$

by

$$\begin{aligned}
 H &= -\frac{1}{2} \sum_i^\eta \nabla_i^2 - \sum_i^\eta \sum_A^{N_a} \frac{Z_A}{r_{iA}} + \sum_{i>j}^\eta \frac{1}{r_{ij}} \\
 &= \sum_i^\eta h_i(\mathbf{r}) + \sum_{i>j}^\eta \frac{1}{r_{ij}}, \quad (2)
 \end{aligned}$$

which defines the electronic Hamiltonian for a molecular system with  $N_a$  atomic nuclei and  $\eta$  electrons. The first term in Eq. (2),  $h(\mathbf{r}) = \sum_i^\eta h_i(\mathbf{r})$ , is the one-electron term, which is the sum of the electronic kinetic energy and the interaction energy of the electrons with the nuclei with  $\mathbf{r}$  being a position vector in real space, while the second term is the Coulomb two-electron interaction energy. Here  $\nabla^2$  is the Laplacian with respect to the electronic coordinates,  $Z_A$  is the  $A$ th nuclear charge, and  $r_{iA} = |\mathbf{r}_i - \mathbf{R}_A|$  and  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$  are the Euclidean distances between the  $i$ th electron and the  $A$ th nucleus, and between the  $i$ th and  $j$ th electrons, respectively.

The molecular electronic Schrödinger equation is a function of the electronic coordinates  $\mathbf{r}$  with a parametric dependence on the nuclear coordinates  $\mathbf{R}$ , e.g.,

$$H(\mathbf{r}, \mathbf{R})|\Psi(\mathbf{r}, \mathbf{R})\rangle = E(\mathbf{R})|\Psi(\mathbf{r}, \mathbf{R})\rangle. \quad (3)$$

The total energy within the Born-Oppenheimer approximation for fixed nuclear positions is given as (for simplicity we omit the explicit position dependence)

$$E_{\text{tot}} = E + V_{\text{nuc}} = E + \sum_{A>B}^{N_a} \frac{Z_A Z_B}{R_{AB}}, \quad (4)$$

where  $V_{\text{nuc}}$  is the nuclear-nuclear repulsion energy with  $R_{AB} = |\mathbf{R}_A - \mathbf{R}_B|$  the Euclidean distance between the  $A$ th and  $B$ th nucleus.

Although the above provides a basis for molecular quantum mechanics and is sufficient for computing molecular energies, it is desirable to also be able to compute different molecular properties. Time-independent molecular properties can be expressed as gradients of the *ab initio* electronic energy  $E$  with respect to a suitable perturbation. For example, the first derivative of the energy with respect to an external electric or magnetic field evaluated at zero field strength yields the electric and magnetic dipole moments, respectively. Further, the first derivative of the energy with respect to the nuclear spin (internal magnetic field) yields the hyperfine coupling

constants, which are important for multiple spectroscopy techniques such as nuclear magnetic resonance (NMR). Similarly, molecular forces are computed as gradients of the energy with respect to nuclear displacements. These molecular properties are summarized in Table II. Although higher-order and mixed derivatives of the energy lead to additional properties, herein we will focus our attention on first-order derivatives of the total energy. Obtaining analytic formulas for these gradients is a rich research area in classical quantum chemistry and we refer the interested reader to Refs. [62–64] for more background.

In this paper, we will analyze two separate classes of methods for computing energy gradients; computing via the Hellmann-Feynman theorem [65,66], and computing via higher-order finite difference techniques. The Hellmann-Feynman theorem relates the energy derivative to the expectation value of the derivative of the Hamiltonian with respect to that same parameter

$$\frac{dE}{dx} = \langle \Psi | \frac{dH}{dx} | \Psi \rangle. \quad (5)$$

Here,  $\Psi$  is a normalized eigenstate of the Hamiltonian  $H$ , and the lower case  $x$  represents a general parameter with respect to which derivatives are taken (e.g., a single nuclear coordinate  $R_i$ , an electric field  $\mathbf{E}$ , or another quantity in Table II). In practice, the process of calculating the correct total derivative of the Hamiltonian  $H$  can be challenging. All explicit and implicit dependencies for the derivative have to be accounted for. In the remainder of this section we detail the analytic form of these operators in second quantized atomic-centered basis sets, and in arbitrary plane wave basis sets. However, neither of these methods are necessary to implement finite difference calculations.

TABLE II. Examples of properties, which can be computed as gradients of the total energy.

Gradient	Perturbation	Property
$dE/d\mathbf{E}$	electric field	electric dipole moment
$dE/d\mathbf{B}$	magnetic field	magnetic dipole moment
$dE/d\mathbf{I}$	nuclear spin	hyperfine coupling constant
$dE/d\mathbf{R}$	nuclear displacement	nuclear forces

### A. Force operators in second quantization for atomic-centered basis orbitals

To obtain the force operators the Schrödinger equation, Eq. (3), needs to be solved. From the atomic-centered basis orbitals (AO) the Hartree-Fock approximation is typically invoked to obtain first a set of molecular orbitals (MO). However, these MOs yield no analytic form and depend on the set of AOs. Therefore, it is not straightforward to calculate a total derivative operator  $\frac{dH}{dx}$  to allow force calculations through Eq. (5). However, through the relations of the AOs and MOs to each other, the derivatives can be calculated by the orbital connection theory of Helgaker [59].

A Hamiltonian represented on a quantum computer is conceptually different from the Hamiltonian on a classical computer. The overlap integrals are all precomputed beforehand in the given MO basis. The wave function on the quantum computer only gives the coefficients of all possible determinants. On a classical computer the MO basis is typically part of the wavefunction and not part of the Hamiltonian itself.

We use the following notational conventions. Lower case italics  $\{p, q, r, s\}$  index general (either occupied or virtual) are used for the MOs. Lower case Greek letters  $\{\mu, \nu, \lambda, \sigma\}$  index are used for the AOs. To distinguish vectors and tensors from their elements, they will be written in a bold typeface.

After a Hartree-Fock computation, the  $\eta$ -electron wave function is represented as a single Slater determinant, that is, an antisymmetric product of spin orbitals  $\{\phi_p(\mathbf{r})\}$ . These spin orbitals are discretized over a set of basis functions  $\{\chi_\mu(\mathbf{r})\}$ , commonly Gaussian atomic orbitals or plane waves. The spin orbitals are expanded as

$$\phi_p(\mathbf{r}) = \sum_{\mu} C_{\mu p} \chi_{\mu}(\mathbf{r}) \quad (6)$$

where  $C_{\mu p}$  denotes an element of the molecular orbital (MO) coefficient matrix. Without loss of generality, we will only consider real-valued MO coefficients. The electronic Hamiltonian from Eq. (2) can be cast in matrix form in the atomic orbital basis, with one-body integrals represented as

$$h_{\mu\nu} = \int d\mathbf{r}_1 \chi_{\mu}^*(\mathbf{r}_1) h(\mathbf{r}_1) \chi_{\nu}(\mathbf{r}_1) = \langle \chi_{\mu} | h | \chi_{\nu} \rangle, \quad (7)$$

and two-body integrals

$$\begin{aligned} g_{\mu\nu\lambda\sigma} &= \iint d\mathbf{r}_1 d\mathbf{r}_2 \chi_{\mu}^*(\mathbf{r}_1) \chi_{\lambda}^*(\mathbf{r}_2) \frac{1}{r_{12}} \chi_{\nu}(\mathbf{r}_1) \chi_{\sigma}(\mathbf{r}_2) \\ &= \langle \chi_{\mu} \chi_{\lambda} | \frac{1}{r_{12}} | \chi_{\nu} \chi_{\sigma} \rangle. \end{aligned} \quad (8)$$

In the general case, the set of AO basis functions is not orthogonal. It is therefore necessary to consider their overlap matrix,

$$S_{\mu\nu} = \int d\mathbf{r}_1 \chi_{\mu}^*(\mathbf{r}_1) \chi_{\nu}(\mathbf{r}_1) = \langle \chi_{\mu} | \chi_{\nu} \rangle. \quad (9)$$

These three integrals and their total derivatives (the so-called “skeleton” or “core” derivative integrals) are the fundamental building blocks of the molecular gradients. Expressions for the total derivatives of these integrals with respect to an arbitrary parameter  $x$  have been derived elsewhere and may

be easily computed with most electronic structure software packages.

In the orthonormal MO basis, it is useful to introduce the *second quantization* formalism, which is developed in terms of fermionic creation (annihilation) operators  $a_p^\dagger$  ( $a_q$ ) that satisfy the anticommutation relations,  $\{a_p^\dagger, a_q^\dagger\} = \{a_p, a_q\} = 0$ , and  $\{a_p^\dagger, a_q\} = \delta_{pq}$ . In second quantization the electronic structure Hamiltonian Eq. (2) in the MO basis is then given by

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \sum_{pqrs} g_{pqrs} a_p^\dagger a_r^\dagger a_s a_q, \quad (10)$$

with one- and two-body terms in the MO basis

$$\begin{aligned} h_{pq} &= \sum_{\mu\nu} C_{\mu p} C_{\nu q} h_{\mu\nu}, \\ g_{pqrs} &= \sum_{\mu\nu\lambda\sigma} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s} g_{\mu\nu\lambda\sigma}. \end{aligned} \quad (11)$$

At times it is useful to consider the overlap matrix also in the MO basis, which is given by

$$S_{pq} = \sum_{\mu\nu} C_{\mu p} C_{\nu q} S_{\mu\nu}, \quad (12)$$

where the use of lower case italic and Greek subscripts distinguishes between the MO and AO representations, respectively. We note that the typical overlap matrix relation  $S_{pq} = \delta_{pq}$  only holds at the reference configuration.

For the molecular electronic Hamiltonian, the energy is given by

$$\begin{aligned} E &= \sum_{pq} \gamma_{pq} \sum_{\mu\nu} C_{\mu p} C_{\nu q} h_{\mu\nu} \\ &+ \sum_{pqrs} \Gamma_{pqrs} \sum_{\mu\nu\lambda\sigma} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s} g_{\mu\nu\lambda\sigma}, \end{aligned} \quad (13)$$

where  $\gamma_{pq}$  and  $\Gamma_{pqrs}$  are the matrix elements of the one- and two-body reduced density matrices (RDMs) respectively.

Energies from *ab initio* calculations depend on several parameters: the one- and two-body AO integrals  $\mathbf{h}$  and  $\mathbf{g}$ , the molecular orbital rotation matrix  $\Theta$ , the set of determinant amplitudes  $\mathbf{c}$ , and any other parameters, which we denote as  $\Omega$ . Given this, and using the chain rule, a general first derivative of the energy  $E$  with respect to from any *ab initio* calculation can be written as

$$\begin{aligned} \frac{dE}{dx} &= \frac{\partial E}{\partial x} + \frac{\partial E}{\partial \mathbf{h}} \frac{d\mathbf{h}}{dx} + \frac{\partial E}{\partial \mathbf{g}} \frac{d\mathbf{g}}{dx} + \frac{\partial E}{\partial \Theta} \frac{d\Theta}{dx} + \frac{\partial E}{\partial \mathbf{S}} \frac{d\mathbf{S}}{dx} \\ &+ \frac{\partial E}{\partial \mathbf{c}} \frac{d\mathbf{c}}{dx} + \frac{\partial E}{\partial \Omega} \frac{d\Omega}{dx}. \end{aligned} \quad (14)$$

The remaining challenge is to fill in explicit expressions for the above elements. As the exact energy is independent of the orbital rotational parameters  $\Theta$  and CI coefficients  $\mathbf{c}$ , the

corresponding partial derivatives are identically zero,

$$\frac{\partial E}{\partial \Theta} = \frac{\partial E}{\partial \mathbf{c}} \equiv 0. \quad (15)$$

Several of the partial derivatives of the exact energy are trivially evaluated

$$\begin{aligned} \left( \frac{\partial E}{\partial \mathbf{h}} \right)_{\mu\nu} &= \sum_{pq} \gamma_{pq} C_{\mu p} C_{\nu q}, \\ \left( \frac{\partial E}{\partial \mathbf{g}} \right)_{\mu\nu\lambda\sigma} &= \sum_{pqrs} \Gamma_{pqrs} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s}. \end{aligned} \quad (16)$$

Because there is no dependence on other parameters, e.g.,  $\Omega$ , the only remaining partial derivative of the energy is the one with respect to the overlap of the AO basis functions,

$$\begin{aligned} \frac{dE}{dx} &= \sum_{pq} \sum_{\mu\nu} \gamma_{pq} C_{\mu p} C_{\nu q} \frac{dh_{\mu\nu}}{dx} \\ &+ \sum_{pqrs} \sum_{\mu\nu\lambda\sigma} \Gamma_{pqrs} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s} \\ &\times \frac{dg_{\mu\nu\lambda\sigma}}{dx} + \frac{\partial E}{\partial \mathbf{S}} \frac{d\mathbf{S}}{dx}. \end{aligned} \quad (17)$$

where the only terms that depend on  $\mathbf{S}$  are the MO coefficients  $\mathbf{C}$ .

With the density matrices  $\gamma_{pq}$  and  $\Gamma_{pqrs}$  being given, the first two terms of Eq. (17) are easy to evaluate: they require the evaluation of atomic orbital *core* derivatives. The last term of Eq. (17) is a little more involved as we need to find an expression for  $\partial E/\partial \mathbf{S}$  in terms of the one- and two-body reduced density matrices. The core derivative overlap integrals  $d\mathbf{S}/dx$  can be computed by most electronic structure packages. We obtain for the last term

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{S}} \frac{d\mathbf{S}}{dx} &= \sum_{\eta\zeta} \sum_{pq} \gamma_{pq} \sum_{\mu\nu} h_{\mu\nu} \frac{\partial}{\partial S_{\eta\zeta}} (C_{\mu p} C_{\nu q}) \frac{dS_{\eta\zeta}}{dx} \\ &+ \sum_{\eta\zeta} \sum_{pqrs} \Gamma_{pqrs} \sum_{\mu\nu\lambda\sigma} g_{\mu\nu\lambda\sigma} \\ &\times \frac{\partial}{\partial S_{\eta\zeta}} (C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s}) \frac{dS_{\eta\zeta}}{dx}. \end{aligned} \quad (18)$$

The quantities in the above expression depend on the MO coefficients  $\mathbf{C}$ . Because the MO coefficients depend on  $\mathbf{S}$  we need to derive the explicit expressions for  $\partial \mathbf{C}/\partial \mathbf{S}$ . The step-by-step derivations are presented in Appendix A, and we find

$$\frac{\partial C_{\mu p}}{\partial S_{\lambda\sigma}} \frac{dS_{\lambda\sigma}}{dx} = -\frac{1}{2} \sum_q C_{\mu q} C_{\lambda q} C_{\sigma p} \frac{dS_{\lambda\sigma}}{dx}. \quad (19)$$

With these, after a derivation presented in Appendix A, we find for derivatives of the one- and two-body terms with respect to the overlap matrix  $\mathbf{S}$

$$\frac{\partial E}{\partial \mathbf{S}} \frac{d\mathbf{S}}{dx} = -\sum_{pqm} \gamma_{qm} h_{pm} \frac{dS_{pq}}{dx} - 2 \sum_{pq} \sum_{rst} \Gamma_{qrst} g_{prst} \frac{dS_{pq}}{dx}. \quad (20)$$

The final expression for the energy derivative, after reindexing, is given by

$$\begin{aligned} \frac{dE}{dx} &= \left( \sum_{pq} \gamma_{pq} \frac{dh_{pq}}{dx} + \sum_{pqrs} \Gamma_{pqrs} \frac{dg_{pqrs}}{dx} \right) \\ &- \left( \sum_{pqm} \gamma_{qm} h_{mq} \frac{dS_{mp}}{dx} + 2 \sum_{pqrst} \Gamma_{pqrs} g_{tqrs} \frac{dS_{tp}}{dx} \right). \end{aligned} \quad (21)$$

To calculate the force on the *A*th nucleus,  $\mathbf{F}_A = -\frac{dE}{d\mathbf{R}_A}$ , we use the above expression to find the energy gradient at nuclear position  $\mathbf{R}_A$ . Following the Hellmann-Feynman theorem Eq. (5), we then convert the problem of calculating the energy gradient to the problem of calculating the expectation value of a derivative operator. We find that in second quantization the derivative operator is given by

$$\begin{aligned} \frac{dH}{d\mathbf{R}_A} &= \sum_{pq} a_p^\dagger a_q \left[ \frac{dh_{pq}}{d\mathbf{R}_A} - \sum_m h_{mq} \frac{dS_{mp}}{d\mathbf{R}_A} \right] \\ &+ \sum_{pqrs} a_p^\dagger a_r^\dagger a_q a_s \left[ \frac{dg_{pqrs}}{d\mathbf{R}_A} - 2 \sum_t g_{tqrs} \frac{dS_{tp}}{d\mathbf{R}_A} \right]. \end{aligned} \quad (22)$$

In this equation, the second terms in each bracket, which include the derivative of the overlap matrix  $\mathbf{S}$  correspond to the Pulay force [59]. For later reference, we write coefficients of this operator in the same form as the Hamiltonian

$$T_{pq}^{(F_A)} = \left[ \frac{dh_{pq}}{d\mathbf{R}_A} - \frac{1}{2} \sum_m (h_{mq} \frac{dS_{pm}}{d\mathbf{R}_A} + h_{pm} \frac{dS_{mq}}{d\mathbf{R}_A}) \right], \quad (23)$$

$$\begin{aligned} V_{pqrs}^{(F_A)} &= \left[ \frac{dg_{pqrs}}{d\mathbf{R}_A} - \frac{1}{2} \sum_t \left( g_{tqrs} \frac{dS_{pt}}{d\mathbf{R}_A} \right. \right. \\ &\left. \left. + g_{ptrs} \frac{dS_{tq}}{d\mathbf{R}_A} + g_{pqt s} \frac{dS_{rt}}{d\mathbf{R}_A} + g_{pqrt} \frac{dS_{ts}}{d\mathbf{R}_A} \right) \right]. \end{aligned} \quad (24)$$

## B. Force operators in plane wave bases

Plane waves are one of the most common basis sets used to model condensed matter systems. They are a natural basis for periodic systems and are independent of the atomic positions. However, their drawback is that many plane waves are typically needed to describe the wave functions accurately. When defined on a cubic reciprocal lattice, the plane wave basis functions take the form

$$\phi_{\mathbf{p}}(\mathbf{r}) = \sqrt{\frac{1}{\Omega}} e^{-i\mathbf{k}_{\mathbf{p}} \cdot \mathbf{r}}, \quad (25)$$

where  $\Omega$  is the computational cell volume and the reciprocal lattice vector in three dimensions is defined as

$$\mathbf{k}_{\mathbf{p}} = \frac{2\pi\mathbf{p}}{\Omega^{1/3}}, \quad \mathbf{p} \in G, \quad G = \left[ -\frac{N^{1/3}-1}{2}, \frac{N^{1/3}-1}{2} \right]^3 \subset \mathbb{Z}^3, \quad (26)$$

with  $N$  being the number of plane waves. The molecular integrals can be evaluated analytically for the case of plane wave basis functions, leading to the following representation of the second-quantized electronic structure Hamiltonian in

Eq. (10):

$$H = \underbrace{\sum_{\mathbf{p}} \frac{|\mathbf{k}_{\mathbf{p}}|^2}{2} a_{\mathbf{p}}^{\dagger} a_{\mathbf{p}} - \frac{4\pi}{\Omega} \sum_{\mathbf{p} \neq \mathbf{q}} \sum_A^{\eta} \left( Z_A \frac{e^{i\mathbf{k}_{\mathbf{q}-\mathbf{p}} \cdot \mathbf{R}_A}}{|\mathbf{k}_{\mathbf{q}-\mathbf{p}}|^2} \right) a_{\mathbf{p}}^{\dagger} a_{\mathbf{q}}}_{\text{one-electron term}} + \underbrace{\frac{2\pi}{\Omega} \sum_{\mathbf{p} \neq \mathbf{q}, \mathbf{s} \neq 0} \frac{1}{|\mathbf{k}_{\mathbf{s}}|^2} a_{\mathbf{p}}^{\dagger} a_{\mathbf{q}}^{\dagger} a_{\mathbf{q}+\mathbf{s}} a_{\mathbf{p}-\mathbf{s}}}_{\text{two-electron term}}. \quad (27)$$

Here and in what follows, we have omitted the electron spin for simplicity. An equivalent expression in first quantization can be written down [57],

$$H = \underbrace{\sum_{i=1}^{\eta} \sum_{\mathbf{p} \in G} \frac{|\mathbf{k}_{\mathbf{p}}|^2}{2} |\mathbf{p}\rangle \langle \mathbf{p}|_i - \frac{4\pi}{\Omega} \sum_{A=1}^{N_a} \sum_{i=1}^{\eta} \sum_{\mathbf{p} \neq \mathbf{q} \in G} \left( Z_A \frac{e^{i\mathbf{k}_{\mathbf{q}-\mathbf{p}} \cdot \mathbf{R}_A}}{|\mathbf{k}_{\mathbf{q}-\mathbf{p}}|^2} \right) |\mathbf{p}\rangle \langle \mathbf{q}|_i}_{\text{one-electron term}} + \underbrace{\frac{2\pi}{\Omega} \sum_{i,j=1}^{\eta} \sum_{\mathbf{p}, \mathbf{q} \in G} \sum_{\substack{\mathbf{s} \in G_0 \\ (\mathbf{p}+\mathbf{s}) \in G \\ (\mathbf{q}-\mathbf{s}) \in G}} \frac{1}{|\mathbf{k}_{\mathbf{s}}|^2} |\mathbf{p}+\mathbf{s}\rangle \langle \mathbf{p}|_i |\mathbf{q}-\mathbf{s}\rangle \langle \mathbf{q}|_j}_{\text{two-electron term}}, \quad (28)$$

where  $|\mathbf{p}\rangle \langle \mathbf{q}|_j$  is a shorthand for  $I_1 \otimes \dots \otimes |\mathbf{p}\rangle \langle \mathbf{q}|_j \otimes \dots \otimes I_{\eta}$  and  $G_0$  is  $G$  from Eq. (26) excluding the zero mode.

While a number of papers have analyzed the viability of quantum algorithms for simulating chemistry in second quantization with plane waves [10,67–70], that approach faces some significant challenges. In particular, in second quantization the number of qubits required scales as the number of plane waves. This is a problem because often hundreds of thousands of plane waves might be required to obtain a suitable wave function accuracy. However, there have been proposals for fault-tolerant algorithms using plane waves in first quantization [57,58]. In first quantization the number of qubits required scales only as the logarithm of the number of plane waves  $N$  and linearly in the number of electrons,  $\eta$ . Algorithms have been demonstrated [57] for time-evolution or state preparation of molecular systems that scale only as

$$\tilde{O}(\eta^3 N^{1/3} \Omega^{-1/3}). \quad (29)$$

Due to the sublinear dependence on  $N$ , with these approaches one can conceivably perform simulations with millions of plane waves.

An additional advantage of the plane wave basis is that the overlap matrix elements in the plane wave representation are reduced to

$$S_{\mathbf{p}\mathbf{q}} = \frac{1}{\Omega} \int_{\Omega} d\mathbf{r} e^{i\mathbf{k}_{\mathbf{q}-\mathbf{p}} \cdot \mathbf{r}} = \delta_{\mathbf{p}\mathbf{q}}, \quad (30)$$

and thus the overlap matrix contributions to the derivative operator from Eq. (22) are identically zero. This suggests that representing the electronic structure Hamiltonian in first-quantized plane waves basis is a promising avenue for calculating energy derivatives of chemical systems.

In a plane wave basis, the only dependence of  $H$  on the nuclear positions  $\mathbf{R}_A$  is in the one-body term, which implies

that (as expected for a nonatomic centered basis set) the force operator in plane waves is a strictly one-body operator. (The same is true for other first-order derivatives that do not affect the electron-electron Coulomb, such as an applied electric or magnetic field.) This operator may be further simply diagonalized by the fermionic fast Fourier transform (FFFT) [10,71,72], in a similar manner to the potential term of the original Hamiltonian. This is simplest to demonstrate in second quantization, so we will perform the calculation there first and then transform to our target first-quantized form. Differentiating Eq. (27) with respect to the  $A$ th nuclear coordinate  $\mathbf{R}_A$  gives us

$$\begin{aligned} \frac{dH}{d\mathbf{R}_A} &= -\frac{4\pi i Z_A}{\Omega} \sum_{\mathbf{p} \neq \mathbf{q}} \frac{\mathbf{k}_{\mathbf{q}-\mathbf{p}} e^{i\mathbf{k}_{\mathbf{q}-\mathbf{p}} \cdot \mathbf{R}_A}}{|\mathbf{k}_{\mathbf{q}-\mathbf{p}}|^2} a_{\mathbf{p}}^{\dagger} a_{\mathbf{q}} \\ &= \sum_{\mathbf{p} \neq \mathbf{q}} f(\mathbf{q} - \mathbf{p}, A) e^{i\mathbf{k}_{\mathbf{q}-\mathbf{p}} \cdot \mathbf{R}_A} a_{\mathbf{p}}^{\dagger} a_{\mathbf{q}}, \\ f(\mathbf{s}, A) &= -\frac{4\pi i Z_A \mathbf{k}_{\mathbf{s}}}{\Omega |\mathbf{k}_{\mathbf{s}}|^2}. \end{aligned} \quad (31)$$

Note that this is a vector-valued derivative, here  $\mathbf{R}_A$  is the 3-dimensional nuclear position vector [individual components of this vector may be obtained by taking individual components of the wave vector  $\mathbf{k}_{\mathbf{s}}$  in  $f(\mathbf{s}, A)$ ]. In second quantization, the FFFT performs the following single-particle rotation,

$$\begin{aligned} a_{\mathbf{s}}^{\dagger} &= \text{FFFT}^{\dagger} c_{\mathbf{s}}^{\dagger} \text{FFFT} = \sqrt{\frac{1}{N}} \sum_{\mathbf{p}} c_{\mathbf{p}}^{\dagger} e^{-i\mathbf{k}_{\mathbf{s}} \cdot \mathbf{r}_{\mathbf{p}}}, \\ a_{\mathbf{s}} &= \text{FFFT}^{\dagger} c_{\mathbf{s}} \text{FFFT} = \sqrt{\frac{1}{N}} \sum_{\mathbf{p}} c_{\mathbf{p}} e^{i\mathbf{k}_{\mathbf{s}} \cdot \mathbf{r}_{\mathbf{p}}} \end{aligned} \quad (32)$$

where  $\mathbf{r}_p = \mathbf{p}(\Omega/N)^{1/3}$ . Under this transformation, the gradient of the electronic structure Hamiltonian becomes

$$\begin{aligned}
\frac{dH}{d\mathbf{R}_A} &= \sum_{\mathbf{p} \neq \mathbf{q}} f(\mathbf{q} - \mathbf{p}, A) e^{i\mathbf{k}_{q-p} \cdot \mathbf{R}_A} a_{\mathbf{p}}^\dagger a_{\mathbf{q}} \\
&= \sum_{\mathbf{p} \neq \mathbf{q}} f(\mathbf{q} - \mathbf{p}, A) e^{i\mathbf{k}_{q-p} \cdot \mathbf{R}_A} \left( \sqrt{\frac{1}{N}} \sum_{\mathbf{p}'} c_{\mathbf{p}'}^\dagger e^{-i\mathbf{k}_p \cdot \mathbf{r}_{p'}} \right) \left( \sqrt{\frac{1}{N}} \sum_{\mathbf{q}'} c_{\mathbf{q}'} e^{i\mathbf{k}_q \cdot \mathbf{r}_{q'}} \right) \\
&= \frac{1}{N} \sum_{\mathbf{p} \neq \mathbf{q}} f(\mathbf{q} - \mathbf{p}, A) e^{i\mathbf{k}_{q-p} \cdot \mathbf{R}_A} \sum_{\mathbf{p}', \mathbf{q}'} c_{\mathbf{p}'}^\dagger c_{\mathbf{q}'} e^{i\mathbf{k}_q \cdot \mathbf{r}_{q'} - \mathbf{p}' \cdot \mathbf{r}_{p'}} e^{-i\mathbf{k}_{p-q} \cdot \mathbf{r}_{p'}} \\
&= \frac{1}{N} \sum_{\mathbf{p}', \mathbf{q}'} \sum_{\mathbf{p} \neq \mathbf{q}} f(\mathbf{q} - \mathbf{p}, A) e^{i\mathbf{k}_{q-p} \cdot (\mathbf{R}_A + \mathbf{r}_{p'})} (c_{\mathbf{p}'}^\dagger c_{\mathbf{q}'} e^{i\mathbf{k}_q \cdot \mathbf{r}_{q'} - \mathbf{p}' \cdot \mathbf{r}_{p'}}). \tag{33}
\end{aligned}$$

Recognizing that  $(\mathbf{q} - \mathbf{p})$  spans the full set of momentum vectors in our system due to aliasing, we can replace the sum over  $\mathbf{q} \neq \mathbf{p}$  and the indices  $\mathbf{q} - \mathbf{p}$  and  $\mathbf{q}$  with a sum over  $\mathbf{s} \neq 0$  and  $\mathbf{p}$ . Following this re-indexing, our gradient operator diagonalizes immediately,

$$\begin{aligned}
\frac{dH}{d\mathbf{R}_A} &= \frac{1}{N} \sum_{\mathbf{p}', \mathbf{q}'} \left( \sum_{\mathbf{s} \neq 0} f(\mathbf{s}, A) e^{i\mathbf{k}_s \cdot (\mathbf{R}_A + \mathbf{r}_{p'})} \right) \\
&\quad \times \left( c_{\mathbf{p}'}^\dagger c_{\mathbf{q}'} \sum_{\mathbf{q}} e^{i\mathbf{k}_q \cdot \mathbf{r}_{q'} - \mathbf{p}' \cdot \mathbf{r}_{p'}} \right) \\
&= \sum_{\mathbf{p}} \left( \sum_{\mathbf{s} \neq 0} f(\mathbf{s}, A) e^{i\mathbf{k}_s \cdot (\mathbf{R}_A + \mathbf{r}_p)} \right) c_{\mathbf{p}}^\dagger c_{\mathbf{p}} \\
&= -\frac{4\pi i Z_A}{\Omega} \sum_{\mathbf{p}} \sum_{\mathbf{s} \neq 0} \frac{\mathbf{k}_s e^{i\mathbf{k}_s \cdot (\mathbf{R}_A + \mathbf{r}_p)}}{|\mathbf{k}_s|^2} c_{\mathbf{p}}^\dagger c_{\mathbf{p}} \tag{34}
\end{aligned}$$

where we have used the fact that the summation grouped on the right side of the first equation is equal to zero unless  $\mathbf{p}' = \mathbf{q}'$ . This is because the negative modes of  $\mathbf{k}_q$  will have exactly the opposite phase as the positive modes of  $\mathbf{k}_q$ .

We now transform the derivative operator into a first-quantized representation. In first quantization, we store our wave function by having a computational basis that encodes configurations of the electrons in  $N$  basis functions such that a configuration is specified as  $|\phi_1 \phi_2 \dots \phi_\eta\rangle$  where each  $\phi_j$  encodes the index of an occupied basis function. Each  $\phi_j$  may be specified in binary, making the space complexity only  $\mathcal{O}(\eta \log N)$ . We can translate Eq. (34) into first quantization to give

$$\begin{aligned}
\frac{dH}{d\mathbf{R}_A} &= -\frac{4\pi i Z_A}{\Omega} \sum_{i=1}^{\eta} \sum_{\substack{\mathbf{p}, \mathbf{q} \in G \\ \mathbf{p} \neq \mathbf{q}}} \frac{\mathbf{k}_{q-p} e^{i\mathbf{k}_{q-p} \cdot \mathbf{R}_A}}{|\mathbf{k}_{q-p}|^2} |\mathbf{p}\rangle \langle \mathbf{q}|_i \\
&= \text{QFT} \left( -\frac{4\pi i Z_A}{\Omega} \sum_{i=1}^{\eta} \sum_{\mathbf{p}, \mathbf{s} \in G_0} \frac{\mathbf{k}_s e^{i\mathbf{k}_s \cdot (\mathbf{R}_A - \mathbf{r}_p)}}{|\mathbf{k}_s|^2} |\mathbf{p}\rangle \langle \mathbf{p}|_i \right) \text{QFT}^\dagger \tag{35}
\end{aligned}$$

where the QFT is the quantum Fourier transform with aliased frequencies [the first quantized version of the FFFT from Eq. (34)]. The QFT can be implemented with Toffoli gate complexity  $\tilde{\mathcal{O}}(\eta)$  [10]. Note that this guarantees that all force

operators are mutually diagonal under the FFFT/QFT, which in turn implies that all force operators commute.

### C. Error tolerance for applications

The most widely-used energy derivatives of a molecular system are nuclear forces. The first application we consider is geometry optimization, where nuclear derivatives are used to find the geometry of the molecule with the lowest energy on the potential energy surface. The second application is molecular dynamics, where the nuclear positions are propagated through time by a classical differential equation within the Born-Oppenheimer approximation. At each time, the forces on the nuclei determine their next position. Both of these applications rely on the nuclear derivatives of the energy to repeatedly update the positions of the nuclei. This is a process where a small error in each step can quickly accumulate. The tolerable error on the forces is an important parameter in the scaling of the quantum algorithms to calculate them. In the next two subsections we investigate the error level that is acceptable.

#### 1. Geometry optimization

The error tolerance of the forces for the geometric relaxation of a structure depends strongly on the system. The geometries of systems with a rather steep potential energy surface can be determined with relative low accuracy of the forces. For example, Gaussian sets the default thresholds for convergence of the maximum force to 0.9 mHa/Å and the RMS of the error of single force component to 0.6 mHa/Å [73]. However, for systems where forces are smaller because the potential energy surface is shallow, typically the geometries need to be determined by relaxing the atomic positions until the forces are one magnitude smaller [74].

#### 2. Molecular dynamics

Error bounds on forces required for MD simulations will again depend strongly on the system studied. To find a simple baseline for a target accuracy for force components in this paper, we focus on the required error tolerance for performing semiclassical MD simulations of a water system. A quantum device would be used in this situation as a subroutine to provide accurate estimates of the classical potential, employing the TIP3P water model [75]. Here, the MD simulations



were performed with Atomic Simulation Environment software package [76].

As a proxy for simulation convergence, we study the 2-particle radial distribution function

$$g^{(2)}(\mathbf{R}_1, \mathbf{R}_2) = \left(\frac{V}{N_a}\right)^2 \frac{N_a(N_a - 1)}{Z_{N_a}} \int_{V^{N_a-2}} \exp[-\beta E_{\text{tot}}(\mathbf{R}_1, \dots, \mathbf{R}_{N_a})] d\mathbf{R}_3 \dots d\mathbf{R}_{N_a}, \quad (36)$$

of a  $N_a = 3 \times 216$ -atom system in a periodic box of volume  $V$ . Here,  $Z_{N_a}$  is the partition function of the  $N_a$  particles, and  $\beta = 1/299$  K is the inverse temperature of the system. As the radial distribution function  $g^{(2)}$  is independent of translations and rotations of  $\mathbf{R}_1$  and  $\mathbf{R}_2$  about the origin, the data it contains can be found solely in the radial term,

$$g^{(2)}(R) = \int_{V^2} \delta(\|\mathbf{R}_1 - \mathbf{R}_2\| - R) g^{(2)}(\mathbf{R}_1, \mathbf{R}_2) d\mathbf{R}_1 d\mathbf{R}_2. \quad (37)$$

To obtain a physically-relevant quantity, we isolate the pair distribution of the  $N_a/3 = 216$  oxygen atoms and ignore the hydrogen atoms. The 2-particle radial distribution function is a macroscopic quantity and is often used to benchmark different water models. The deviation from the ideal radial distribution function is a measure of the quality of the run. We test for which size of errors in the force we can still reproduce the error free radial distribution function. To achieve this we performed micro-canonical MD simulations with 36 000 time steps, and at each 1 fs time step we added a random error term to the forces of water molecules. This error was sampled from a Gaussian distribution with a given RMS of the force error, with separate error terms drawn independently. The radial distribution function of Eq. (36) is averaged over all time steps of the simulation. From these MD runs we can determine the size of errors for which we can still reproduce the error-free radial distribution function, which we take as the ground truth. Fig. 1 shows that the radial distribution function rapidly converges to the error-free radial distribution function. As a metric for convergence, we focus on the largest feature in the system (the peak around 2.8 Å, and plot the error in the peak position, Fig. 1 inset). From this we conclude that for water the RMS error for molecular dynamics needs to be smaller than 6.4 mHa/Å to reproduce macroscopic properties as the radial distribution function. Our findings are in agreement with studies where *ab initio* MD simulations based on quantum Monte Carlo calculations were employed [77]. Beyond radial distribution functions, quantities such as vibrational density of states are expected to require higher precision of the force evaluation [78]. We demonstrate how a quantum computer would be used in a full molecular dynamics simulation in Fig. 2.

The geometry optimization seems to be more stringent by an order of magnitude on required accuracy of the forces than MD simulations. In MD simulations errors can average out. Later in this paper we consider the 2-norm as the parameter for accuracy of the forces. For the 2-norm the sum is taken over all forces and its components. Therefore, the 2-norm is an extensive quantity and depends on the number of atoms. We can convert the RMS error to the 2-norm by multiplying RMS error with  $\sqrt{3N_a}$ . For example, for the 216 water molecules we obtain a 2-norm of the error of 282.2 mHa/Å.

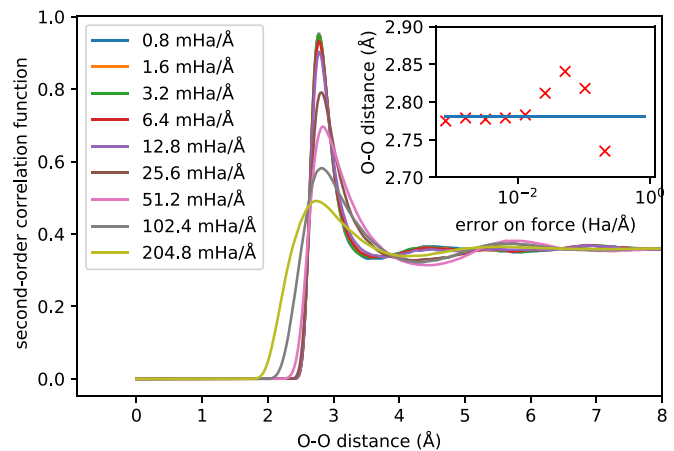


FIG. 1. The radial distribution function, which is second-order correlation function of two oxygen atoms as function of their distance. The MD run contained 216 water molecules in a periodic box with a temperature of 299 K. The radial distribution function is plotted for different RMS errors on the force from 409.6 mHa/Å to 0.8 mHa/Å. In the inset the peak positions of radial distribution function is plotted as function of the error on the force.

### III. COMPUTATION OF FORCE VECTORS IN NISQ

To optimize quantum algorithms for NISQ quantum computers, we must reduce quantum circuit depth wherever practical. Near-term proposals for quantum chemistry achieve this by preparing approximate ground states [17,81,82], from which energies may be extracted by state tomography. As long as the approximate state is variationally optimized within the active space considered on the device, the energy derivatives yielded by the Hellman-Feynman theorem are accurate for the variational energy; no further corrections need to be made to the methods outlined in Sec. II. In this paper, we assume access to the preparation of an initial state  $|\psi\rangle$  that satisfies the Hellman-Feynman theorem Eq. (5) ( $\frac{dE}{dR_i} = \langle \psi | \frac{dH}{dR_i} | \psi \rangle$ ), and focus on the optimization of the measurement of this state to extract an estimate of  $\langle \psi | \frac{dH}{dR_i} | \psi \rangle$ . While many approaches exist for reconstructing the expectation, tomographic approaches have a major role in recent quantum computing works [19,20,54,60,83]. However, to the best of our knowledge no-one has optimized estimation methods for the measurement of vectors of operators prior to now.

State tomography in NISQ is complicated by the fact that simultaneous direct measurement of multiple operators is only possible in quantum mechanics when all operators mutually commute. More broadly, the parameter estimation or partial tomographic protocols used to estimate a gradient consist of three key steps:

- (1) Define a set of basis rotations  $\{Q_j\}$  for which low-depth quantum circuits are known.
- (2) For each basis rotation  $Q_j$ , prepare the state  $|\psi\rangle$   $M_j$  times, apply the quantum circuit, and then destructively measures the system in the computational basis.
- (3) Estimate the set  $\{\langle \psi | \frac{dH}{dR_i} | \psi \rangle\}$  from the observed measurement data.

In a NISQ cost model, the target is to reduce the total number  $M = \sum_j M_j$  of preparations of  $|\psi\rangle$ , or experiment

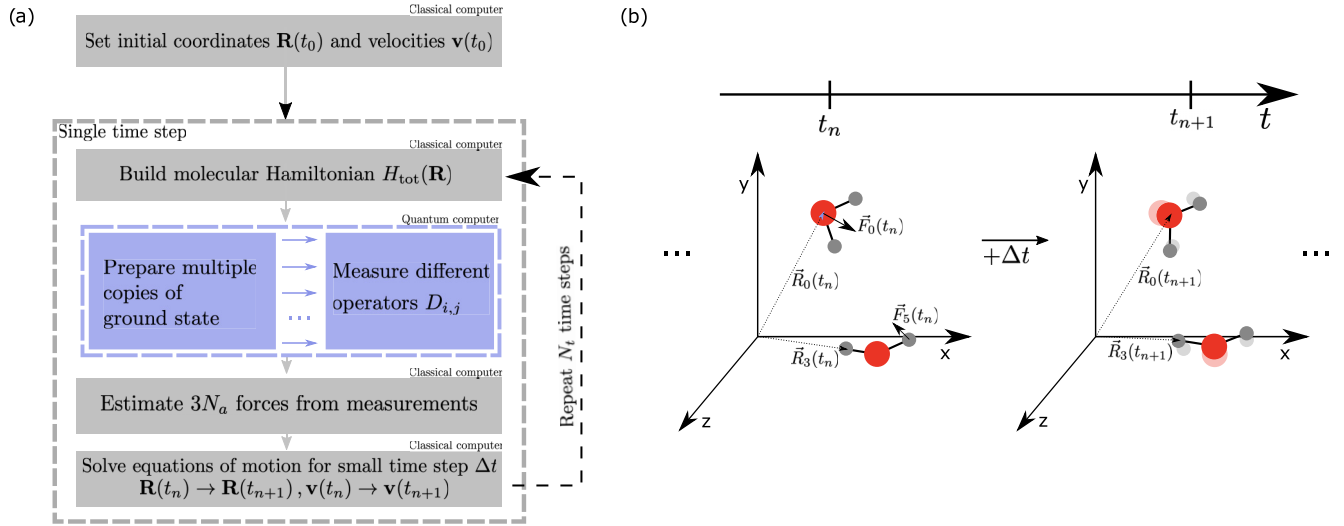


FIG. 2. Schematic representation of molecular dynamics enhanced by a NISQ device. (a) Flowchart highlighting the hybrid setup in which a NISQ device is used to calculate the forces, while a classical computer updates the nuclei coordinates  $\mathbf{R}$ , velocities  $\mathbf{v}$ , and the molecular Hamiltonian. A typical MD simulation requires  $N_t = \mathcal{O}(10^6 - 10^9)$  time steps with a step size of  $\Delta t = \mathcal{O}(10^{-15})$ s [79,80]. (b) Example of a single update of the nuclei coordinates  $\mathbf{R}$  of two water molecules with red balls and gray balls denoting oxygen and hydrogen atoms respectively.  $\vec{F}_i$  denotes the three-dimensional force vector on the  $i$ th atom.

“shots”, while targeting some error bound on the set of energy gradient estimates.

If the expectation value estimation (step 3 above) is linear and unbiased, the error on individual estimates  $\langle \psi | \frac{dH}{dR_i} | \psi \rangle$  may be calculated by variance propagation. Such an estimation corresponds to the decomposition of the gradient operator as a linear combination

$$\frac{dH}{dR_i} = \sum_j D_{i,j}, \quad (38)$$

where the  $D_{i,j}$  are operators that are diagonalized by the  $j$ th basis rotation  $Q_j$ . If  $Q_j$  diagonalizes  $D_{i,j}$ ,  $\langle \psi | D_{i,j} | \psi \rangle$  may be estimated by averaging over the  $M_j$  destructive measurements taken in the  $Q_j$  basis. The variance in this estimator is given by

$$\text{Var}[\langle \psi | D_{i,j} | \psi \rangle] = \frac{\sigma_{i,j}^2}{M_j}, \quad (39)$$

where  $\sigma_{i,j}$  is the quantum-mechanical variance of the observable  $D_{i,j}$  with respect to the state  $|\psi\rangle$  (which is distinct from the variance of the estimator),

$$\sigma_{i,j}^2 = \langle \psi | D_{i,j}^2 | \psi \rangle - \langle \psi | D_{i,j} | \psi \rangle^2. \quad (40)$$

As expectation values are linear, we have

$$\frac{dE}{dR_i} = \langle \psi | \frac{dH}{dR_i} | \psi \rangle = \sum_j \langle \psi | D_{i,j} | \psi \rangle. \quad (41)$$

Then, as each  $D_{i,j}$  is measured independently, the variance of the estimation propagates in the usual way to the variance in an estimation of  $\frac{dE}{dR_i}$ ,

$$\epsilon_i^2 = \text{Var}\left[\frac{dE}{dR_i}\right] = \sum_j \frac{\sigma_{i,j}^2}{M_j}. \quad (42)$$

These errors may be captured within a  $3N_a$ -dimensional error vector  $\epsilon$ . In practice estimates of  $\sigma_{i,j}^2$  are not known in advance, making exact estimation of  $\epsilon_i$  and subsequent parameter optimization difficult. Instead, bounds on  $\sigma_{i,j}$  are often substituted; we will introduce various such methods throughout this section. Some of these bounds are in practice quite weak, which implies that fair comparison of the results described in this section may not be possible.

The general state tomography method described above leaves open a large number of parameters for optimization: the rotations  $Q_j$ , the shot allocation  $M_j$ , and the choice of operators  $D_{i,j}$  in the decomposition of  $\frac{dH}{dR_i}$ . In the following sections, we will describe and compare various methods that attempt to optimize these choices. Complete optimization of each of these choices is not practical due to the sheer number of parameters and the (classical) cost of evaluating cost functions. Optimizing basis rotations  $Q_j$  to diagonalize multiple operators is in general an NP-hard problem [84]. Moreover, the lack of precise knowledge of  $\sigma_{i,j}$  implies that the cost function may be difficult to estimate for the purposes of optimization. Various heuristic techniques are widely known [19,20,60,85,86], and many of these can be shown to achieve asymptotically optimal results in terms of the number of groups (up to constant factors). However, the risk that Pauli grouping may significantly increase covariances for measurements within said groups needs to be taken into account for proper shot distribution [17,87].

### A. Basis rotation choices

When choosing the set of basis rotations  $Q_j$  for a state tomography protocol, one must try find operators  $D_{i,j}$  that are diagonal in the  $Q_j$  basis. Calculating such operators is typically as difficult as simulating the circuit, so rotations  $Q_j$  are typically chosen to be classically easy to simulate. One must take further care that the  $D_{i,j}$  are not exponentially

difficult to express in order for step 3 in the general method above to be computationally feasible. Drawing  $Q_j$  from one of a few well-known sets of quantum circuits defined below typically solves this problem.

The commonly used Clifford circuits form the first example. These circuits preserve the Pauli group  $\mathbb{P}^N = \{I, X, Y, Z\}^N$  modulo complex phases. If  $Q_j$  is a Clifford circuit, so is  $Q_j^\dagger$  and the algebra formed by the set  $Q_j^\dagger Z_n Q_j$  yields all possible operators  $D_{i,j}$  that are diagonal in this basis. Moreover, Clifford circuits can in principle be constructed to simultaneously diagonalize any set of mutually commuting elements of  $\mathbb{P}^N$ . It is relatively easy to design a set of Clifford basis rotations  $Q_j$  in this manner:

(1) Decompose all operators  $\frac{dH}{dR_i}$  into a linear combination of Pauli operators following the Jordan-Wigner, Bravyi-Kitaev, or alternative fermion-to-qubit transformation.

(2) Subdivide the set  $\mathcal{S}$  of Pauli operators that appear in at least one linear combination into commuting subsets  $\mathcal{S}_j$  (i.e., so that all Pauli operators within each  $\mathcal{S}_j$  commute).

(3) For each subset, find an appropriate basis rotation  $Q_j$ .

A disadvantage to the above is that Clifford circuits to diagonalize mutually-commuting operators can be relatively deep [53,88]. This can be simplified by adding the requirement that the subsets  $\mathcal{S}_j$  contain not commuting Pauli operators, but amenable Pauli operators. Two Pauli operators are amenable if, on each qubit the tensor factor  $I, X, Y, Z$  of both operators is the same or the tensor factor of at least one operator is the identity. (For example,  $YY$  and  $YI$  are amenable, but  $YY$  and  $XX$  are not.) Amenable Pauli operators can be mutually diagonalized by single-qubit basis rotations  $Q_j$ , making this a practical subdivision  $\mathcal{S}_j$ . In general subdividing  $\mathcal{S}$  into the minimum number of  $\mathcal{S}_j$  is a known NP-hard problem [84], although relative success has been found in heuristics [20] or brute-force optimization methods [88].

A second set of circuits relevant for diagonalizing Hamiltonians and force operators in chemistry are Givens rotation circuits. These correspond to evolution by a one-body fermionic operator

$$Q_j = e^{i \sum_{n,m} Q_j^{n,m} a_n^\dagger a_m}, \quad (43)$$

and are classically tractable to calculate as they map single creation and annihilation operators to each other,

$$Q_j a_n Q_j^\dagger = \sum_m [e^{iq_j}]_{n,m} a_m, \quad (44)$$

where  $q_j$  is the  $N \times N$  Hermitian matrix with elements taken from Eq. (43). Givens rotation circuits are relatively low depth; an arbitrary Givens rotation may be implemented in depth  $2N$  on a linear array using precisely  $N^2$  two qubit gates [28,89]. The above may be slightly generalized to the set of fermionic Gaussian unitaries [60]

$$Q_j = e^{\sum_{p,q} g_j^{p,q} \gamma_p \gamma_q}, \quad (45)$$

where  $\gamma_m$  and  $\gamma_n$  are anticommuting Majorana operators

$$\gamma_{2n} = a_n + a_n^\dagger \quad \gamma_{2n+1} = -i(a_n - a_n^\dagger). \quad (46)$$

This strictly contains the set of Givens rotations, and also allows for Bogoliubov-style rotations between  $a_n$  and  $a_n^\dagger$ .

A low-cost method for constructing Givens rotation circuits  $Q_j$  to target a two-body fermionic operator is to factorize the operator [11,19]. Starting from the operator in its chemist formulation

$$A = \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p,q} T_{pq} a_{p,\sigma}^\dagger a_{q,\sigma} + \underbrace{\sum_{\alpha, \beta \in \{\uparrow, \downarrow\}} \sum_{p,q,r,s} V_{pqrs} a_{p,\alpha}^\dagger a_{q,\alpha} a_{r,\beta}^\dagger a_{s,\beta}}_{=V}, \quad (47)$$

we reshape the 4-rank tensor  $V_{pqrs}$  into a 2-rank tensor  $M_{(pq),(rs)}$ . A direct diagonalization (or a Cholesky decomposition) of the flattened version of  $V_{pqrs}$  yields

$$V = \sum_{\ell=1}^L w_\ell \left( \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p,q=1}^{N/2} g_{pq}^{(\ell)} a_{p,\sigma}^\dagger a_{q,\sigma} \right)^2 = \sum_{\ell=1}^L \left( \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p,q=1}^{N/2} W_{pq}^{(\ell)} a_{p,\sigma}^\dagger a_{q,\sigma} \right)^2 = \sum_{\ell=1}^L W^{(\ell)2} \quad (48)$$

with  $g_{pq}^{(\ell)}$  and  $w_\ell$  representing the eigenvectors and eigenvalues respectively. Further diagonalization of the squared single-body operators yields [15]

$$V = \sum_{\ell=1}^L U^{(\ell)} \left( \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p=1}^{M_\ell} f_p^{(\ell)} n_{p,\sigma} \sum_{\sigma' \in \{\uparrow, \downarrow\}} \sum_{q=1}^{M_\ell} f_q^{(\ell)} n_{q,\sigma'} \right) U^{(\ell)\dagger}, \quad (49)$$

with  $f_p^{(\ell)}$  the eigenvalues of  $W_{pq}^{(\ell)}$  and  $U_\ell$  the unitaries performing the diagonalization, which can be expressed as a single-particle change of basis unitary

$$U^{(\ell)} = \exp \left( - \sum_{p=1}^N \sum_{q=1}^{M_\ell} \kappa_{pq}^{(\ell)} (a_p^\dagger a_q - a_q^\dagger a_p) \right) \\ U^{(\ell)} W^{(\ell)} U^{(\ell)\dagger} = \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p=1}^{M_\ell} f_p^{(\ell)} n_{p,\sigma}, \quad (50)$$

and the  $\kappa_{p,q}$  are obtained from the Givens rotation procedure in [89]. Similarly, the one-body fermionic operator may be diagonalized by a single Givens rotation  $Q_j$ , as one simply takes the rotation  $q_j$  that diagonalizes the corresponding  $N \times N$  one-body matrix [following Eq. (44)].

If the operator  $A$  given in Eq. (47) is the electronic structure Hamiltonian in a generic second-quantized basis, we have that  $L = \tilde{\mathcal{O}}(N)$  and  $M < N$  and in some special cases  $M_\ell = \mathcal{O}(\log N)$  [11]. This implies that one may estimate the expectation value of a Hamiltonian with  $\tilde{\mathcal{O}}(N)$  basis rotations  $Q_j$ . As we show in Sec. IV B, this extends to a bound on the number of Givens rotations required to factorize a single derivative operator

$$\frac{dH}{dR_i} = \sum_{\ell=1}^L W^{(i,\ell)2}, \quad W^{(i,\ell)} = \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p,q=1}^{N/2} W_{pq}^{(i,\ell)} a_{p,\sigma}^\dagger a_{q,\sigma}. \quad (51)$$

However, factorizations do not typically parallelize; the set of Givens rotation circuits that measure  $\frac{dH}{dR_i}$  will not typically

allow estimation of  $\frac{dH}{dR_q}$ . Moreover, it was found recently that the  $L \sim \tilde{O}(N)$  scaling is relatively delicate; subtracting operators from one derivative will tend to yield an operator that is no longer low rank [90]. This implies that it is likely not possible to significantly parallelize factorized methods, and the number of Givens rotations required to measure  $3N_a$  derivative operators likely scales as  $\tilde{O}(N_a N)$ .

An obvious question to ask is whether the set of fermionic Gaussian unitaries and Clifford circuits intersect. The answer to this question is yes: The intersection of these operators are the fermionic Gaussian Clifford unitaries, which are generated by the set of Majorana swap operators

$$e^{\frac{\pi}{4}\gamma_p\gamma_q}, \quad (52)$$

and correspond to the symmetric permutation group  $\text{Sym}(2N)$  (being permutations of indices of Majoranas). For the sake of measurement, the effect of a Majorana permutation  $Q_j$  is to pair the set of  $2N$  Majorana operators: to choose a set of disjoint pairs  $\{(\gamma_p, \gamma_q)\}$  containing all operators, and permute  $p \rightarrow 2n, q \rightarrow 2n+1$  for some  $n$ . This maps the Hermitian operator  $i\gamma_p\gamma_q \rightarrow Z_n$ , which implies that any linear combination of products of the pairs are diagonalized by  $Q_j$ . This allows simultaneous measurement of  $\binom{2N}{2} = \frac{N(N-1)}{2}$  linearly-independent 2-body fermionic terms, which is optimal, and the basis for the best-known measurement schemes for the estimation of arbitrary 2-body fermionic operators [20,60].

### B. Parallelized importance sampling

Once an optimal set of basis rotations  $Q_j$  and operators  $D_{i,j}$  have been chosen, it remains to allocate the number of shots  $M_j$  to each  $Q_j$ . Here, we target minimizing a given cost function  $f(\{M_j\}, \{\sigma_{i,j}\})$  while keeping the total number of measurements  $M = \sum_j M_j$  constant (or vice versa). This may be achieved by Lagrangian methods [91,92], this methodology being a form of importance sampling over the expectation values  $\langle D_{i,j} \rangle$ . Such methods entail adding the total number of measurements as a constraint to the cost function with a Lagrangian multiplier  $\lambda$ , giving a Lagrangian

$$\mathcal{L} = f(\{M_j\}, \{\sigma_{i,j}\}) + \lambda \left( \sum_j M_j - M \right). \quad (53)$$

The solution to the problem is then achieved by minimizing  $\mathcal{L}$  with respect to all free parameters:  $M_j$  and  $\lambda$ . (See Appendix B 1 for more details and explicit calculations of the optimizations used in the text.) Crucially,  $\sigma_{i,j}$  is typically not known *a priori*. In principle  $\sigma_{i,j}$  can be estimated during the expectation value estimation procedure, which could be used to adaptively optimize the distribution of the  $M_j$ . However, typically in the literature a range of bounds  $\bar{\sigma}_{i,j} > \sigma_{i,j}$  are used instead. We will discuss the known bounds in detail in the next section.

In order to perform the above minimization procedure, we must define the cost function  $f(\{M_j\}, \{\sigma_{i,j}\})$ . This is complicated by the fact that we estimate  $3N_a$  force components, and must combine the error on each into a single-cost function. This can be achieved by defining a norm on the error vector  $\epsilon = \mathbb{E}(\frac{dH}{d\mathbf{R}} - \frac{dH}{d\mathbf{R}})$ . In Sec. II C, we saw that the 2-norm

is a reasonable proxy to bound the error in molecular dynamics simulations. An additional issue presents itself as we should take into account the covariance between different force components (assuming that we do not measure these independently). However, this may be circumvented if we take the 2-norm squared as our cost function,

$$\begin{aligned} f(\{M_j\}, \{\sigma_{i,j}\}) &= \mathbb{E} \left( \left\| \frac{d\tilde{E}}{d\mathbf{R}} - \frac{dE}{d\mathbf{R}} \right\|^2 \right) \\ &= \mathbb{E} \left( \sum_i \left( \frac{d\tilde{E}}{dR_i} - \frac{dE}{dR_i} \right)^2 \right) \\ &= \sum_i \epsilon_i^2 = \sum_{i,j} \frac{\sigma_{i,j}^2}{M_j}. \end{aligned} \quad (54)$$

We finally write  $f(\{M_j\}, \{\sigma_{i,j}\}) = \epsilon^2$ , where  $\epsilon$  is the RMS error in our final force vector.

The advantage of targeting the norm of the error vector for importance sampling is not just that we can allocate different numbers of shots to different gradient components depending on their relative need, but that we can account for basis rotations  $Q_j$  that allow for multiple measurements. Substituting Eq. (54) into Eq. (53) and replacing the true deviation  $\sigma_{i,j}$  with our estimate  $\bar{\sigma}_{i,j}$  yields the Lagrangian

$$\mathcal{L} = \sum_j \frac{\sum_i \bar{\sigma}_{i,j}^2}{M_j} + \lambda \left( \sum_j M_j - M \right). \quad (55)$$

Minimizing with respect to  $M_j$  and solving for  $M_j$  yields a shot allocation with respect to  $\lambda$ , which may be simplified by enforcing our constraint  $\sum_j M_j = M$ ,

$$M_j = \frac{\sqrt{\sum_i \bar{\sigma}_{i,j}^2}}{\sqrt{\lambda}} = M \frac{\sqrt{\sum_i \bar{\sigma}_{i,j}^2}}{\sum_{j'} \sqrt{\sum_i \bar{\sigma}_{i,j'}}}. \quad (56)$$

Resubstituting this into our definition of  $\epsilon^2$  and solving for  $M$  then achieves a relatively compact result,

$$M \leq \epsilon^{-2} \Gamma_2^{(\text{par})}, \quad \Gamma_2^{(\text{par})} = \left( \sum_j \sqrt{\sum_i \bar{\sigma}_{i,j}^2} \right)^2. \quad (57)$$

In Appendix C 4 we repeat this calculation to find a bound on the measurement count required to estimate the error vector  $\epsilon$  to constant 1-norm instead of 2-norm.

It is instructive here to consider the effect of parallelization; what do we gain from the ability to use one basis rotation  $Q_j$  to measure components  $D_{i,j}$  of multiple force operators? This is important as this ability is lost in schemes such as low-rank factorization, where basis rotations to diagonalize factors from  $\frac{dH}{dR_i}$  and  $\frac{dH}{dR_q}$  cannot be made to easily overlap while keeping all operators low rank [90]. This can be studied by replacing  $M_j \rightarrow M_{i,j}$ , and performing the same Lagrangian minimization as before. The effect of this minimization can be immediately written down, as we are effectively losing the  $i$  index from the second sum in Eq. (57) and replacing the  $j$

index by a pair  $(i, j)$ . Thus, we can write

$$M \leq \epsilon^{-2} \Gamma_2^{(\text{sep})}, \quad \Gamma_2^{(\text{sep})} = \left( \sum_{i,j} \bar{\sigma}_{i,j} \right)^2. \quad (58)$$

As  $\sum_j x_j \geq \sqrt{\sum_j x_j^2}$ , parallelization is clearly always favorable when possible (which we expect). The gain in efficiency going from Eqs. (58) to (57) depends on how well parallel measurements can be grouped. The case with the largest difference in efficiency is when a set of basis rotations can be chosen for all  $3N_a$  force operators such that the magnitude of all errors in each component are roughly equal for each rotation, i.e.,  $\bar{\sigma}_{i,j} \sim \bar{\sigma}_j$ . In this case, we have

$$\Gamma_2^{(\text{sep})} = 3N_a \Gamma_2^{(\text{par})}. \quad (59)$$

However, in a real setting the asymptotic gain may be significantly smaller.

We can also consider the gain obtained from importance sampling in the parallel estimation case. This will be useful to predict the improvement that might be gained from importance sampling in methods where this is not natively performed. In the absence of importance sampling, we replace  $M_j \rightarrow \frac{M}{N_r}$ , where  $N_r$  is the total number of basis rotations ( $N_r = \sum_j 1$ ). The 2-norm of the error in Eq. (54) then becomes

$$\epsilon^2 \leq \frac{N_r}{M} \sum_{i,j} \bar{\sigma}_{i,j}^2, \quad (60)$$

and rearranging yields

$$M \leq \epsilon^{-2} \Gamma_2^{(\text{par, n.i.})}, \quad \Gamma_2^{(\text{par, n.i.})} = N_r \sum_{i,j} \bar{\sigma}_{i,j}^2. \quad (61)$$

As one would expect, in the limit that  $\bar{\sigma}_{i,j} = \bar{\sigma}$  Eqs. (61) and (57) are identical. However, when  $\sum_i \bar{\sigma}_{i,j}^2$  varies significantly as a function of  $j$ , the gain can be up to a factor of  $N_r$ ; the number of basis rotations used. As full tomography of the fermionic 2-RDM requires  $N_r \sim N_a^2$  (and naive tomography  $N_r \sim N_a^4$ ), this can be a significant gain.

### C. Fermionic shadow tomography

An alternative method for choosing basis rotations  $Q_j$  and allocating shots  $M_j$  is to choose them at random. This idea has been recently formalized by the notion of classical shadows [54]. Here, one considers the action of randomly drawing a basis rotation  $Q_j$  from an ensemble  $\mathcal{Q}$ , measuring in the computational basis, and observing basis state  $|b_j\rangle$ . One could in principle now invert  $Q_j$  on the measured state  $|b_j\rangle$  to give

a new state  $Q_j^\dagger |b_j\rangle \langle b_j| Q_j$ . Assuming that the ensemble  $\mathcal{Q}$  is informationally/tomographically complete (i.e., that every marginal of  $\rho$  is measured by at least one element of  $\mathcal{Q}$ ), the map

$$\mathcal{M} : \rho \rightarrow \mathbb{E}[Q_j^\dagger |b_j\rangle \langle b_j| Q_j] \quad (62)$$

is invertible. Moreover, the expectation value of the inverse map  $\mathcal{M}^{-1}$  across sampled rotations  $Q_j$  and consequently measured states  $|b_j\rangle$  must be the initial state

$$\mathbb{E}[\mathcal{M}^{-1}(Q_j^\dagger |b_j\rangle \langle b_j| Q_j)] = \rho. \quad (63)$$

Given a finite set of basis rotations  $Q_j$  and subsequent measurements  $|b_j\rangle$ , this gives an estimator for  $\frac{dE}{dR_i} = \text{Tr}[\psi^\dagger \psi \frac{dH}{dR_i}]$  [54,60]

$$\widehat{\frac{dE}{dR_i}} = \mathbb{E} \left\{ \text{Trace} \left[ \frac{dH}{dR_i} \mathcal{M}^{-1}(Q_j^\dagger |b_j\rangle \langle b_j| Q_j) \right] \right\}. \quad (64)$$

The key advantages to this method are that the ensemble  $\mathcal{Q}$  may be easier to design than a specific set of rotations  $R_j$ , and by averaging over the entire ensemble we may reduce the covariance between different terms. To suppress the tails on the distribution of the estimator  $\widehat{\frac{dE}{dR_i}}$  and achieve optimal scaling a median-of-means technique was originally used in [54], but it was shown in [60] that this is unnecessary for fermionic systems. A provably optimal choice of  $\mathcal{Q}$  to estimate arbitrary 2-RDM elements is the ensemble  $\mathcal{Q}_{\text{FGU}}$  of fermionic Clifford Gaussian unitaries described in Sec. III A. We label the corresponding channel  $\mathcal{M}_{\text{FGU}}$ .

The variance of the above estimator may be calculated by representing the force operator in the algebra generated by the Majorana operators  $\gamma_p$  [Eq. (46)],

$$\frac{dH}{dR_i} = \sum_{k=1}^2 \sum_{\mu \in C(2N, 2k)} f_\mu^{(i)} \Gamma_\mu, \quad \Gamma_\mu = i^k \gamma_{\mu_1} \dots \gamma_{\mu_{2k}}, \quad (65)$$

where  $C(2N, 2k)$  is the set of all possible combinations of  $2k$  elements drawn from  $\{1, \dots, 2N\}$ . With this defined, the variance on the estimator constructed from a single choice of basis rotation  $Q_j$  and observation of  $|b_j\rangle$  is calculated in [60] to be

$$\text{Var}_{\text{FS}} \left( \frac{dE}{dR_i} \right) = \sum_{k=1}^2 \sum_{\mu \in C(2N, 2k)} \|f_\mu^{(i)} \Gamma_\mu\|_{\text{FGU}}^2 - \text{Tr}(F_i \rho)^2, \quad (66)$$

where here  $\|\cdot\|_{\text{FGU}}$  is the shadow norm [54] under the fermionic Gaussian Clifford ensemble

$$\|O\|_{\text{FGU}}^2 = \max_{\text{state } \rho} \left\{ \mathbb{E}_{W \sim \mathcal{Q}_{\text{FGU}}} \sum_{b \in \{0,1\}^n} \langle b|W \rho W^\dagger |b\rangle \langle b|W \mathcal{M}^{-1}(O) W^\dagger |b\rangle^2 \right\}. \quad (67)$$

The shadow norm of a product of  $2k$  Majorana operators in this ensemble was calculated in [60] to be  $\binom{2n}{2k} / \binom{n}{k}$ . Thus, by the central limit theorem, the variance of the estimator  $\widehat{\frac{dE}{dR_i}}$  after  $M$  different rotations  $Q_j$  and measurements is

$$\text{Var}_{\text{FS}} \left( \frac{dE}{dR_i} \right) = \frac{1}{M} \left\{ \sum_{k=1}^2 \sum_{\mu \in C(2N, 2k)} \binom{2N}{2k} \binom{N}{k}^{-1} |f_\mu^{(i)}|^2 - \text{Tr} \left( \frac{dH}{dR_i} \rho \right)^2 \right\}. \quad (68)$$

Note that unlike other methods where one must take a bound on the variance of individual estimators, Eq. (68) is exact. It is also significantly easier to calculate than the variance on other estimation methods, as it does not require access to higher-order correlators that come with expectation values of  $H^2$ .

We now extend the above analysis to estimate the error in the 2-norm  $\epsilon_2$  [see Eq. (54)] of the force vector  $\frac{dE}{d\mathbf{R}}$ . As we are drawing  $M$  basis rotations from our distribution at random, this time we do not need to optimize the distribution of our measurements via importance sampling. (Importance sampling over shadow tomography may be introduced by locally biasing the classical shadows [93], which has been seen to yield significant improvements for Hamiltonian tomography [60].) As we do not encounter covariances between terms when calculating the 2-norm (see Sec. III B), we have immediately that

$$\epsilon_2^2 = \frac{1}{M} \sum_i \text{Var}_{\text{FS}} \left( \frac{dE}{dR_i} \right), \quad (69)$$

and so to bound  $\epsilon_2^2 < \epsilon^2$  requires that we set

$$M = \epsilon^{-2} \Gamma_2^{(\text{FS})}, \quad \Gamma_2^{(\text{FS})} = \sum_i \text{Var}_{\text{FS}} \left( \frac{dE}{dR_i} \right). \quad (70)$$

#### D. Numerical results

We now attempt to summarize and estimate the cost of the different optimizations made in this section for NISQ tomography of force vectors. The cost of estimating force vectors depends critically on the studied system. In this section we consider two different scenarios; one where we make the assumption that our force operators are relatively uniformly distributed across our system (allowing us to make general asymptotic estimates), and a set of numerical cost estimates for hydrogen chains of varying length, calculated in STO-6G using localized orbitals, see Appendix B 2. Although near optimal methods of grouping fermionic operators are known [20], to simplify the results in this paper we choose a naive grouping to compare the parallelized vs serial importance sampling discussed in Sec. III B. That is, we consider a set of rotations  $Q_j$  that measure independent Pauli terms  $D_{i,j} = h_{i,j} P_j$ , where  $\frac{dH}{dR_i} = \sum_j h_{i,j} P_j$  following a Jordan-Wigner transformation. We compare this in turn to results obtained for fermionic shadow tomography and to results for a low-rank factorization  $\frac{dH}{dR_i} = \sum_l W^{(i,\ell)^2}$  [Eq. (51)].

To avoid the need to calculate expectation values of four-body terms (which are required for exact variance estimates using the methods of Sec. III B), we use standard methods for upper bounding the variance contributions  $\sigma_{i,j}^2 \leq \bar{\sigma}_{i,j}^2$  instead. Our variance bound for naive measurements is very simple;  $\sigma_{i,j} \leq h_{i,j}$ . Note that this implies  $\Gamma_2^{(\text{sep})}$  is just the square of the induced 1-norm of the force operator (in a qubit representation). By comparison, for the basis rotation grouping method, we take the worst-case bound on the variance for each given operator

$$\sigma_{i,\ell}^2 \leq \frac{r_{i,\ell}^2}{4}, \quad (71)$$

where  $r_{i,\ell} = |\lambda_{i,\ell}^{(\max)} - \lambda_{i,\ell}^{(\min)}|$  is the range of the spectrum of the corresponding factor  $\sum_{\sigma,\sigma'} \sum_{pq} f_p^{(\ell)} f_q^{(\ell)} n_{p,\sigma} n_{q,\sigma'}$  constrained to the appropriate particle number sector. In Fig. 3, we plot the resulting scaling coefficients  $\Gamma$  for these techniques, alongside the classical fermionic shadow scaling coefficient from Sec. III C. As the bounds used for different methods differ in their tightness, it is not possible to make a direct comparison between the lines in different plots. Instead, for comparison we plot the equivalent scaling factors for the Hamiltonians of the same system, allowing us to compare the cost of energy and derivative estimation. As each line is approximately straight on a log-log plot, we may extract scaling coefficients by a linear fit, which we report in Table III. We contrast this in the same table with an asymptotic analysis under the assumption that all forces are the same magnitude. We see that, the shift from separate to parallel force measurement in the naive case (left plot) decreases the cost of estimation by a factor  $N_a^{0.2} \ll N_a$ . When parallelized, we further observe that the cost to estimate the force operator to a target precision (in Ha/Å) is roughly the same the cost to estimate energies to a target precision (in Ha). Similar results are found for the shadow tomography case (b). As the error required on forces in molecular dynamics (in the given units) is roughly 20 times larger than chemical accuracy (Sec. II C), we suggest that a single-shot force estimation using shadow tomography may be already a factor 100 cheaper than the corresponding energy estimation on a reasonably-sized system. However, semiclassical molecular dynamics simulations typically requires millions of such estimations, which presents a significant additional multiplicative cost. Similarly, we see that the scaling of the cost of estimating forces via basis rotation grouping (c) is smaller for force operators as compared to the Hamiltonian. We note, however, that the here presented bounds are not tight (see [19] for a comparison) and that it is an open question how these numbers change when using the explicit variances  $\sigma_{i,j}$  of the ground state.

Previous results [42,43,49] have focused their analysis on the scaling of the quantum circuits required to compute gradients without considering the contribution coming from the properties of the gradient operator and without performing any optimization of the measurements. In contrast, here we have computed bounds on the number of measurements required while also optimizing the measurement number with three different methods. For this reason, performing a direct quantitative comparison between this paper and previous results is not possible.

## IV. BLOCK ENCODING HAMILTONIANS AND FORCE OPERATORS

### A. General block encoding of the Hamiltonian and force operator

In the next section, we will present several methods to calculate energy gradients using a fault-tolerant quantum computer. These methods all require access to block encodings of the Hamiltonian and derivative operators. Block encoding is a method of encoding a nonunitary operator on a quantum device; one block-encodes an operator  $O$  by adding additional qubits to a system and finding a unitary  $U$  that

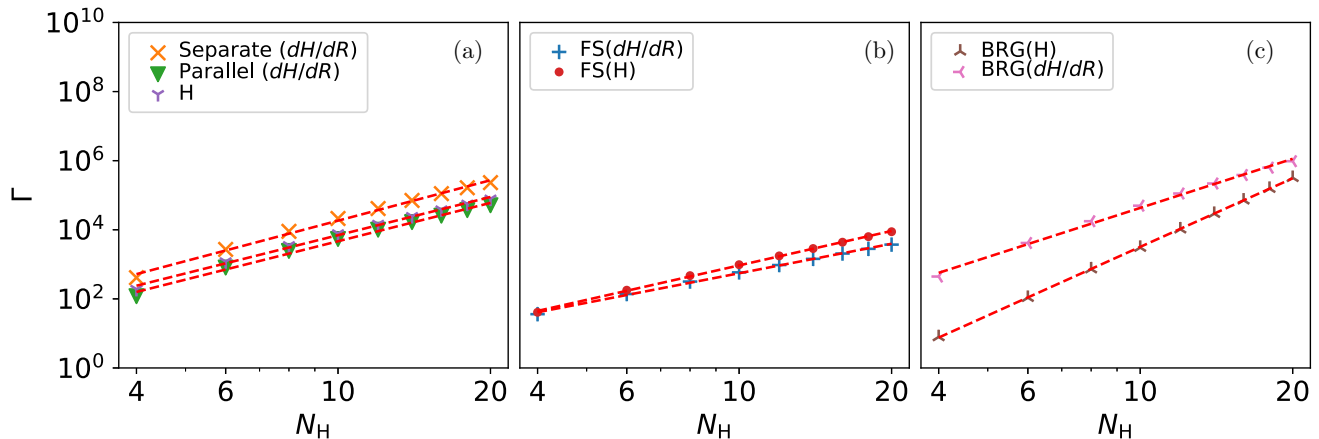


FIG. 3. Numerical evaluation of upper bounds on the cost of force estimation for hydrogen chains up to  $N_H = 20$  atoms with spacing of 0.74084 Å calculated in STO-6G using localized orbitals. We show the upper bounds of the Hamiltonian and the force operators using (a) Pauli measurements, (b) fermionic shadows, and (c) the basis rotation grouping technique.

contains  $O$  in the block corresponding to the  $|0\rangle$  state of the additional qubits. That is, if

$$U = \begin{pmatrix} u_0 & u_1 & \dots \\ u_1^\dagger & \ddots & \\ \vdots & & \end{pmatrix}, \quad (72)$$

where the subscript on  $u_i$  denotes the computational basis state of the additional qubits, we require

$$\left\| u_i - \frac{1}{\lambda_O} O \right\|_\infty \leq \epsilon. \quad (73)$$

This encoding typically requires the rescaling factor  $\lambda_O \gg 1$ , as subblocks of a unitary matrix must have eigenvalues with norm  $\leq 1$ . These rescaling factors typically then appear as multiplicative costs for implementing a given algorithm. For example, one can calculate immediately given Eq. (73) that

$$\left| \langle \psi | \otimes \langle 0 | U | 0 \rangle \otimes | \psi \rangle - \frac{1}{\lambda_O} \langle \psi | O | \psi \rangle \right| \leq \epsilon. \quad (74)$$

In this paper, we use  $\lambda_H$  and  $\lambda_{F_i}$  respectively to refer to the rescaling factors for a Hamiltonian and force operator (or  $\lambda_F$  to refer to the rescaling factor averaged across all forces). The cost of implementing these block encodings as quantum circuits is stated as  $T_H$  and  $T_{F,i}$  respectively (with  $T_F$  the cost averaged over all force components again). In this section, we give circuit implementations of these block encodings, and the corresponding  $T_H$ ,  $T_{F,i}$ ,  $\lambda_H$ , and  $\lambda_{F,i}$  costs. These differ depending on the type of basis set used to solve the electronic

structure problem, and whether the problem is solved in first or second quantization; we will give results for multiple commonly considered systems.

The most commonly used methods to block encode an arbitrary operator  $O$  are linear combination of unitaries (LCU) methods [94]. We give a brief overview of this class of techniques here. LCU methods involve writing  $O$  as a linear combination of unitary operators

$$O = \sum_i h_i P_i, \quad P_i^\dagger P_i = I, \quad h_i \geq 0. \quad (75)$$

The condition that  $h_i$  be real and positive can be accounted for by placing any complex phase on the operator  $P_i$  (as  $P_i e^{i\phi}$  is unitary whenever  $P_i$  is unitary). The LCU block encoding then requires two oracles; a PREPARE oracle that constructs the state

$$\text{PREPARE}|0\rangle = \frac{1}{\sqrt{\sum_i h_i}} \sum_i \sqrt{h_i} |i\rangle \quad (76)$$

on an additional LCU register, and the SELECT oracle defined by

$$\text{SELECT} = \sum_i |i\rangle \langle i| \otimes P_i. \quad (77)$$

One can then confirm that given states  $|\psi_1\rangle, |\psi_2\rangle$  on the system register,

$$\begin{aligned} & \langle \psi_1 | \otimes \langle 0 | \text{PREPARE}^\dagger \text{SELECT PREPARE} | 0 \rangle \otimes | \psi_2 \rangle \\ &= \frac{1}{\sum_i h_i} \langle \psi_1 | O | \psi_2 \rangle, \end{aligned} \quad (78)$$

TABLE III. Comparison of the asymptotic scaling for the case that all force operators are equal ( $F_i = F \forall i$ ) and the empirical scaling inferred from Fig. 3. The value  $\beta$  refers to the one-norm of the coefficients of the force operator after a Jordan-Wigner transformation,  $\beta_{\text{FS}} = \text{Var}_{\text{FS}}(F)$  and  $\beta_{\text{BRG}} = (\sum_i \text{Var}(F_i)^{1/2})^2$ .

Method	Asymptotic scaling ( $F_i = F, \forall i$ )	Empirical scaling (H chain)
Separate measurements	$\mathcal{O}(N_a^2 \beta (N_a) \epsilon^{-2})$	$\mathcal{O}(N_a^{3.89 \pm 0.1} \epsilon^{-2})$
Parallel measurements	$\mathcal{O}(N_a \beta (N_a) \epsilon^{-2})$	$\mathcal{O}(N_a^{3.69 \pm 0.11} \epsilon^{-2})$
Fermionic shadows	$\mathcal{O}(N_a \beta_{\text{FS}} (N_a) \epsilon^{-2})$	$\mathcal{O}(N_a^{2.84 \pm 0.05} \epsilon^{-2})$
Basis rotation grouping	$\mathcal{O}(N_a^2 \beta_{\text{BRG}} (N_a) \epsilon^{-2})$	$\mathcal{O}(N_a^{4.72 \pm 0.10} \epsilon^{-2})$

as required for this to be a block encoding, with a rescaling factor

$$\lambda = \sum_i h_i. \quad (79)$$

One has a great degree of freedom in choosing the decomposition [Eq. (75)] and constructing the PREPARE and SELECT oracles. Optimizing these is a critical requirement to lower fault-tolerant quantum costs.

### B. Block encoding in second quantization with local basis sets

The critical factor in block-encoding Hamiltonians or derivative operators with a local basis is the requirement to transfer all of the data about the operator onto the device. A generic two-body fermionic operator requires  $O(N^4)$  pieces of information to describe, and in the absence of additional structure this bounds the cost of block encoding below. Luckily, the electronic structure problem has additional structure that can be used to reduce this cost. As discussed briefly in Sec. III A, the two-body electron tensor may be factorized as the product of lower-order tensors that may be expressed with fewer indices. Two of the most well-known factorization methods are low-rank factorization [12] and tensor hypercontraction [13]. Both of these methods incur additional truncation errors, but numerically these have been shown to converge at a much lower cost than the  $O(N^4)$  bound above.

Tensor hypercontraction holds the current record as the best scaling method for block-encoding Hamiltonians. We can write the THC form of the Coulomb operator as

$$V \approx G = \frac{1}{2} \sum_{\alpha, \beta \in \{\uparrow, \downarrow\}} \sum_{p, q, r, s=1}^{N/2} G_{pqrs} a_{p, \alpha}^\dagger a_{q, \alpha} a_{r, \beta}^\dagger a_{s, \beta},$$

$$G_{pqrs} = \sum_{\mu, \nu=1}^M \chi_p^{(\mu)} \chi_q^{(\mu)} Z_{\mu\nu} \chi_r^{(\nu)} \chi_s^{(\nu)}, \quad (80)$$

where  $\chi$  is an  $M \times N$  dimension matrix and  $\zeta$  is an  $M \times M$  dimension matrix. Crucially, for the electronic structure problem it is possible to truncate  $M = O(N \text{polylog}(1/\epsilon_{\text{THC}}))$  while achieving a truncation error  $\epsilon_{\text{THC}}$ . Reference [13] demonstrated a method for block encoding  $H/\lambda_{\tilde{H}}$  using a quantum walk with gate complexity  $T_H \in \tilde{O}(M) = \tilde{O}(N)$ .

We now argue that the same method can be used to block encode the force operator, with the same scaling. One can represent the derivative of the THC Hamiltonian with respect to nuclear coordinates as

$$\frac{d}{dR_i} G_{pqrs} = \sum_{\mu, \nu=1}^M \left( \frac{d}{dR_i} \chi_p^{(\mu)} \right) \chi_q^{(\mu)} \zeta_{\mu\nu} \chi_r^{(\nu)} \chi_s^{(\nu)}$$

$$+ \left( \frac{d}{dR_i} \chi_q^{(\mu)} \right) \chi_p^{(\mu)} \zeta_{\mu\nu} \chi_r^{(\nu)} \chi_s^{(\nu)}$$

$$+ \chi_p^{(\mu)} \chi_q^{(\mu)} \zeta_{\mu\nu} \left( \frac{d}{dR_i} \chi_r^{(\nu)} \right) \chi_s^{(\nu)}$$

$$+ \chi_p^{(\mu)} \chi_q^{(\mu)} \zeta_{\mu\nu} \left( \frac{d}{dR_i} \chi_s^{(\nu)} \right) \chi_r^{(\nu)}$$

$$+ \chi_p^{(\mu)} \chi_q^{(\mu)} \left( \frac{d}{dR_i} \zeta_{\mu\nu} \right) \chi_r^{(\nu)} \chi_s^{(\nu)}, \quad (81)$$

We see that the dimension of these operators is not increased; however, there are now four types of tensors in the expression ( $\chi$ ,  $\zeta$ ,  $\frac{d\chi}{dR_i}$  and  $\frac{d\zeta}{dR_i}$ ) rather than just two ( $\chi$ ,  $\zeta$ ). However, in order to pursue the same strategy for realizing the block encoding that is employed by [13], it is necessary to yet define two more types of tensor  $\chi^{(+)} = \frac{d\chi}{dR_i} + \chi$  and  $\chi^{(-)} = \frac{d\chi}{dR_i} - \chi$ . With these definitions we can rewrite the expression as

$$\frac{d}{dR_i} G_{pqrs} = \frac{1}{2} \sum_{\mu, \nu=1}^M (\chi_p^{(\mu)})^{(+)} (\chi_q^{(\mu)})^{(+)} \zeta_{\mu\nu} \chi_r^{(\nu)} \chi_s^{(\nu)}$$

$$- (\chi_p^{(\mu)})^{(-)} (\chi_q^{(\mu)})^{(-)} \zeta_{\mu\nu} \chi_r^{(\nu)} \chi_s^{(\nu)}$$

$$+ \chi_p^{(\mu)} \chi_q^{(\mu)} \zeta_{\mu\nu} (\chi_r^{(\nu)})^{(+)} (\chi_s^{(\nu)})^{(+)}$$

$$- \chi_p^{(\mu)} \chi_q^{(\mu)} \zeta_{\mu\nu} (\chi_r^{(\nu)})^{(-)} (\chi_s^{(\nu)})^{(-)}$$

$$+ 2\chi_p^{(\mu)} \chi_q^{(\mu)} \left( \frac{d}{dR_i} \zeta_{\mu\nu} \right) \chi_r^{(\nu)} \chi_s^{(\nu)}. \quad (82)$$

We see that this expression also involves five tensors, ( $\chi$ ,  $\zeta$ ,  $\chi^{(+)}$ ,  $\chi^{(-)}$ ,  $\frac{d}{dR_i} \zeta$ ). There are now five terms instead of one but we see that within each of these terms, the two tensors with the  $\mu$  index are always the same and the two tensors with the  $\nu$  index are always the same. This is a critical requirement for the following step so that creation and annihilation operators transform in the same way. Associated with each of the three tensors related to  $\chi$ , is the following projection of the fermionic ladder operators into three different larger auxiliary bases:

$$c_{\mu, \sigma}^\dagger = \sum_{p=1}^{N/2} \chi_p^{(\mu)} a_{p, \sigma}^\dagger, \quad c_{\mu, \sigma} = \sum_{p=1}^{N/2} \chi_p^{(\mu)} a_{p, \sigma}, \quad (83)$$

$$c_{\mu, \sigma}^{\dagger, (+)} = \sum_{p=1}^{N/2} (\chi_p^{(\mu)})^{(+)} a_{p, \sigma}^\dagger, \quad c_{\mu, \sigma}^{(+)} = \sum_{p=1}^{N/2} (\chi_p^{(\mu)})^{(+)} a_{p, \sigma}, \quad (84)$$

$$c_{\mu, \sigma}^{\dagger, (-)} = \sum_{p=1}^{N/2} (\chi_p^{(\mu)})^{(-)} a_{p, \sigma}^\dagger, \quad c_{\mu, \sigma}^{(-)} = \sum_{p=1}^{N/2} (\chi_p^{(\mu)})^{(-)} a_{p, \sigma}, \quad (85)$$

where these creation and annihilation operators act on a larger space of  $2M$  spin orbitals rather than  $N$  spin orbitals. Without loss of generality, we are taking the  $\chi^{(\mu)}$ ,  $(\chi^{(\mu)})^{(+)}$ , and  $(\chi^{(\mu)})^{(-)}$  to be normalized vectors for each  $\mu$  (because constant factors can be absorbed into  $\zeta$ ). Using this, we rewrite the full force operator as

$$\frac{d}{dR_i} G = \frac{1}{4} \sum_{\alpha, \beta \in \{\uparrow, \downarrow\}} \sum_{\mu, \nu=1}^M \left( (n_{\mu, \alpha}^{(+)} n_{\nu, \alpha} - n_{\mu, \alpha}^{(-)} n_{\nu, \alpha} + n_{\mu, \alpha} n_{\nu, \alpha}^{(+)} - n_{\mu, \alpha} n_{\nu, \alpha}^{(-)}) \zeta_{\mu\nu} + 2n_{\mu, \alpha} n_{\nu, \alpha} \frac{d}{dR_i} \zeta_{\mu\nu} \right) \quad (86)$$

where

$$n_{\mu, \alpha} = c_{\mu, \alpha}^\dagger c_{\mu, \alpha}, \quad n_{\mu, \alpha}^{(+)} = c_{\mu, \alpha}^{\dagger, (+)} c_{\mu, \alpha}^{(+)}, \quad n_{\mu, \alpha}^{(-)} = c_{\mu, \alpha}^{\dagger, (-)} c_{\mu, \alpha}^{(-)}, \quad (87)$$

are the number operators in the auxiliary bases defined earlier. Having reduced the Hamiltonian to this diagonal form, it is



now clear that one can use the same block encoding strategy as that defined in [13] with minimal additional modification. In particular, because there are now five different terms, we will need several additional ancilla bits that flag which term we mean to block encode. We will need a slightly larger QROM because we now need to access five tensors rather than just four. However, the  $\zeta$  tensors, which are dimension  $M \times M$  tend to dominate the cost since the  $\chi$  tensors only have dimension  $M \times N$ . Furthermore, the Toffoli cost of the QROM increases only as the square root of the database size. Thus, the quantum walk will have a Toffoli cost that is increased by roughly a factor of  $T_F \sim \sqrt{2}T_H$ . The rescaling factor should now be defined as

$$\lambda_{F_i} = \sum_{\mu, \nu} \left( |\zeta_{\mu\nu}| + \frac{1}{2} \left| \frac{d}{dR_i} \zeta_{\mu\nu} \right| \right). \quad (88)$$

Using this method,  $\lambda_{F_i}$  will be larger than the original  $\lambda_H$  by at least a factor of two. We do not believe  $\lambda_{F_i}$  will be asymptotically larger than  $\lambda_H$  because we would expect the  $\frac{d}{dR_i} \zeta$  values to generally be smaller than the  $\zeta$  values. Thus, we have shown that the block encodings of Ref. [13] also work for the force operator, with the same asymptotic complexity.

Our expectation is that this procedure would be suboptimal and that a better approach would be to use *exactly* the same procedure as [13] in order to block encode the force operator. By this, we mean that one could apply THC directly to the coefficients  $G_{pqrs}$ , rather than take the derivatives of the THC tensors. Then, one ends up with a single  $\chi$  and a single  $\zeta$ , which compress the force operator, rather than a mix of the  $\chi$  and  $\zeta$  compressing the original Hamiltonian and their derivatives. Doing this will likely lead to even smaller values of  $\lambda_{F_i}$  than for the Hamiltonian simulation (perhaps even asymptotically so). However, it is more difficult to show analytically that  $M$  would be as small or smaller than the original  $M$ . We expect this to be the case due to the sparseness of the derivative operator; studying this numerically (or analytically) would be a clear target for future work. One additional complication that presents itself here is that the truncation error from factorizing the derivative operator is different to the truncation error in the Hamiltonian, which implies that the forces estimated by this approach do not quite match the energy manifold of the THC-factorized Hamiltonian. (This error is avoided when differentiating the THC-factorized Hamiltonian, as the truncation here is identical.) This implies that a lower tolerance for the truncation error on both the Hamiltonian and force operator may be necessary to make the systematic error here negligible; this would need to be properly bounded in any future work.

Although the THC algorithm is expected to be more efficient, a very similar argument can be made for the low-rank block encodings of Ref. [12]. The form of the low-rank Hamiltonian is given in Sec. III A, in Eq. (49). Reference [12] shows that one can use this form of the Hamiltonian in conjunction with qubitization [95] to block encode  $H/\lambda_H$  with complexity scaling as  $T_H \sim \mathcal{O}(\sqrt{NL})$  and in this case

$$\lambda_H = \sum_{\ell=1}^L \sum_{p,q=1}^{M_\ell} |f_p^{(\ell)} f_q^{(\ell)}| = \sum_{\ell=1}^L \left[ \sum_{p=1}^{M_\ell} f_p^{(\ell)} \right]^2. \quad (89)$$

Let us now extend this factorization to the derivative. We define  $V$  as in Eq. (49),  $V = \sum_{l=1}^L W^{(l)} W^{(l)\dagger}$ . Then, we can write the derivative as a difference of squares,

$$\begin{aligned} \frac{dV}{dR_i} &= \sum_{l=1}^L \left[ \frac{dW^{(l)}}{dR_i} W^{(l)\dagger} + W^{(l)} \frac{dW^{(l)\dagger}}{dR_i} \right] \\ &= \frac{1}{2} \sum_{l=1}^L \left\{ \left[ W^{(l)} + \frac{dW^{(l)}}{dR_i} \right] \left[ W^{(l)} + \frac{dW^{(l)}}{dR_i} \right]^\dagger \right. \\ &\quad \left. - \left[ W^{(l)} - \frac{dW^{(l)}}{dR_i} \right] \left[ W^{(l)} - \frac{dW^{(l)}}{dR_i} \right]^\dagger \right\}. \end{aligned} \quad (90)$$

The  $W^{(l)}$  (Eq. (51)) and their derivatives are one-body operators, so their sum is a one-body operator and can be block diagonalized following Ref. [12]. This gives a constant factor overhead and requires a single additional qubit to flag between two choices of basis rotations [in place of Eq. (50)],

$$\begin{aligned} U^{(+)} \left[ W^{(l)} + \frac{dW^{(l)}}{dR_i} \right] U^{(+)} &= \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p=1}^{M_l} f_p^{(+)} n_{p,\sigma}, \\ U^{(-)} \left[ W^{(l)} - \frac{dW^{(l)}}{dR_i} \right] U^{(-)} &= \sum_{\sigma \in \{\uparrow, \downarrow\}} \sum_{p=1}^{M_l} f_p^{(-)} n_{p,\sigma}. \end{aligned} \quad (92)$$

The corresponding scaling factor of the block encoding is in this case

$$\lambda_{F_i} = \sum_{l=1}^L \frac{1}{2} \left\{ \left[ \sum_{p=1}^{M_l} f_p^{(+)} \right]^2 + \left[ \sum_{p=1}^{M_l} f_p^{(-)} \right]^2 \right\}. \quad (93)$$

Assuming that  $\frac{dW^{(l)}}{dR_i}$  is a similar scale to  $W^{(l)}$ , we can expect that  $f_p^{(\pm)} \sim 2f_p^{(l)}$ , and that  $\lambda_{F_i} \sim 4\lambda_H$ . However, as noted before for the THC derivative factorization, we expect these scalings to be improved significantly by a direct factorization of the derivative operator instead of differentiating the factorization. This may yet lead to an asymptotic improvement, but with the same systematic bias as described for the THC method that comes from a different truncation error between the Hamiltonian and force operators.

### C. Block encoding in first quantization in a plane wave basis

The relatively simple structure of the electronic structure Hamiltonian in a plane wave basis implies that the cost to block encode it is far below the  $\mathcal{O}(N^4)$  bound for an arbitrary two-body operator. Block encodings of the first-quantized plane wave Hamiltonian are already well known [58], with a number of gates required scaling as

$$T_H \in \tilde{\mathcal{O}}((\log N)^2 \eta), \quad (94)$$

where  $\eta$  is the number of electrons in the system, and  $N$  is the number of plane waves. The corresponding scaling factor is similarly known to be [57]

$$\lambda_H \in \mathcal{O}(\eta N^{2/3} / \Omega^{2/3}). \quad (95)$$

We now consider how to block encode the force operator in its first-quantized plane wave form. We first rewrite Eq. (35)

slightly to place it in a linear combination of unitaries form

$$\frac{dH}{dR_{l,\alpha}} = \frac{4\pi Z_l}{\Omega} \sum_{j=1}^{\eta} \sum_{v \in G_0} \frac{|k_v^\alpha|}{\|\mathbf{k}_v\|^2} \left( \text{QFT} \sum_{p \in G_0} - i \text{sign}(k_v^\alpha) e^{i\mathbf{k}_v \cdot (\mathbf{R}_l - \mathbf{r}_p)} |p\rangle \langle p|_j \text{QFT}^\dagger \right). \quad (96)$$

Here, we split our indexing of the  $3N_a$ -dimensional nuclear coordinate vector  $\mathbf{R}$  into the nuclei index  $l$  and the position index  $\alpha$ , and write  $\mathbf{R}_l = (R_{l,x}, R_{l,y}, R_{l,z})$  for the 3-dimensional position of nuclei  $l$ . Then, we write  $k_v^\alpha$  for the projection of the 3-dimensional vector  $\mathbf{k}_v$  onto the unit vector in the direction of  $R_{l,\alpha}$ . To see that the operator inside each bracket is unitary, recall that  $|p\rangle \langle p|_j$  projects the  $j$ th electron register onto the  $p$ th basis state, but acts as the identity on all other electron registers. This implies that  $|p\rangle \langle p|_j |q\rangle \langle q|_j = \delta_{p,q} |p\rangle \langle p|_j$ , and  $\sum_p |p\rangle \langle p|_j = I$ , as required. This implies that our PREPARE operator requires us to prepare the state

$$\sqrt{\frac{4\pi Z_l}{\lambda_{F_{l,\alpha}} \Omega}} \sum_{j,v} \frac{\sqrt{|k_v^\alpha|}}{\|\mathbf{k}_v\|} |v, j\rangle, \quad (97)$$

with a rescaling constant

$$\begin{aligned} \lambda_{F_{l,\alpha}} &= \frac{4\pi \eta Z_l}{\Omega} \sum_{v \in G_0} \frac{|\mathbf{k}_v \cdot \mathbf{R}_l|}{\|\mathbf{k}_v\|^2 \|\mathbf{R}_l\|} \\ &= \mathcal{O} \left( \frac{\eta Z_l}{\Omega} \int_0^{2\pi} d\phi \int_0^\pi d\theta \int_0^{N^{1/3}} dr \frac{\Omega^{1/3} r |\cos \theta|}{r^2} r^2 \sin \theta \right) \\ &= \mathcal{O}(\eta N^{2/3} / \Omega^{2/3}). \end{aligned} \quad (98)$$

In Appendix D, we describe a circuit for the PREPARE operator that generates an initial approximation via a nested boxes approach, using inequality testing and a single amplitude amplification round to achieve  $\epsilon$  precision at a cost  $\mathcal{O}(\log N \log(N/\epsilon))$ . The SELECT operator for our block encoding takes the form

$$\begin{aligned} \text{SELECT} &= -i \text{QFT} \sum_{v,j} \text{sign}(\mathbf{k}_v^\alpha) \sum_p \\ &\quad \times e^{i\mathbf{k}_v \cdot (\mathbf{R}_l - \mathbf{r}_p)} |v, j\rangle \langle v, j| \otimes |p\rangle \langle p|_j \text{QFT}^\dagger. \end{aligned} \quad (99)$$

Normally the QFT operation would be performed on each of the  $\eta$  electron registers. However, we only need to act on the momentum register  $j$ . We therefore first swap the  $j$ th electron register into the 0th electron register conditional on  $|j\rangle$ . That can be achieved with  $\eta$  Toffolis for unary iteration [67], together with  $\eta \log N$  Toffolis for controlled swaps. At the end we invert the procedure to replace the register. The QFT on this register can be implemented with gate complexity  $\mathcal{O}((\log N)^2)$ .

To implement the phase  $e^{i\mathbf{k}_v \cdot (\mathbf{R}_l - \mathbf{r}_p)}$ , one would first subtract  $\mathbf{r}_p$  from the classically given value  $\mathbf{R}_l$ . The number of bits needed for  $\mathbf{R}_l$  can be found in the case of energy estimation to be  $\mathcal{O}(\log(N/\epsilon))$ , using Lemma 2 of Ref. [58]. Here we are estimating force rather than energy, but the problem is sufficiently similar that the same scaling is needed. The complexity of the subtraction is therefore  $\mathcal{O}(\log(N/\epsilon))$ . The components of both  $\mathbf{R}_l$  and  $\mathbf{r}_p$  would be given in two's

complement, then one would convert the difference to signed integers.

The phasing can then be performed by using a phase gradient state [96,97]. Let us assume that the  $|v\rangle$  register contains three separated components  $|v_x, v_y, v_z\rangle$  written in binary form  $k_v^\alpha = \frac{2\pi v_\alpha}{\Omega^{1/3}}$  for  $\alpha = x, y, z$ . Then each bit of each component of  $v$  can be used to control addition of a component of  $\mathbf{R}_l - \mathbf{r}_p$  into the phase gradient register. There are sign bits of both, since they are given as signed integers. The product of the sign bits can be used to give an overall sign, and this sign can be used to control addition versus subtraction of  $\mathbf{R}_l - \mathbf{r}_p$  without further non-Clifford cost. Since there are  $\mathcal{O}(\log N)$  bits of  $v$  and  $\mathcal{O}(\log(N/\epsilon))$  bits of  $\mathbf{R}_l - \mathbf{r}_p$ , the overall cost is  $\mathcal{O}(\log N \log(N/\epsilon))$  for the phasing. This is similar to the complexity of the state preparation.

Taking into account the  $\mathcal{O}((\log N)^2)$  complexity of the QFT and the  $\mathcal{O}(\eta \log N)$  complexity of swapping the momentum register, the complexity to block encode the derivative operator is

$$T_F = \mathcal{O}[\eta \log N + \log N \log(N/\epsilon)]. \quad (100)$$

## D. Numerical studies of the force operator in molecular systems

The Hamiltonian simulation cost depends on the representation of the Hamiltonian and can be, among other things, quantified by its spectral bandwidth, which we upper bound with a value  $\lambda$ . For electronic structure Hamiltonians there exist numerical insights into the upper bounds for various systems and representation techniques and with respect to their scaling when increasing the system size or the number of orbitals [12,13,15]. For the force operator, see Eq. (22), such results do not exist. In the following, we compare the values of  $\lambda$  obtained for the force operator and the Hamiltonian using two methods, the sparse method [12] and the double low-rank factorization [15]. The formula for the rescaling constant to block-encode an operator via the sparse method is

$$\begin{aligned} \lambda_1^{(\text{sparse})} &= \sum_p \left| T_{pq} + \sum_r V_{pqrr} \right|, \\ \lambda_2^{(\text{sparse})} &= \frac{1}{2} \sum_r |V_{pqrs}| \quad \text{and} \\ \lambda^{(\text{sparse})} &= \lambda_1^{(\text{sparse})} + \lambda_2^{(\text{sparse})}, \end{aligned} \quad (101)$$

where the coefficients  $T_{pq}$  and  $V_{pqrs}$  are taken from from Eq. (47) for the Hamiltonian, and using  $T_{pq}^{(F_A)}$  and  $V_{pqrs}^{(F_A)}$  from Eq. (24) for the force operators. These constants are induced 1-norms of the Hamiltonian (or force) operator, which makes them highly dependent on the choice of molecular orbital basis. As it has been observed that localizing the basis orbitals reduces the scaling for Hamiltonians in Refs. [13,98], we also use localized orbitals for all of our calculations, see also Appendix B 2. In Fig. 9, we provide the same results for canonical molecular orbitals.

Instead of directly following the method used in Sec. IV B to differentiate the double-low-rank-factorized Hamiltonian to prove a bound on the cost of block-encoding the derivative operator, in this section we directly block encode the derivative operator instead. This is practically easier, and should in principle yield a lower rescaling factor  $\lambda$ ; however, care

should be taken in practice to suppress the systematic error this produces (as described in Sec. IV B). Within the double low-rank factorization the lambdas are given by

$$\begin{aligned}\lambda_1^{(\text{DF})} &= \sum_i |t_i|, \\ \lambda_2^{(\text{DF})} &= \frac{1}{4} \sum_l^L \left( \sum_p |f_p^{(l)}| \right)^2 \quad \text{and} \quad (102) \\ \lambda^{(\text{DF})} &= \lambda_1^{(\text{DF})} + \lambda_2^{(\text{DF})},\end{aligned}$$

with  $t_i$  the eigenvalues of  $T_{pq} + \sum_r V_{pqrr}$ , cf. Eq. (49) for the Hamiltonian, and the eigenvalues of  $T_{pq}^{(F_A)} + \sum_r V_{pqrr}^{(F_A)}$  [Eq. (24)] for the force operators.

As the force is a vector with  $3N_a$  elements for a system with  $N_a$  atoms, we find  $3N_a$  lambdas. To compare the simulation cost of the Hamiltonian and the force operator, we use the mean of the lambdas of the force operators,

$$\lambda_F = \frac{\sum_{i=1}^{3N_a} \lambda_{F_i}}{3N_a}, \quad (103)$$

with  $\lambda_{F_i}$  the lambda of the  $i$ th force operator. We note that, in contrast to electronic structure Hamiltonians, performing a double low-rank factorization of the force operator yields negative eigenvalues  $w_i$  in Eq. (48). This can be accounted for in a block encoding by shifting the sign from the eigenvalue  $w_i$  to the encoding itself, so we replace the eigenvalues by their absolute values  $|w_i|$ . For all shown plots, we use localized orbitals (LMOs) from a Hartree-Fock calculation, see Appendix B 2.

In the following we provide numerical results on one-dimensional chains of hydrogen atoms with spacings of 1.4 Bohr radii (0.74084 Å), calculated in STO-6G, a system often used to infer the scaling in the thermodynamic limit [12,13]. Moreover, to provide a benchmark for systems closer to relevance for biological systems, we investigate the scaling of  $\lambda^{(\text{sparse})}$  and  $\lambda^{(\text{DF})}$  for water clusters of increasing size. Water clusters are prototypes of hydrogen bonded networks and therefore of fundamental interest, and thoroughly investigated benchmarks by experimental and theoretical methods [99,100]. Hydrogen bonds are regarded as the key to life in general, due to the directionality, cooperativity and versatility [101], being crucially involved in the chemical basis of the genetic code and all catalytic processes. The strength of hydrogen bonds is highly dependent on the local surrounding and requires accurate quantum treatment [102]. As input for the calculations, we use the data from [103], presented also in Fig. 8, which include the geometries of water clusters, which were optimized by using a TTM2.1-F *ab initio* based interaction potential together with subsequent MP2 calculations in the aug-cc-pVTZ basis.

To study the scaling of  $\lambda^{(\text{DF})}$  in the continuum limit, i.e., when increasing the basis set size, we follow [12,13] and use  $H_4$  on a square plaquette with a side length of 2 Bohr radii. We first calculate the Hamiltonian and force operators using the cc-pvtz basis, which yields  $N = 56$  spatial orbitals. To study the effect of an increasing basis set we artificially truncate the

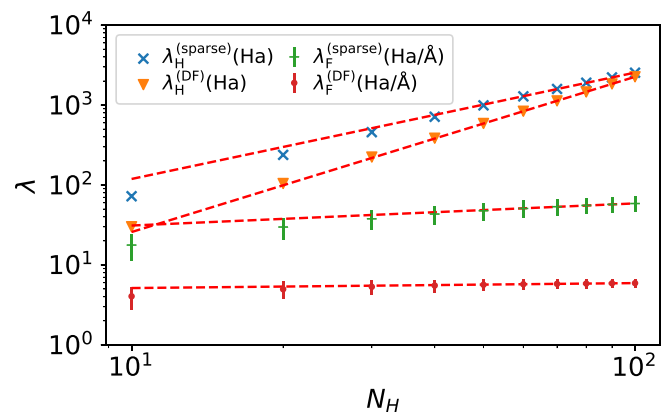


FIG. 4. Rescaling factors for the Hamiltonian ( $\lambda_H$ ) and mean rescaling factors for the force operators [ $\lambda_F$ , Eq. (103)] using the sparse method with localized molecular orbitals, and a double low-rank factorization of the force operators (as opposed to differentiating the double-low-rank-factorized Hamiltonian described in Sec. IV B) for one-dimensional hydrogen chains with up to  $N_H = 100$  atoms separated by 1.4 Bohr (0.74084 Å). All calculations were performed in STO-6G. The plot shows  $\lambda_1 + \lambda_2$  of the Hamiltonian and force operators, see Eq. (103). The fits suggest scalings of  $\lambda_H^{(\text{sparse})} = \mathcal{O}(N_H^{1.331 \pm 0.012})$ ,  $\lambda_F^{(\text{sparse})} = \mathcal{O}(N_H^{0.277 \pm 0.011})$ ,  $\lambda_H^{(\text{DF})} = \mathcal{O}(N_H^{1.942 \pm 0.002})$  and  $\lambda_F^{(\text{DF})} = \mathcal{O}(N_H^{0.062 \pm 0.004})$ . Only force operators with  $\lambda_{F_i} \neq 0$  were included in this calculation. For all fits only the last five points were taken into account. The error bars show the standard deviation of the distribution of the lambdas of single force operators  $\lambda_{F_i}$ .

two-body coefficients  $V_{pqrs}$ ,

$$V_{pqrs}^{\text{truncated}} = \sum_{pqrs=1}^{\tilde{N}} V_{pqrs} \quad (104)$$

for  $10 \leq \tilde{N} \leq 56$  and show the  $\lambda^{(\text{DF})}$  in Fig. 5.

In Fig. 4, we show  $\lambda^{(\text{sparse})}$  and  $\lambda^{(\text{DF})}$  of the force operators and Hamiltonian for the hydrogen chains of length between  $N_H = 10$  and  $N_H = 100$ . For both methods, we recognize that the force operator shows a better scaling compared to the scaling of the Hamiltonian, and indeed tends towards a constant or  $\log(N_H)$  cost. This makes sense, as we expect the force on a given nuclei in a 1D chain to scale logarithmically in the system size.

In Fig. 5(a), we show the scaling of  $\lambda^{(\text{sparse})}$  and  $\lambda^{(\text{DF})}$  for the water clusters in STO-3G with increasing system size. For technical details on the calculation we refer to Appendix E 1. Similarly to the hydrogen chain, we find a significant better scaling of the average  $\lambda$  of the force operators in comparison to the scaling of the Hamiltonian.

In Fig. 5(b), we show the scaling of  $\lambda^{(\text{DF})}$  when increasing the basis set size for  $H_4$ . In this case, the scaling of  $\lambda^{(\text{DF})}$  for the Hamiltonian and the force is comparable, which is contrary to the results in the previous section. This makes sense: increasing the basis set increases the range of local energy densities on each atom, which suggests the spectrum

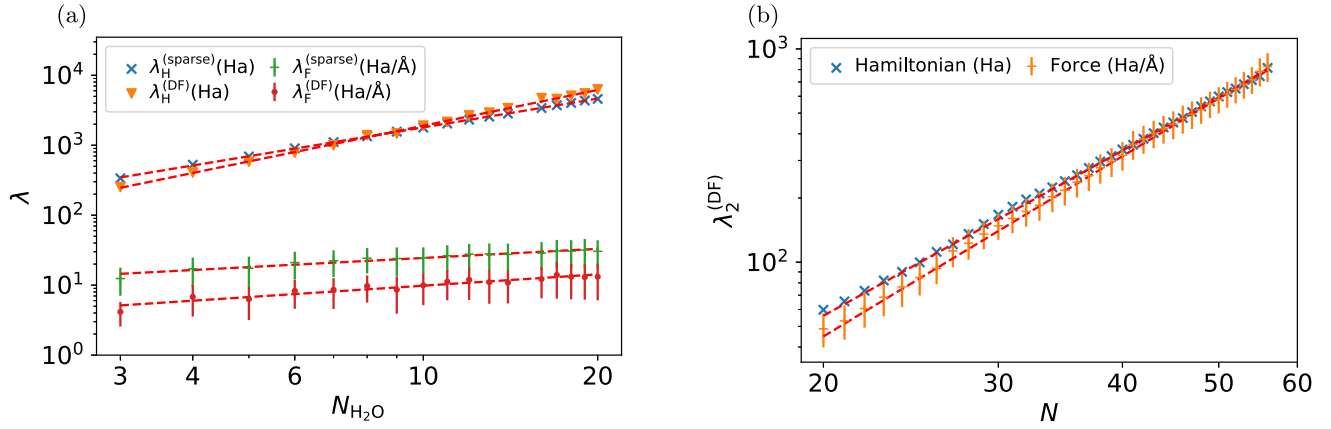


FIG. 5. (a) Rescaling factors  $\lambda$  of the Hamiltonian and force operators for the water clusters presented in Fig. 8, for the sparse method (using localized molecular orbitals, see Appendix B 2), and following a double low-rank factorization of the corresponding operators (as opposed to differentiating the double-low-rank-factorized Hamiltonian described in Sec. IV B). Fits suggest scalings of  $\lambda_H^{(\text{sparse})} = \mathcal{O}(N_{H_2O}^{1.368 \pm 0.005})$ ,  $\lambda_F^{(\text{sparse})} = \mathcal{O}(N_{H_2O}^{0.434 \pm 0.027})$ ,  $\lambda_H^{(\text{DF})} = \mathcal{O}(N_{H_2O}^{1.697 \pm 0.019})$  and  $\lambda_F^{(\text{DF})} = \mathcal{O}(N_{H_2O}^{0.539 \pm 0.042})$ . (b) The rescaling factor of double-low-rank-factorized Hamiltonians and force operators with respect to the number of spin orbitals  $N$  in an STO-6G basis for a  $H_4$  plaquette with side length of 2 Bohr radii. Fits suggest scalings of  $\lambda_H^{(\text{DF})} = \mathcal{O}(N^{2.549 \pm 0.014})$  and  $\lambda_F^{(\text{DF})} = \mathcal{O}(N^{2.749 \pm 0.03})$ . As in Fig. 4, in both plots, the error bars show the standard deviation of the distribution of the lambdas of single force operators  $\lambda_{F_i}$ .

of local operators such as derivatives should be increased also.

## V. ALGORITHMS TO COMPUTE FORCES IN FAULT TOLERANCE

We detail three different methods in this section by which one might estimate forces in a fault-tolerant setting. We first study a semiclassical finite difference calculation, where the quantum computer is called as a subroutine to estimate energies at different positions  $\mathbf{R}$ . We then study the direct estimation of forces as the expectation value of the derivative of the Hamiltonian via the Hellman-Feynman theorem, using the overlap estimation algorithm to make this estimation at the Heisenberg limit. We finally study expectation value estimation of the Hamiltonian derivative by the new expectation value estimation algorithm of [56], which re-encodes this as a separate derivative estimation problem that can be solved by the quantum derivative estimation algorithm of [104]. We estimate the asymptotic cost for each method (up to logarithmic factors) for plane waves, hydrogen chains, and small water cluster systems.

### A. Numerical differentiation of the energy by higher-order finite differences

Finite difference methods estimate gradients as a linear combination of the energies  $E(\mathbf{R})$  from different atomic configurations  $\mathbf{R}$ . Central finite differences formulas offer a quadratic advantage in the discretization error compared to backward and forward differences, so here we will focus only on the central difference forms. The simplest central finite difference formula is the first-order form, which requires energy calculations at two points  $\mathbf{R}$ ,

$$\frac{dE}{dR_i} = \frac{E(\mathbf{R} + \frac{dR \cdot \mathbf{v}_i}{2}) - E(\mathbf{R} - \frac{dR \cdot \mathbf{v}_i}{2})}{dR} + \epsilon_{fd}^{(1)}. \quad (105)$$

Here,  $\epsilon_{fd}^{(1)} \sim dR^3$  is the error due to the finite difference approximation, and  $\mathbf{v}_i$  is the unit vector along the component  $R_i$ . This can be generalized to the degree- $2m$  (central) finite difference formula [104],

$$\begin{aligned} \frac{dE}{dR_i} &= \frac{1}{dR} \sum_{\substack{l=-m \\ l \neq 0}}^m a_l^{(m)} E(\mathbf{R} + l dR \cdot \mathbf{v}_i) + \epsilon_{fd}^{(m)} \\ &:= \frac{dE^{(m)}}{dR_i} + \epsilon_{fd}^{(m)}, \end{aligned} \quad (106)$$

where the coefficients  $a_l^{(m)}$  are given by

$$a_l^{(m)} := \begin{cases} \frac{(-1)^{|l|-1}}{l} \frac{\binom{m}{|l|}}{\binom{m+|l|}{|l|}} & l \neq 0 \\ 0 & \text{otherwise} \end{cases}, \quad (107)$$

and  $\epsilon_{fd}^{(m)} \sim dR^{2m+1}$  is the degree- $2m$  finite difference error.

The easiest way to implement higher-order finite difference methods on a quantum device is to use the device as a subroutine that calculates the required energies  $E(\mathbf{R} + l dR \cdot \mathbf{v}_i)$ . This then sets a competition between the finite-difference error and the error  $\epsilon_{PE}$  from phase estimation, as the latter scales inversely in  $dR$ . Specifically,  $\epsilon_{PE} \sim dR^{-1}$  due to the division in Eq. (106). Optimizing higher-order finite difference methods on a quantum device then requires balancing these two sources of error by optimizing our choices of  $dR$  and the degree  $m$  of the finite-difference method used. It also requires that we optimize our quantum circuits for finite difference estimation: As we are repeatedly estimating eigenvalues of similar (but not identical) quantum states, under certain conditions it may be preferable to reuse our state register in between phase estimation routines. In this section we will optimize all the above, and calculate bounds on the asymptotic scaling of the resulting algorithm.

In Appendix C 2 we calculate the error  $\epsilon_{PE}$  due to quantum phase estimation on a single component  $\frac{dE}{dR_i}$ . We draw on re-

sults from [67,105,106], and then propagate this error through Eq. (106). Then, in Appendix C 4, we optimize our resource allocation to the estimation of different  $E(\mathbf{R} + l dR \cdot \mathbf{v}_i)$  via Lagrangian techniques. This yields a bound

$$\epsilon_{\text{PE}} \leq \frac{\pi \lambda_{\text{H}} 6^{3/2} m^{1/2}}{2dR T}, \quad (108)$$

where  $\lambda_{\text{H}}$  is the cost of block-encoding the Hamiltonian (Sec. IV A). In order for this bound to hold, we require that  $\lambda_{\text{H}}$  be the worst-case block-encoding cost over all points  $\mathbf{R} + l dR \cdot \mathbf{v}_i$ .

Equation (108) gives the first part of the total error for estimating the gradient with higher-order finite differences. To complete the estimation of the error we need also to upper bound the second component  $\epsilon_{\text{fd}}^{(m)}$ . To do this, we invoke Theorem 24 of [104]. This requires imposing the condition that the energy and its derivatives are bounded,

$$\left| \frac{d^k E}{dR_i^k} \right| \leq \epsilon c^k (k)^{\frac{k}{2}}, \quad \forall i, \quad (109)$$

with  $c$  a constant with units  $\text{length}^{-1}$  and  $\epsilon$  a constant with the units of energy. When this condition is satisfied, the discretization error  $\epsilon_{\text{fd}}^{(m)}$ , for one component of the gradient can be upper bounded by

$$\begin{aligned} \epsilon_{\text{fd}}^{(m)} &= \left| \frac{dE}{dR_i} - \frac{dE^{(m)}}{dR_i} \right| \leq \frac{\epsilon}{dR} \sum_{k=2m+1}^{\infty} (8 dR c m)^k \\ &= \frac{\epsilon (8 dR c m)^{2m+1}}{dR (1 - (8 dR c m))}. \end{aligned} \quad (110)$$

As the error shrinks exponentially with  $m$ , the value of  $dR$  needed to ensure that the error is at most  $\mathcal{O}(\epsilon)$  can be taken to be much larger as  $m$  increases. Since the cost of phase estimation scales inversely with  $dR$ , we will see below that using higher-order difference formulas will actually reduce the cost of phase estimation.

### 1. Optimization of the finite difference step size

It remains to optimize the remaining free parameters in our finite difference method. Substituting Eqs. (110) and (108) into  $\epsilon^2 = \epsilon_{\text{PE}}^2 + (\epsilon_{\text{fd}}^{(m)})^2$ , we obtain

$$\epsilon^2 = \epsilon_{\text{PE}}^2 + (\epsilon_{\text{fd}}^{(m)})^2 \leq \left( \frac{\pi \lambda_{\text{H}} \Gamma}{2T dR} \right)^2 + \left( \epsilon \frac{(8cm)^{2m+1} dR^{2m}}{(1 - 8mc dR)} \right)^2, \quad (111)$$

If we simply make the assumption that  $8 m c dR \leq 1/2$  and we combine the two errors quadratically, then we have the following upper bound on the total error of the numerically estimated gradient:

$$\begin{aligned} \epsilon^2 &= \epsilon_{\text{PE}}^2 + (\epsilon_{\text{fd}}^{(m)})^2 < \left( \frac{\pi \lambda_{\text{H}} 6^{3/2} (m)^{1/2}}{2T dR} \right)^2 \\ &\quad + (2\epsilon(8cm)^{2m+1} dR^{2m})^2. \end{aligned} \quad (112)$$

We can also find a second upper bound on the value of  $dR$ , by imposing that the  $\epsilon_{\text{fd}}^2$  is at most  $\epsilon^2/2$  in the upper bound of

Eq. (112) and obtain

$$dR \leq \min \left\{ \frac{1}{8cm} \left( \frac{\epsilon}{16\sqrt{2}\epsilon mc} \right)^{\frac{1}{2m}}, \frac{1}{16cm} \right\}, \quad (113)$$

whereas the second requirement is a direct consequence of the requirement that  $8 dR cm \leq 1/2$ . Substituting this into Eq. (108) will yield a comparable bound for the time.

One remaining point that needs to be addressed is the sensitivity of the estimation protocol on the ground-state preparation. In particular, the above discussion holds if the ground-state preparation success probability per point is  $1 - \mathcal{O}(1/N_a)$ . This is because we assume in the above analysis that the probability of the algorithm projecting out of the ground space at each of the  $3N_a$  estimations is asymptotically negligible. From the union bound our previous claim on the success probability implies this because  $(1 - \mathcal{O}(1/N_a))^{3N_a} \in \Omega(1)$ .

While this may appear to be a substantial drawback of the finite difference method, the ground state will sometimes only need to be prepared once for the FD gradient estimation process. This is because if the eigenvalue gap is large relative to the perturbation in the energy needed over our finite sized mesh then the ground states at two near by configurations will be nearly identical. We show this fact in Appendix C 3 and specifically claim that for any  $\Delta_1, \Delta_2 \in \mathbb{R}^{3N_a}$  satisfying  $\max(\|\Delta_1\|, \|\Delta_2\|) < (E_1(\mathbf{R}) - E_0(\mathbf{R})) / (4 \max_{\mathbf{R}', i} \|\frac{dH(\mathbf{R}')}{dR_i}\|)$  (where the max over  $\mathbf{R}'$  is taken over a ball of radius  $\max(\|\Delta_1\|, \|\Delta_2\|)$  around  $\mathbf{R}$ ), the difference between the approximate ground states  $|\tilde{\psi}_0(\mathbf{R} + \Delta)\rangle$  and the true ground state at  $\Delta = 0$  is

$$\begin{aligned} &|\langle \tilde{\psi}_0(\mathbf{R} + \Delta_1) | \tilde{\psi}_0(\mathbf{R} + \Delta_2) \rangle| \\ &\geq 1 - 3\delta - \frac{6 \max\{\|\Delta_1\|, \|\Delta_2\|\} \max_{\mathbf{R}', i} \|\frac{dH(\mathbf{R}')}{dR_i}\|}{E_1(\mathbf{R}) - E_0(\mathbf{R})} \end{aligned} \quad (114)$$

where  $E_i(\mathbf{R})$  refers to the  $i^{\text{th}}$  eigenvalue of  $H(\mathbf{R})$  and

$$\begin{aligned} &\max(\| |\tilde{\psi}_0(\mathbf{R} + \Delta_1)\rangle - |\psi_0(\mathbf{R} + \Delta_1)\rangle \|, \\ &\| |\tilde{\psi}_0(\mathbf{R} + \Delta_2)\rangle - |\psi_0(\mathbf{R} + \Delta_2)\rangle \|) \leq \delta. \end{aligned} \quad (115)$$

We will choose  $\delta$  to be in  $\Theta(\epsilon)$  since the logarithmic costs of making the state preparation accurate are negligible relative to the costs of the phase estimation step. Under these assumptions we have that the  $\delta$  contribution is negligible compared to the other two terms and so we can use the following lower

bound:

$$|\langle \tilde{\psi}_0(\mathbf{R} + \Delta_1) | \tilde{\psi}_0(\mathbf{R} + \Delta_2) \rangle| \geq 1 - \frac{12 \max\{\|\Delta_1\|, \|\Delta_2\|\} \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|}{E_1(\mathbf{R}) - E_0(\mathbf{R})}. \quad (116)$$

In order to ensure that the probability of failing to project is at most  $1/3$  during the protocol it suffices to take the failure probability of each step to be  $1/9mN_a$ . An appropriate probability can be met if we choose

$$\max\{\|\Delta_i\|\} \leq \frac{E_1(\mathbf{R}) - E_0(\mathbf{R})}{36mN_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|}. \quad (117)$$

Then, as our choices of  $\mathbf{y}$  in our finite difference algorithm are at max  $m dR$  from  $\mathbf{R}$ , a sufficient choice of  $|dR|$  to guarantee this scaling is

$$\max_i \|\Delta_i\| = m|dR| \Rightarrow |dR| \leq \frac{E_1(\mathbf{R}) - E_0(\mathbf{R})}{36m^2N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|} := \frac{\gamma}{36m^2N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|}. \quad (118)$$

This potentially creates a problem since in order to ensure that the ground state remains stable over all the measurements (and thereby making the ground-state projection cost additive rather than multiplicative) we need to have  $\|\Delta_i\|$  constant as a function of  $\epsilon$ . We can therefore choose  $\Delta_i$  to be a constant by taking

$$dR \leq \min \left\{ \frac{1}{8cm} \left( \frac{\epsilon}{16\sqrt{2}\epsilon mc} \right)^{\frac{1}{2m}}, \frac{1}{16cm}, \frac{\gamma}{36m^2N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|} \right\}. \quad (119)$$

## 2. Optimization of the finite difference order

By substituting this expression for  $dR$  in Eq. (108), and inverting in  $T$  we have that the total amount of time (in number of simulations) needed to compute a component of the gradient with error  $\epsilon$  is bounded by

$$T \leq 48 \frac{\sqrt{2}\sqrt{6} \lambda_H \pi c \sqrt{m}}{\epsilon} \times \max \left\{ \left( \frac{16\sqrt{2}\epsilon cm}{\epsilon} \right)^{\frac{1}{2m}} m^{3/2}, 2m^{3/2}, \frac{9m^2N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|}{2\gamma} \right\}, \leq 48 \frac{\sqrt{2}\sqrt{6} \lambda_H \pi c m^{5/2}}{\epsilon} \times \max \left\{ \left( \frac{16\sqrt{2}\epsilon cm}{\epsilon} \right)^{\frac{1}{2m}}, \frac{9}{2} \left( 1 + \frac{N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|}{\gamma c} \right) \right\}. \quad (120)$$

After the above choices have been made, the only variable that we have access to, impacting the number of simulation steps needed in the phase estimation is  $m$ . This is clearly

optimized in Eq. (120) by setting the two terms in the curly brackets equal

$$\frac{16\sqrt{2}\epsilon cm}{\epsilon} = \left( \frac{9}{2} \left( 1 + \frac{N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|}{\gamma c} \right) \right)^{2m}. \quad (121)$$

The solution to this equation is given by a Lambert  $W$ -function and the asymptotics of the function reveals that it suffices to choose

$$m \in \tilde{\mathcal{O}} \left( \frac{\log \left( \frac{\epsilon c}{\epsilon} \right)}{\log \left( \frac{9(1+N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|/\gamma c)}{2} \right)} \right). \quad (122)$$

This leads to the conclusion that the query complexity of computing one of the components of the gradient within error  $\epsilon$  is at most

$$T \in \tilde{\mathcal{O}} \left( \frac{\lambda_H c m^{5/2} (1 + N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|/\gamma c)}{\epsilon} \right) \subseteq \tilde{\mathcal{O}} \left( \frac{\lambda_H \log^{5/2}(\epsilon) (c + N_a \max_{\mathbf{R}', i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|/\gamma)}{\epsilon} \right). \quad (123)$$

## 3. Gradient vector estimation

In the above we have considered the cost to estimate single components of the gradient vector. It remains to convert our resulting expression for the query complexity for these single components [Eq. (123)] into one for the error on the gradient vector. This involves repeating this calculation for each of the  $3N_a$  components of the gradient vector. If we want to ensure that the error in the gradient calculation (as quantified by the 2-norm of the error vector) is at most  $\epsilon$  it suffices to take

$$\max_i \left| \frac{dE}{dR_i} - \frac{dE^{(m)}}{dR_i} \right| \leq \frac{\epsilon}{\sqrt{3N_a}}, \quad (124)$$

because

$$\left\| \frac{dE}{d\mathbf{R}} - \frac{dE^{(m)}}{d\mathbf{R}} \right\|_2 \leq \sqrt{3N_a} \max_i \left| \frac{dE}{dR_i} - \frac{dE^{(m)}}{dR_i} \right| \leq \epsilon. \quad (125)$$

(In principle this could be improved by importance sampling, however we do not believe this will provide a significant gain here.) We then choose the error in individual gradient component estimations to be  $\epsilon' \leq \epsilon/\sqrt{3N_a}$ . The asymptotic cost of the overall number of queries made in the simulation,

excluding the cost of initial state preparation, is then

$$T_{\text{FD}}\left(\frac{dE}{d\mathbf{R}}\right) \in \tilde{\mathcal{O}}\left(\frac{N_a^{3/2}\lambda_H \log^{5/2}(\epsilon)(c + N_a \max_{\mathbf{R}',i} \|\frac{dH(\mathbf{R}')}{dR_i}\|/\gamma)}{\epsilon}\right). \quad (126)$$

If we define the initial state preparation process to be an algorithm that prepares a state using no queries (such as a Hartree-Fock state) and has a probability of success  $a_0^2$  then the cost of implementing an approximate ground-state projector using [107] is in

$$T_{\text{Prep}} \in \tilde{\mathcal{O}}\left(\frac{\lambda_H \log(1/\epsilon)}{|a_0|\gamma}\right). \quad (127)$$

Recall that state preparation is only needed on average  $O(1)$  times, due to the choice of  $\|\Delta_i\|$  made in Eq. (117). This implies that the total query complexity of using the finite difference method is

$$T_{\text{FD-Tot}}\left(\frac{dE}{d\mathbf{R}}\right) \in \tilde{\mathcal{O}}\left(\lambda_H \left(\frac{N_a^{3/2} \log^{5/2}(\epsilon)(c + N_a \max_{\mathbf{R}',i} \|\frac{dH(\mathbf{R}')}{dR_i}\|/\gamma)}{\epsilon} + \frac{\log(1/\epsilon)}{|a_0|\gamma}\right)\right). \quad (128)$$

The above cost is given in the number of oracle calls to a block encoding of the Hamiltonian (Sec. IV A). To convert this into wall-clock time, we have to multiply this by the corresponding oracle cost in first or second quantization. The results of [15] show that the PREPARE operation for second quantized simulations of systems in a nonplanewave basis can be implemented using  $\tilde{\mathcal{O}}(N^2)$  Toffoli gates; however, with the use of tensor hypercontraction this becomes  $\tilde{\mathcal{O}}(N)$  asymptotically [13]. The SELECT operation can be implemented in  $\tilde{\mathcal{O}}(N)$  operations, where  $N$  is the number of spin orbitals used in each simulation [13]. Thus the overall cost of the walk operator is in  $\tilde{\mathcal{O}}(N)$ , and we have

$$\text{ToffCount}_{2\text{nd}} \in \tilde{\mathcal{O}}\left(\lambda_H N \left(\frac{N_a^{3/2} \log^{5/2}(\epsilon)(c + N_a \max_{\mathbf{R}',i} \|\frac{dH(\mathbf{R}')}{dR_i}\|/\gamma)}{\epsilon} + \frac{\log(1/\epsilon)}{|a_0|\gamma}\right)\right) \quad (129)$$

The scaling of this algorithm depends on the scaling of the value of  $\lambda_H$ . In first quantization, the cost of the simulation using qubitization is based on the discussion in Sec. IV C is in [58]

$$\text{ToffCount}_{1\text{st}} \in \tilde{\mathcal{O}}\left(\lambda_H \eta \log(N) \left(\frac{N_a^{3/2} \log^{5/2}(\epsilon)(c + N_a \max_{\mathbf{R}',i} \|\frac{dH(\mathbf{R}')}{dR_i}\|/\gamma)}{\epsilon} + \frac{\log(1/\epsilon)}{|a_0|\gamma}\right)\right) \quad (130)$$

Here the result arises from the fact that the select operation incurs a cost that is in  $O(\eta \log(N))$  and the prepare operation has a cost that is in  $O(\eta \log^2(N) + \log^3(N))$ .

As a final point, we can consider the case where we make no attempt to reuse the quantum state evaluated for the finite difference formulas. The advantage of this approach is that we do not need to reduce  $dR$  to ensure that the ground state is sufficiently stable under perturbations of the Hamiltonian across the different points of interest. However, it will often be less efficient due to the many state preparation steps needed and also it will require a set of different state. We assume below that each such state has success probability at least  $|a_i|^2$ . Repeating the same analysis given above leads to a number of oracle queries that scales as

$$T_{\text{FD}}\left(\frac{dE}{d\mathbf{R}}\right) \in \tilde{\mathcal{O}}\left(N_a^{3/2}\lambda_H \left[\frac{\log^{5/2}(\epsilon)c}{\epsilon} + \frac{\log(1/\epsilon)}{\min_i |a_i|\gamma}\right]\right). \quad (131)$$

Combining Eq. (131) with Eq. (128) yields a joint expression

$$T_{\text{FD}}\left(\frac{dE}{d\mathbf{R}}\right) \in \tilde{\mathcal{O}}\left(\lambda_H N_a^{3/2} \left[\frac{\log^{5/2}(\epsilon)c}{\epsilon} + \min\left\{\frac{\log(1/\epsilon)}{\min_i |a_i|\gamma}, \frac{N_a \max_{\mathbf{R}',i} \|\frac{dH(\mathbf{R}')}{dR_i}\| \log^{5/2}(\epsilon)}{\epsilon\gamma}\right\}\right]\right). \quad (132)$$

We examine this scaling in Sec. IV D and the results of the scalings given either empirically or analytically is given in Table IV. In all cases we assume that  $\min_i |a_i|\gamma \gg \epsilon$ . While these choices are made for simplicity, they do privilege the state preparation method given here because it means that we do not need to artificially reduce the value of  $dR$  to ensure that the state preparation costs are additive. This point is important to consider when comparing these costs with the methods in the following sections.

The analytic results given in Table IV are straightforward to verify. In the plane-wave basis, it is difficult to reason about ionic systems because of the periodic nature of the basis

functions leading to infinite energy. For this reason we focus on the case where the number of electrons in the system is equal to the nuclear charge. This implies that if we define  $Z_i$  to be the charge of nucleus  $i$  then the number of electrons can be bounded above by

$$\eta = \sum_{i=1}^{N_a} Z_i \leq N_a \mathbb{E}(|Z_i|). \quad (133)$$

Next the value of  $\lambda_H$  for such systems is given by Eq. (25) of [58] to be in  $\mathcal{O}(N^{2/3}\eta^{1/3})$  in the limit where the number of plane waves is much greater than the number of electrons.

TABLE IV. Scaling of finite difference gradient estimation for a system with  $N_a$  atoms and  $N$  orbitals and  $\eta$  electrons within the Born-Oppenheimer approximation for constant  $c$  and  $\epsilon$  as well as  $\epsilon$  error as measured by the 2-norm of the gradient vector. Numerical estimates follow from the data in Sec. IV D. The analytic results for plane waves follows from taking the volume of the unit cell  $\Omega$  to be proportional to the number of plane waves  $N$ . Note that for the hydrogen chains and water clusters we choose  $N \propto N_a$  because a minimal basis is used within these simulations. The cost of state preparation is ignored in these scalings, which is true if  $|a_0| \gamma \gg \epsilon$ .

System	Analytic scaling	Empirical scaling
First quantized plane waves	$\tilde{\mathcal{O}}(N^{2/3} N_a^{17/6} (\mathbb{E}( Z_i ))^{1/3} \epsilon^{-1})$	
Second quantized plane waves	$\tilde{\mathcal{O}}(N^3 N_a^{3/2} \epsilon^{-1})$	
Hydrogen chains (sparse)		$\tilde{\mathcal{O}}(N_a^{3.833} \epsilon^{-1})$
Hydrogen chains (double factorization)		$\tilde{\mathcal{O}}(N_a^{4.442} \epsilon^{-1})$
Water clusters (sparse)		$\tilde{\mathcal{O}}(N_a^{3.868} \epsilon^{-1})$
Water clusters (double factorization)		$\tilde{\mathcal{O}}(N_a^{4.197} \epsilon^{-1})$

Thus

$$\lambda_H \in \mathcal{O}(N^{2/3} N_a^{1/3} (\mathbb{E}(|Z_i|))^{1/3}). \quad (134)$$

This implies that the cost of performing this process within the required error (under the assumptions on  $\|\frac{dH}{dR_i}\|$ ,  $c$ ,  $\epsilon$  made above) is

$$\text{ToffCount}_{1, \text{st}} \in \tilde{\mathcal{O}}\left(\frac{N^{2/3} N_a^{17/6} (\mathbb{E}(|Z_i|))^{1/3}}{\epsilon}\right). \quad (135)$$

The case of second quantized plane wave simulations is much simpler to analyze. From Eq. (53) of [108] we find that  $\lambda_H \in \tilde{\mathcal{O}}(N^2)$  when we take  $\Omega \propto N$ . Eq. (129) then gives us that the Toffoli count scales as  $\tilde{\mathcal{O}}(N^3 N_a^{3/2} / \epsilon)$ . This completes our justification of the analytic scalings given in Table IV.

#### 4. Considerations on system dependent constants

One of the biggest disadvantages about using high-order finite difference formulas is that their performance depends on the higher-order derivatives of the energy. Specifically, Eq. (109) gives the requirements that these constants much satisfy for each of the derivatives. While a simple expression for  $c$  and  $\epsilon$  cannot be easily extracted from such an expression, lower bounds can be extracted from the requirements. Using Eq. (109) we have from substituting  $k = 1$  into the expression that if  $|\psi_0\rangle$  is the true ground state (after choosing the global phases of the eigenstates such that their derivatives are orthogonal to themselves) we then have that

$$\langle \psi_0 | \frac{dH}{dR_i} | \psi_0 \rangle \leq \epsilon c. \quad (136)$$

From this expression we see that while the product of the two constants is lower bounded by the expected derivative in the ground state, that does not determine either of the two constants that constitute it. However, we can see that dimensionally,  $\epsilon$  can be thought of as an energy scale and  $c$  can be thought of an inverse length scale for the system.

A reasonable approximation may be made for the constant  $c$  in the case of first-quantized plane waves. If we assume that the eigenvalue gap is large and the high-order derivatives of the Hamiltonian dominate the expectation value of the energy then it is straightforward to see from perturbative arguments that

$$\frac{d^k E}{dR_i^k} \sim \langle \psi_0 | \frac{d^k H}{dR_i^k} | \psi_0 \rangle \leq \epsilon c^k k^{k/2}. \quad (137)$$

Thus we have in this restricted case that

$$c \geq \frac{1}{\sqrt{k}} \left( \frac{\langle \psi_0 | \frac{d^k H}{dR_i^k} | \psi_0 \rangle}{\epsilon} \right)^{1/k} \quad \forall k. \quad (138)$$

If we choose  $\epsilon = \langle \psi_0 | H | \psi_0 \rangle$  to be the ground-state energy then it follows that any such scaling that satisfies this must obey

$$c \gtrsim \sup_k \left( \frac{1}{\sqrt{k}} \left( \frac{\langle \psi_0 | \frac{d^k H}{dR_i^k} | \psi_0 \rangle}{\langle \psi_0 | H | \psi_0 \rangle} \right)^{1/k} \right). \quad (139)$$

While the derivative with respect to the coordinate of the Hamiltonian in general is difficult to compute, we can argue about it in plane wave bases, which have an analytical expression as illustrated by Eq. (34). Specifically, we have that if  $R_i = R_{i,\alpha}$  is one of the position coordinates of a nuclei with charge  $Z_i$  we have (using the same notation as Sec. IV C) that

$$\frac{d^k H}{dR_{i,\alpha}^k} = -\frac{4\pi Z_i}{\Omega} \sum_{\mathbf{p}} \sum_{\mathbf{s} \neq 0} \frac{(i\mathbf{k}_s^\alpha [i])^k e^{i\mathbf{k}_s \cdot (\mathbf{R}_i + \mathbf{r}_p)}}{|\mathbf{k}_s|^2} c_{\mathbf{p}}^\dagger c_{\mathbf{p}}. \quad (140)$$

Terms in this result scale like  $\mathcal{O}(N^{k/3})$ , but as the phase  $e^{i\mathbf{k}_s \cdot (\mathbf{R}_i + \mathbf{r}_p)}$  oscillates rapidly as we change  $\mathbf{p}$  and  $\mathbf{s}$ , the final sum over both variables can be as small as  $o(1)$ . This suggests that physical circumstances might exist where  $c \in o(1)$ , but it could in other cases scale as  $c \in \mathcal{O}(N^{1/3})$  or potentially higher. Numerical explorations will likely be needed to better understand the scaling of  $\epsilon$  and  $c$ .

#### B. Expectation value estimation of the force operator at the Heisenberg limit with overlap estimation

On a fault-tolerant quantum device one can estimate expectation values at the Heisenberg limit scaling as  $\epsilon^{-1}$  with the unbiased error  $\epsilon$  (which is tight [106,109]). This allows us to estimate gradients via the Hellman-Feynman theorem, similarly to how one would achieve this in a NISQ setting (Sec. III). The Heisenberg limit was originally achieved by the overlap estimation algorithm (OEA) of [55]. In our case the expectation values that we wish to measure are real valued, enabling a slight optimization over the original algorithm. In this section we perform an asymptotic cost analysis of our implementation of the OEA on a block-encoded derivative operator, including optimizing the preparation and reflection



subroutines for the purposes of derivative estimation, and compare it to previous methods.

The OEA aims to perform unbiased estimation of  $\langle \psi | U | \psi \rangle$  for an arbitrary unitary operator  $U$  and state  $|\psi\rangle$ . It does so by calling a subroutine, the amplitude estimation algorithm (AEA), which estimates  $|\langle \psi | U | \psi \rangle|$  via phase estimation of the Szegedy walk unitary

$$S = \mathcal{R}U\mathcal{R}U^\dagger, \quad \mathcal{R} = I - 2|\psi\rangle\langle\psi|. \quad (141)$$

The eigenvalues of this walk operator, constrained to the subspace spanned by  $|\psi\rangle$  and  $U|\psi\rangle$  provide us with all the information we need to estimate the amplitude. To see this, let

$$P_\psi = |\psi\rangle\langle\psi| + (1 - |\psi\rangle\langle\psi|)U|\psi\rangle\langle\psi|U^\dagger(1 - |\psi\rangle\langle\psi|). \quad (142)$$

We have from Jordan's Lemma that the action of  $S$  can be broken into a direct sum of irreducible two-dimensional subspaces. Therefore  $P_\psi S P_\psi = S P_\psi$ . It then follows that for any  $|\phi\rangle$  such that  $P_\psi|\phi\rangle = |\phi\rangle$  that phase estimation will return an eigenvalue of the form  $e^{\pm i\theta}$  for some angle  $\theta \in [-\pi, \pi]$ . Further analysis in [110] shows that this eigenphase satisfies  $\cos(\theta) = 2|\langle U \rangle|^2 - 1$ . This implies that if  $\hat{\theta}$  is our estimate of theta, then

$$\begin{aligned} |\widehat{\langle U \rangle}| &= \sqrt{\frac{\cos(\hat{\theta}) + 1}{2}} = |\cos(\hat{\theta}/2)| = |\cos((\theta + \Delta)/2)| \\ &= |\cos(\theta/2)\cos(\Delta/2) - \sin(\theta/2)\sin(\Delta/2)|, \end{aligned} \quad (143)$$

where  $\Delta = \hat{\theta} - \theta$  and  $\mathbb{V}(\hat{\theta}) = \mathbb{E}(\Delta^2)$  using the phase estimation method reviewed in Appendix C 2, which is unbiased. Under the assumption that  $\sqrt{\mathbb{V}(\hat{\theta})}/\theta \ll 1$ , the fourth moment of the distribution of  $\hat{\theta}$  is negligible and that the distribution of  $\hat{\theta}$  has negligible support over the branch cut

$$\begin{aligned} \mathbb{V}(|\widehat{\langle U \rangle}|) &= \cos^2(\theta/2)\mathbb{V}(\cos(\Delta/2)) + \sin^2(\theta/2)\mathbb{V}(\sin(\Delta/2)) \\ &\lesssim \mathbb{V}(\sin(\Delta/2)) \\ &\approx \mathbb{V}(\hat{\theta})/4. \end{aligned} \quad (144)$$

Under the above assumptions, an estimator  $\hat{\theta}$  of the eigenphases of  $S$  in the support of the input state  $|\psi\rangle$  that is approximately unbiased and has variance of  $\epsilon_\theta^2$  can be achieved using the techniques of Appendix C 2 in approximately  $\frac{\pi}{2\epsilon_\theta}$  calls to  $\mathcal{S}$ . This variance in  $\hat{\theta}$  then propagates to an estimate of  $|\langle U \rangle|$  with a standard deviation that is approximately bounded by  $\frac{\epsilon_\theta}{2}$  from Eq. (144).

To learn the sign of  $\langle U \rangle$ , we use the fact that

$$\langle + | \langle \psi | c - U | \psi \rangle | + \rangle = \frac{1}{2}(1 + (\langle \psi | U | \psi \rangle)). \quad (145)$$

We can then estimate the expectation value using the fact that, in our context  $\langle \psi | U | \psi \rangle$  is real valued. (Note that this would not be the case if one encoded  $U = e^{-i\mathbf{d}H/\mathbf{d}R_i}$  as was originally proposed in [55].) That  $\langle \psi | U | \psi \rangle$  is real implies  $|\langle + | \langle \psi | c - U | \psi \rangle | + \rangle| = \langle + | \langle \psi | c - U | \psi \rangle | + \rangle$ , and thus

$$\langle \psi | U | \psi \rangle = \frac{1}{2}(4((1 + \langle \psi | U | \psi \rangle)/2)^2 - (\langle \psi | U | \psi \rangle)^2 - 1), \quad (146)$$

and we can find an estimator for the overlap as

$$\widehat{\langle U \rangle} = \frac{1}{2} \left( 4 \left( \frac{1 + \widehat{\langle U \rangle}}{2} \right)^2 - |\widehat{\langle U \rangle}|^2 - 1 \right). \quad (147)$$

Note that one could in principle estimate  $\langle \psi | U | \psi \rangle$  directly from amplitude estimation of  $c - U$ .

The variance of the estimate of  $\widehat{\langle U \rangle}$  is then found from the additive property of variance of independent variables,

$$\mathbb{V}(\widehat{\langle U \rangle}) = 4\mathbb{V} \left( \left( \frac{1 + \widehat{\langle U \rangle}}{2} \right)^2 \right) + \frac{1}{4}\mathbb{V}(|\widehat{\langle U \rangle}|^2). \quad (148)$$

We have from Jensen's inequality that for any convex function  $\Phi$  and random variable  $X$  with variance  $\sigma^2$ ,

$$\mathbb{E}(\Phi(X)) - \Phi(\mathbb{E}(X)) \leq \frac{\sigma^2 \sup_X \Phi''(X)}{2}. \quad (149)$$

Since  $\Phi(X) = X^2$  is a convex function with second derivative 1 for all  $X \in [-1, 1]$  we have from Eq. (149) that

$$\mathbb{V}(\widehat{\langle U \rangle}) \leq 4\mathbb{V} \left( \left( \frac{1 + \widehat{\langle U \rangle}}{2} \right) \right) + \frac{1}{4}\mathbb{V}(|\widehat{\langle U \rangle}|). \quad (150)$$

Specifically, we will choose  $\mathbb{V}(\widehat{\langle U \rangle}) \leq \epsilon_{\text{OEA}}^2$ . This motivates the following choices of the variance targets for the two operations

$$\mathbb{V} \left( \left( \frac{1 + \widehat{\langle U \rangle}}{2} \right) \right) \leq \frac{\epsilon_{\text{OEA}}^2}{8}, \quad \mathbb{V}(|\widehat{\langle U \rangle}|) \leq 2\epsilon_{\text{OEA}}^2. \quad (151)$$

We now estimate the complexity of computing the requisite estimators within the variance requirements of Eq. (151). Our circuit primitives and phase estimation routine for the amplitude estimation algorithm (including control) are given in Fig. 6. We note that control can be added to  $\mathcal{S}$  by either controlling the implementations of  $U$  and  $U^\dagger$  or by controlling the implementation of  $R$ . We assume that the cost to implement controlled- $\mathcal{S}$  compared to the cost to implement  $\mathcal{S}$  without control is negligible in the number of Toffoli gates required (which is typically the case). We make one further improvement to our phase estimation routine in the same vein as [67]: instead of performing phase estimation by controlling  $\mathcal{S}^{2^n}$  by the  $n$ th qubit in the QPE register, we implement the unitary

$$|0\rangle\langle 0|S^{-2^{n-1}} + |1\rangle\langle 1|S^{2^{n-1}}. \quad (152)$$

This may be achieved by realizing that

$$\begin{aligned} \mathcal{R}S^k\mathcal{R} &= \mathcal{R}(U^\dagger\mathcal{R}UR)^k\mathcal{R} = (\mathcal{R}U^\dagger\mathcal{R}U)^k \\ &= (U^\dagger\mathcal{R}UR)^{\dagger k} = S^{-k}, \end{aligned} \quad (153)$$

and so one need only append  $S^{2^n}$  by an application of  $\mathcal{R}$  on either side, controlled by the  $n$ th qubit in the QPE register being in the  $|0\rangle$  state to implement Eq. (152). This technique halves the number of applications of  $\mathcal{S}$  required, and significantly reduces the control overhead. With this implemented, using the variance targets in Eq. (151) for our two separate rounds of amplitude estimation will achieve an estimate of  $\langle \psi | U | \psi \rangle$  with standard deviation approximately

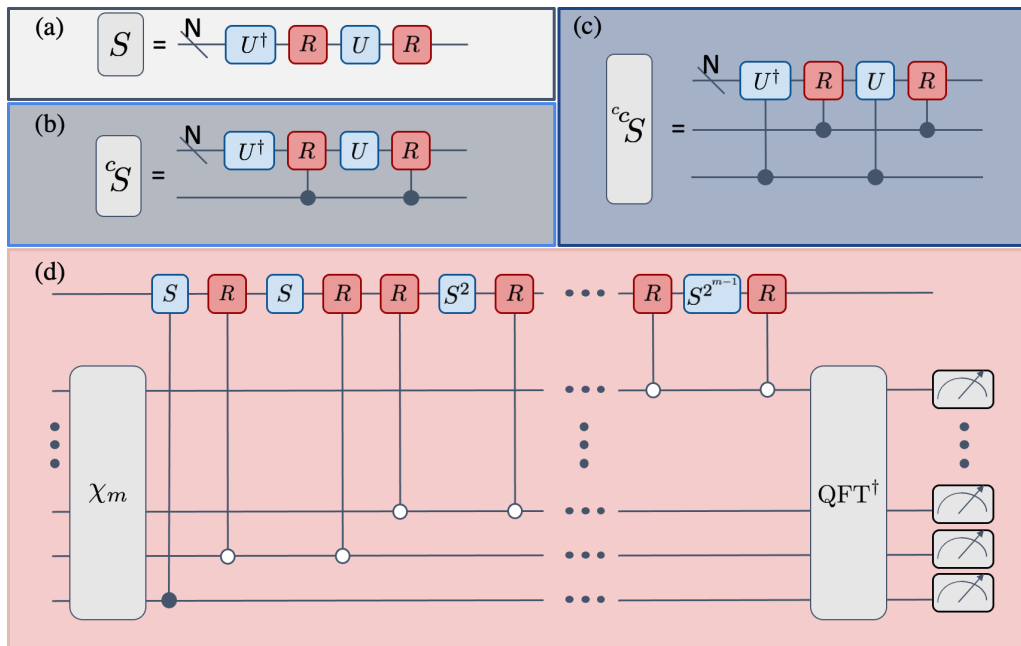


FIG. 6. FT circuits for the overlap estimation algorithm. (a) A circuit implementation of the quantum walk unitary  $S$  without control. (b) A circuit implementation of the quantum walk unitary  $S$  controlled by a single qubit. (c) A circuit implementation of the quantum walk unitary  $S$  controlled by two qubits, required to estimate the sign of  $\langle U \rangle$ . (d) Algorithm for the quantum phase estimation routine, sped up by a factor 2 by improvements discussed in the text. Note that for the purposes of estimating the sign of  $\langle U \rangle$ ,  $S$  should be replaced by  $c - S$  in the QPE circuit (d) and an additional ancilla qubit added. The unitary  $\chi_m$  denotes a preparation of a QPE register state to estimate at the Heisenberg limit [67]. Black dots denote a controlled operation implemented when the control qubit is in the  $|1\rangle$  state, while white dots denote a controlled operation implemented when the control qubit is in the  $|0\rangle$  state.

bounded by  $\epsilon_{\text{OEA}}$ , using

$$N_{\text{calls}} = \frac{\sqrt{2}\pi}{\epsilon_{\text{OEA}}} + \frac{\pi}{2\sqrt{2}\epsilon_{\text{OEA}}} = \frac{5\pi}{2\sqrt{2}\epsilon_{\text{OEA}}} \quad (154)$$

calls to a (controlled) circuit implementation of  $S$ .

To convert the cost of the OEA to the cost of gradient estimation, we need to account for the fact that  $\frac{dH}{dR_i}$  is not a unitary operator. This is an issue, as the OEA explicitly requires unitary  $U$ . This issue may be resolved by block encoding  $\frac{dH}{dR_i}$  (Sec. IV A): adding additional qubits and finding a unitary operator  $U_i$  on the larger Hilbert space such that

$$\begin{aligned} & \left\| \left( \langle 0| \otimes \mathbb{1} \right) U_i \left( |0\rangle \otimes \mathbb{1} \right) - \left( \langle 0| \otimes \mathbb{1} \right) \begin{pmatrix} \frac{1}{\lambda_{F_i}} \frac{dH}{dR_i} & U_{01} \\ U_{10} & U_{11} \end{pmatrix} \left( |0\rangle \otimes \mathbb{1} \right) \right\|_\infty \\ & \leq \epsilon_{\text{BE}} \rightarrow \left| \langle \psi | \langle 0| U_i |0\rangle | \psi \rangle - \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right| \\ & \leq \epsilon_{\text{BE}}, \end{aligned} \quad (155)$$

where  $\epsilon_{\text{BE}}$  is the block-encoding error. This block encoding of  $\frac{dH}{dR_i}$  requires rescaling by some constant  $\lambda_{F_i}$  for  $U_i$  to be unitary. We may propagate this rescaling to the error in estimating  $\frac{dE}{dR_i}$ ; an error  $\epsilon_{\text{OEA}}$  in the estimation of  $\langle \psi | U_i | \psi \rangle$ . However, one complication involved in doing so arises from the fact that many of the estimation errors given above are stated in terms of the variance of the estimate. In contrast, errors such as the block-encoding errors are deterministic errors. To solve this, we can use the Chebyshev inequality to say that for an

estimate  $\mathbf{y}$  with variance  $\sigma^2$

$$\Pr(|\mathbf{y} - \mathbb{E}(\mathbf{y})| \geq k\sigma) \leq \frac{1}{k^2}. \quad (156)$$

Thus, up to a constant factor, we can with probability of failure in  $\mathcal{O}(1)$ , replace the uncertainty in our estimates with their standard deviation. From the above discussion, we then yields an estimate of  $\frac{dE}{dR_i}$ , denoted  $\widehat{\frac{dE}{dR_i}}$  such that with constant probability of failure  $\leq 1/4$  (i.e.,  $k = 2$  above),

$$\left| \widehat{\frac{dE}{dR_i}} - \frac{dE}{dR_i} \right| \leq \lambda_{F_i} (\epsilon_{\text{OEA},i} + \epsilon_{\text{BE}} + \epsilon_{\mathcal{R}}) \quad (157)$$

in the estimation of  $\frac{dE}{dR_i}$ . Here,  $\epsilon_{\mathcal{R}}$  is the error in implementing the reflection operator  $\mathcal{R} = I - 2|\psi\rangle\langle\psi|$ , and we index the errors in the overlap estimation and block encoding by  $i$  (as these will be different for different gradient operators). To bound the error in  $\frac{dE}{dR_i}$  by some  $\epsilon_i$ , we can require the error in both the block encoding and overlap estimation be bounded as  $\epsilon_{\text{BE}}, \epsilon_{\text{OEA}}, \epsilon_{\mathcal{R}} \leq \frac{\epsilon_i}{3\lambda_{F_i}}$ . The bound on  $\epsilon_{\text{OEA}}$  then gives us the cost of executing the gradient estimation algorithm

$$T_{\text{grad-OEA},i} = \frac{15\pi\lambda_{F_i}}{\sqrt{2}\epsilon_i} (T_{F,i} + T_R) + T_{P,i} + \mathcal{O}(\log[\lambda_{F_i}/\epsilon_i]), \quad (158)$$

where  $T_R$  is the cost of implementing the reflection about  $|\psi\rangle$ , and  $T_{P,i}$  is the cost of preparing  $|\psi\rangle$  for the  $i$ th gradient estimation. (We will reuse our system register between gradient estimation applications to significantly reduce the cost of  $T_{P,i}$  for  $i > 1$ .) Note that from the Chernoff bound, the probability

of failure of estimating the component of the gradient within the required error tolerance can be reduced from  $1/4$  to  $\delta$  at cost that is in  $\mathcal{O}(\log(1/\delta))$ . We allow here for the fact that when performing multiple different gradient estimations we may have different block encodings (see Sec. IV A) and state preparation methods (see Sec. VB 2). However, we assume that the reflection oracle (see Sec. VB 2) is implemented the same for each estimation.

### 1. Parallel estimation and importance sampling

The cost estimation in Eq. (158) is the cost of estimating a single gradient element. To expand this cost to a vector of gradients  $\frac{dE}{dR_i}$ , we sum individual terms

$$T_{\text{grad-OEA}} = \sum_i T_{\text{grad-OEA},i}. \quad (159)$$

Following the methods developed in Appendix B, we may select  $\epsilon_i$  to optimize the 1- or 2-norm of the gradient vector  $\epsilon$  by importance sampling, where the resources are allocated

optimally, spending more time on the components that require a higher accuracy. The final term in Eq. (158) is negligible and the second term is not affected by our choice of  $\epsilon_i$  (we will investigate the cost of state preparation in the following section). Thus, we consider the problem of minimizing the sum of the first terms,

$$\sum_i \frac{15\pi\lambda_{F_i}}{\sqrt{2}\epsilon_i} (T_{F,i} + T_R), \quad (160)$$

with either a fixed 1-norm  $\sum_i \epsilon_i$  or 2-norm  $\sum_i \epsilon_i^2$ . This can be solved by the same Lagrangian methods as in previous sections. Solving for a fixed 1-norm yields the condition

$$\epsilon_i = \frac{\epsilon_{\text{grad-OEA}} [\lambda_{F_i} (T_{F,i} + T_R)]^{\frac{1}{2}}}{\sum_{i'} [\lambda_{i'} (T_{F,i'} + T_R)]^{\frac{1}{2}}}. \quad (161)$$

The Cauchy-Schwarz inequality and the Chernoff bound can then be used to bound the total cost of the gradient evaluation within error  $\epsilon_{\text{grad-OEA}}$  with probability of failure  $1/4$  using a total cost of

$$\begin{aligned} T_{\text{grad-OEA}} &= \frac{90\pi \log(12N_a)}{\sqrt{2}\epsilon_{\text{grad-OEA}}} \left[ \sum_i (T_{F,i} + T_R)^{\frac{1}{2}} \lambda_{F_i}^{\frac{1}{2}} \right]^2 + \sum_i T_{P,i} + \sum_i \mathcal{O}(\log [\lambda_{F_i}^{\frac{1}{2}} (T_F + T_R)^{\frac{1}{2}} \epsilon_{\text{grad-OEA}}^{-1}]) \\ &\in \tilde{\mathcal{O}} \left( \frac{1}{\epsilon_{\text{grad-OEA}}} \left( \sum_i \lambda_{F_i} \right) \left( \sum_i T_{F,i} + T_R \right) + \sum_i T_{P,i} \right). \end{aligned} \quad (162)$$

If we promise that  $\lambda_{F_i} \leq \lambda_F$  and  $T_{F,i} \leq T_F$ , this becomes

$$T_{\text{grad-OEA}} \in \tilde{\mathcal{O}} \left( \epsilon_{\text{grad-OEA}}^{-1} (T_F + T_R) \lambda_F N_a^2 + \sum_i T_{P,i} \right). \quad (163)$$

We can repeat the same arguments for the case where error  $\epsilon_{\text{grad-OEA}}$  is desired in the 2-norm,

$$\epsilon_i = \frac{\epsilon_{\text{grad-OEA}} [\lambda_{F_i} (T_{F,i} + T_R)]^{\frac{1}{3}}}{[\sum_{i'} \lambda_{i'}^{2/3} (T_{F,i'} + T_R)^{2/3}]^{\frac{1}{3}}}, \quad (164)$$

for a total cost

$$\begin{aligned} T_{\text{grad-OEA}} &= \frac{90\pi \log(12N_a)}{\sqrt{2}\epsilon_{\text{grad-OEA}}} \left[ \sum_i (T_{F,i} + T_R)^{\frac{2}{3}} \lambda_{F_i}^{\frac{2}{3}} \right]^{\frac{3}{2}} + \sum_i T_{P,i} + \sum_i \mathcal{O}(\log [\lambda_{F_i}^{\frac{2}{3}} (T_F + T_R)^{\frac{2}{3}} \epsilon_{\text{grad-OEA}}^{-1}]) \\ &\in \mathcal{O} \left( \frac{1}{\epsilon_{\text{grad-OEA}}} \left[ \sum_i (T_{F,i} + T_R)^{\frac{4}{3}} \right]^{\frac{3}{4}} \left[ \sum_i \lambda_{F_i}^{\frac{4}{3}} \right]^{\frac{3}{4}} + \sum_i T_{P,i} + \sum_i \mathcal{O}(\log [\lambda_{F_i}^{\frac{2}{3}} (T_F + T_R)^{\frac{2}{3}} \epsilon_{\text{grad-OEA}}^{-1}]) \right), \end{aligned} \quad (165)$$

and in the case where all forces are equal cost and  $\lambda_{F_i} \leq \lambda_F$ ,

$$T_{\text{grad-OEA}} \in \tilde{\mathcal{O}} \left( \epsilon_{\text{grad-OEA}}^{-1} (T_F + T_R) \lambda_F N_a^{\frac{3}{2}} + \sum_i T_{P,i} \right). \quad (166)$$

The cost of implementing the overlap estimation algorithm now depends on the circuit costs  $T_F$ ,  $T_R$ , and  $T_{P,i}$ , and on

the rescaling factor  $\lambda_F$ . Asymptotic costs for  $T_F$  and  $\lambda_F$  were calculated in Sec. IV A. It remains to calculate  $T_R$  and  $T_{P,i}$  in order to obtain a final costing of the algorithm.

### 2. Implementation of preparation and reflection unitaries

As part of the expectation value estimation algorithm, we need to implement a reflection  $I - 2|\psi\rangle\langle\psi|$  about the ground

state  $|\psi\rangle$ . In lieu of other access to  $|\psi\rangle$ , this reflection can be constructed by an inequality test on its energy  $E$  [61]. If we have some  $\mu$  such that  $E < \mu < E + \gamma$ , where  $\gamma$  is the gap between the ground and first excited state with  $E$  denoting the ground-state energy, and we assume the ground state is nondegenerate, we have that we can define an operator of the form

$$\text{sign}[H - \mu] := \sum_j \text{sign}(E_j - \mu)|E_j\rangle\langle E_j| = I - 2|\psi\rangle\langle\psi|. \quad (167)$$

One can approximate the sign function above by approximating it by as a sum of exponentials,  $\sum_i c_i \exp(iHt)$ , and block encoding this using LCU techniques [107], or by making a polynomial approximation and applying the quantum singular value transformation [61]. In this paper we consider the latter. Quantum signal processing requires a block encoding of  $H - \mu$ ,

$$U_H = \begin{pmatrix} \frac{1}{\lambda_H}(H - \mu) & \cdot \\ \cdot & \cdot \end{pmatrix}, \quad (168)$$

where  $\lambda_H$  is an appropriate rescaling factor; this is at most  $\mu$  larger than a corresponding block-encoding of  $H$ . The technique also requires an approximation to the sign function that is within  $\epsilon$  on the eigenvalues of  $\frac{1}{\lambda_H}(H - \mu)$ . Near  $\mu$  this must break down for a finite polynomial series, so it is crucial that  $\mu$  be chosen near the middle of the gap ( $E, E + \gamma$ ). Known polynomial approximations to the sign function that achieve these requirements exist, with polynomial degree  $d = \mathcal{O}(\frac{\lambda_H}{\gamma} \log(\frac{1}{\epsilon}))$  [95]. Let us specify the smallest such constant by defining  $c_{\text{sgn}}$  such that

$$d \leq c_{\text{sgn}} \lambda_H \gamma^{-1} \log(\epsilon^{-1}). \quad (169)$$

Given this sign function approximation and an exact block encoding of  $H - \mu$ , we can construct a block encoding of  $\text{sign}[H - \mu]$  via quantum phase estimation with exactly  $d$  repetitions of the block encoding (or its inverse),  $d$  single-qubit  $z$  rotations and  $2d$   $N$ -qubit Toffoli gates [61,111]. For this to have an error less than  $\epsilon_r$ , we require the block encoding of  $H$  to be implemented with error less than  $\epsilon_r/d$ . Let the number of Toffoli gates needed to implement the block encoding of the  $H$  to an error  $\epsilon_{\text{BE}} \geq \frac{\epsilon_r}{c_{\text{sgn}} \gamma^{-1} \log(\epsilon^{-1}) \lambda_H}$  be  $T_H$  (for many block encodings  $T_H$  will be logarithmically dependent on  $\epsilon_{\text{BE}}$ , and so the rescaling here will not affect any  $\tilde{O}$  scaling results). Further, assume that we are provided a phase gradient state of the form  $\frac{1}{\sqrt{\lceil \frac{2\pi}{\delta_z} \rceil}} \sum_{j=0}^{\lceil \frac{2\pi}{\delta_z} \rceil - 1} e^{2\pi i k / \lceil \frac{2\pi}{\delta_z} \rceil} |k\rangle$  where  $\delta_z$  is the required precision for the single-qubit  $z$  rotations. The number of Toffoli gates needed to perform these using controlled swaps on the above phase gradient state is

$$T_R \in \tilde{\mathcal{O}}((2N + \log(\delta_z^{-1}) + T_H) \lambda_H c_{\text{sgn}} \gamma^{-1} \log(\epsilon^{-1})), \quad (170)$$

Here we resort to  $\tilde{\mathcal{O}}(\cdot)$  notation to simplify the result given the doubly-logarithmic terms that arise from using the lower bound on  $\epsilon_{\text{BE}}$ .

We construct the reflection operator  $1 - 2|\psi\rangle\langle\psi|$  to prepare the ground state  $|\psi\rangle$  from an initial state  $|\phi\rangle$  through the use of fixed-point amplitude amplification [111,112]. This requires we have access to a preparation unitary  $U_\phi$  that prepares

$|\phi\rangle$ , and a bound  $0 < a_0 \leq |\langle\phi|\psi\rangle|$ . The fixed point amplitude amplification algorithm then prepares an approximation to  $|\psi\rangle$  with error  $\epsilon_A$  with  $\frac{\log(2/\epsilon_A)}{2a_0}$  calls to  $U_\phi$  and  $1 - 2|\phi\rangle\langle\phi|$ . A call to the reflection  $1 - 2|\phi\rangle\langle\phi|$  can be implemented by

$$1 - 2|\phi\rangle\langle\phi| = U_\phi(1 - 2|0\rangle\langle 0|)U_\phi^\dagger, \quad (171)$$

which requires two queries to  $U_\phi$  and a single application of a reflection about  $|0\rangle$ , which can be implemented using  $2N$  Toffoli gates. Therefore, we have

$$T_{P,i=1} \in \tilde{\mathcal{O}}\left(\frac{\log(1/\epsilon_A)}{a_0}(T_R + T_\phi + N)\right), \quad (172)$$

where  $T_\phi$  is the cost of implementing  $U_\phi$ .

This does not require an accurate estimate of  $|\langle\phi|\psi\rangle|$ , but just that  $a_0 \leq |\langle\phi|\psi\rangle|$ . In many cases a bound  $a_0$  will be known that is relatively tight, in which case the scaling of this algorithm is optimal [112]. When the bound is loose, it may be improved by performing initial rounds of amplitude amplification followed by direct estimation of  $a_0$  on the device with little overhead [113].

As we are repeating the overlap estimation algorithm multiple times on the same initial state, we can recycle the system register between each iteration, reducing the dependence of our asymptotic cost on the initial state overlap. Let us define the action of our block encoding  $U_i$  of  $\frac{dH}{dR_i}$  on our ground state  $|\psi\rangle$  as

$$\begin{aligned} U_i|\psi\rangle &= \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} |\psi\rangle + \sqrt{1 - \left[\frac{1}{\lambda_{F_i}} \frac{dE}{dR_i}\right]^2} |\chi_i\rangle, \rightarrow |\chi_i\rangle \\ &= \frac{U_i - \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i}}{\sqrt{1 - \left[\frac{1}{\lambda_{F_i}} \frac{dE}{dR_i}\right]^2}} |\psi\rangle \end{aligned} \quad (173)$$

and then we have the well-known result that the Szegedy walk operator  $S_i$  in Eq. (141) is block diagonal on the two-dimensional subspace spanned by  $|\psi\rangle$  and  $|\chi_i\rangle$ , where it takes the form

$$\begin{aligned} S_i &= \exp[iY \sin^{-1}(\theta_i)], \\ \theta_i &= \frac{2}{\lambda_{F_i}} \frac{dE}{dR_i} \sqrt{1 - \left[\frac{1}{\lambda_{F_i}} \frac{dE}{dR_i}\right]^2}. \end{aligned} \quad (174)$$

Regardless of the value of  $\frac{1}{\lambda_{F_i}} \frac{dE}{dR_i}$ , the eigenstates of  $S_i$  are simply  $|\psi_i^\pm\rangle = \frac{1}{\sqrt{2}}(|\psi\rangle \pm i|\chi_i\rangle)$ . This implies that following each implementation of the overlap estimation algorithm, the overlap of the system register with the ground state is always  $\langle\psi_i^\pm|\psi\rangle = \frac{1}{\sqrt{2}}$ . This probability can be raised to 1 by a round of amplitude amplification, but this is not practical here because we do not have easy access to a reflection about  $|\psi_i^\pm\rangle$ . This is especially true when  $\frac{1}{\lambda_{F_i}} \frac{dE}{dR_i}$  is very small (implying the gap between the eigenvalues of  $S_i$  is quite small), which we expect to be the case. Instead, one may measure the reflection operator  $1 - 2|\psi\rangle\langle\psi|$  via a Hadamard test, which entails performing this operator conditional on a control qubit prepared in the  $|+\rangle$  state, and measuring the control qubit in the  $X$  basis. In terms of Toffoli gates, this costs exactly  $T_R$ . With 50% probability, the control qubit is flipped to the  $|-\rangle$  state and the system register is prepared in  $|\psi\rangle$  (up to the error given

by the reflection operator). When the control qubit remains in the  $|+\rangle$  state, we know that our system register is now in the state  $|\chi_i\rangle$ . To avoid throwing away this state, we consider the operation of  $U_i^\dagger$  on  $|\chi_i\rangle$ . If we define

$$\begin{aligned} U_i^\dagger|\psi\rangle &= \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} |\psi\rangle + \sqrt{1 - \left[ \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right]^2} |\chi_i^\dagger\rangle, \rightarrow |\chi_i^\dagger\rangle \\ &= \frac{U_i^\dagger - \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i}}{\sqrt{1 - \left[ \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right]^2}} |\psi\rangle, \end{aligned} \quad (175)$$

we have

$$\begin{aligned} U_i^\dagger|\chi_i\rangle &= \sqrt{1 - \left[ \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right]^2} |\psi\rangle - \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} |\chi_i^\dagger\rangle, \\ U_i|\chi_i^\dagger\rangle &= \sqrt{1 - \left[ \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right]^2} |\psi\rangle - \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} |\chi_i\rangle. \end{aligned} \quad (176)$$

Thus, if our initial measurement of the reflection operator fails, we may iterate performing either  $U_i^\dagger$  or  $U_i$  to the system register, followed by re-measuring the reflection operator  $1 - 2|\psi\rangle\langle\psi|$  till it reports a success. The probability of this failing till the  $n$ th round (for  $j > 1$ ) and then succeeding is  $p_n = \frac{1}{2} (1 - \left[ \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right]^2) \left[ \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right]^{2(n-1)}$ , and so the average Toffoli gate cost of the operation can be calculated to be

$$\begin{aligned} T_{P,i>1} &= \left( \frac{1}{2} + \sum_{n=1}^{\infty} (n+1)p_n \right) T_R \\ &= \left( 1 + \frac{1}{2} \frac{1}{1 - \left[ \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right]^2} \right) T_R. \end{aligned} \quad (177)$$

This cost approaches its minimal value of  $\frac{3}{2}T_R$  as  $\frac{dE}{dR_i} \rightarrow 0$ , and diverges as  $\frac{dE}{dR_i} \rightarrow \lambda_{F_i}$ . This is curious as it is the opposite behavior that one finds when considering the cost of estimating expectation values in NISQ. The divergence can be avoided by artificially increasing  $\lambda_{F_i}$ , though we expect in practice that this will never be required. Assuming  $\frac{1}{\lambda} \frac{dE}{dR_i} \leq 0.5$ , we have  $T_{P,i>1} < 2T_R$ . This yields a total preparation cost of

$$\sum_i T_{P,i} \in \tilde{\mathcal{O}} \left( \left( N_a + \frac{\log(1/\epsilon_A)}{a_0} \right) T_R + \frac{\log(1/\epsilon_A)}{a_0} (T_\phi + N) \right). \quad (178)$$

The dependence on  $a_0^{-1}$  in this equation is additive to the dependence on  $N_a$ . For practical applications, we expect  $T_R \gg T_\phi + N$ , and so the last term can be neglected.

### 3. Total costing of the overlap estimation algorithm

The full cost of the overlap estimation algorithm can be found by substituting Eq. (178) into Eq. (166). This yields an expression in terms of the costs  $T_R$  and  $T_F$  to implement the reflection algorithm and block-encoded force operator respectively, as well as the rescaling factor  $\lambda_F$  for the block encoding of the force operator, and some other physical parameters. In general the first term of Eq. (178) will dominate the second (we assume  $T_R \gg T_\phi + N$ ), so we drop it for simplicity, yield-

ing

$$\begin{aligned} T_{\text{grad-OEA}} &\in \tilde{\mathcal{O}} \left( \epsilon_{\text{grad-OEA}}^{-1} \lambda_F N_a^{3/2} T_F \right. \\ &\quad \left. + \left[ \epsilon_{\text{grad-OEA}}^{-1} \lambda_F N_a^{3/2} + 2N_a + a_0^{-1} \right] T_R \right). \end{aligned} \quad (179)$$

The last two terms in the square brackets come from the cost of state preparation. We see that the middle term is completely dominated by the first term in the square brackets. In practice, we expect the final term to be similarly dominated. Then, in Eq. (170) we expect  $T_H \gg 2N, \log(\delta_z^{-1})$ . Ignoring constants that do not scale with the system size, we can then write

$$T_R \in \tilde{\mathcal{O}}(T_H \lambda_H \gamma^{-1}), \quad (180)$$

which yields a simplified asymptotic estimate

$$T_{\text{grad-OEA}} \in \tilde{\mathcal{O}} \left( \epsilon_{\text{grad-OEA}}^{-1} \lambda_F N_a^{3/2} [T_F + \lambda_H \gamma^{-1} T_H] \right). \quad (181)$$

As our methods for implementing the block-encoded derivative operator are either similar to or better than our methods for implementing the block-encoded Hamiltonian, we assume

$$T_F \in \mathcal{O}(T_H). \quad (182)$$

This implies the second term in Eq. (181) will be the dominant contribution to an asymptotic resource estimate due the additional multiplicative factors. We tabulate the estimated scalings for the overlap estimation algorithm in Table V.

Let us now compare the asymptotic cost of gradient estimation via the overlap estimation algorithm to the result of the finite difference method, Eqs. (128) and (131), including terms assumed to be constant in Tables IV and V. As we do not have tight lower bounds on either method, a direct universal comparison is not possible. In place of this we compare the asymptotic upper bounds. Given the assumption in Eq. (182), all terms in these bounds depend linearly on  $\lambda_H$  and  $T_H$  so it is relevant to compare

$$\tau_{\text{OEA}} = \epsilon^{-1} \gamma^{-1} \lambda_F \quad (183)$$

against the three different potentially dominant terms in the finite difference method Eq. (132),

$$\begin{aligned} \tau_{\text{FD},1} &= \epsilon^{-1} c \log^{5/2}(\epsilon), \quad \tau_{\text{FD},2} = \epsilon^{-1} N_a \gamma^{-1} \max_{\mathbf{R}',i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|, \\ \tau_{\text{FD},3} &= \log(1/\epsilon) \left( \min_i |a_i| \right)^{-1} \gamma^{-1} \log^{5/2}(\epsilon). \end{aligned} \quad (184)$$

Note that all of the above terms have units of inverse energy (as we have dropped the universal dependence on  $\lambda_H$ ). Equation (132) chooses between two different algorithms depending on which is faster, so we should compare  $\tau_{\text{OEA}}$  to the best-scaling algorithm. The appropriate quantity to compare to is

$$\max(\tau_{\text{FD},1}, \min(\tau_{\text{FD},2}, \tau_{\text{FD},3})). \quad (185)$$

We have from Sec. V A 4 that  $c \leq \frac{dE}{dR_i} \leq \max_{\mathbf{R}',i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\|$ . Then, as we assume  $\epsilon \sim \langle \psi_0 | H | \psi_0 \rangle$ , we have  $\epsilon \gg \gamma$ , which implies  $c \ll \max_{\mathbf{R}',i} \left\| \frac{dH(\mathbf{R}')}{dR_i} \right\| \gamma^{-1}$ . This in turn implies  $\tau_{\text{FD},2} \gg \tau_{\text{FD},1}$ . Let us now consider the comparison between  $\tau_{\text{OEA}}$  and these three terms in turn.

The comparison between  $\tau_{\text{OEA}}$  and  $\tau_{\text{FD},1}$  is stark. Here, the relative comparison is between  $\gamma^{-1} \lambda_F$  and  $c \log(\epsilon)$  respectively. We expect  $c \in o(1)$  (Sec. V A 4), and we have only

TABLE V. Scaling of the overlap estimation algorithm for a system with  $N_a$  atoms and  $N$  orbitals and  $\eta$  electrons within the Born-Oppenheimer approximation, assuming that  $c_{\text{sgn}}, \gamma \in \Theta(1)$  and that the second term of Eq. (181) dominates, and using an error  $\epsilon$  as measured by the 2-norm of the gradient vector. Scalings are calculated by multiplying the results of Table IV by the relevant  $\lambda_F$  quantity.

System	Analytic scaling	Empirical scaling
First quantized plane waves	$\tilde{\mathcal{O}}(N^{4/3}N_a^{23/6}(\mathbb{E}( Z_i ))^{4/3}\epsilon^{-1})$	
Second quantized plane waves	$\tilde{\mathcal{O}}(N^3N_a^{5/2}\epsilon^{-1})$	
Hydrogen chains (sparse)		$\tilde{\mathcal{O}}(N_a^{4.11}\epsilon^{-1})$
Hydrogen chains (double factorization)		$\tilde{\mathcal{O}}(N_a^{4.50}\epsilon^{-1})$
Water clusters (sparse)		$\tilde{\mathcal{O}}(N_a^{4.30}\epsilon^{-1})$
Water clusters (double factorization)		$\tilde{\mathcal{O}}(N_a^{4.74}\epsilon^{-1})$

a logarithmic dependence on  $\epsilon$ . By contrast,  $\lambda_F$  can be quite large and scale badly in the system size [see Sec. IV A, in particular Eqs. (88), (93), and (98)]. The gap  $\gamma$  to the first excited state varies in molecular and condensed matter systems, but it can also be quite small. This implies that the best-scaling  $\tau_{\text{OEA}}$  (e.g., for double-factorized force operators in systems with large gaps) scale at best equally with the corresponding  $\tau_{\text{FD},1}$ . This is a somewhat surprising result:  $\tau_{\text{FD},1}$  contains the cost of performing phase estimation, which is typically expected to dominate the finite difference algorithm, while  $\tau_{\text{FD},2}$  and  $\tau_{\text{FD},3}$  detail the cost of state recycling or state re-preparation. The reason behind this is that the overlap estimation algorithm contains a sub-routine (the reflection operation) that uses exactly the same circuitry as phase estimation, and the OEA requires then repeating this subroutine  $\epsilon^{-1}$  times. This multiplicative factor is interesting as for NISQ methods the cost of derivative estimation is completely independent of the Hamiltonian 1-norm.

The comparison between  $\tau_{\text{FD},2}$  and  $\tau_{\text{OEA}}$  is more complex: here the rescaling factor  $\lambda_F$  competes with  $N_a \max_{\mathbf{R}, i} \|\frac{dH(\mathbf{R})}{dR_i}\|$ . For a fixed  $\mathbf{R}$ , the spectral norm  $\|\frac{dH(\mathbf{R})}{dR_i}\|$  is a lower bound for  $\lambda_{F,i}(\mathbf{R})$  (being the rescaling factor for a block encoding of  $\frac{dH}{dR_i}$  at point  $\mathbf{R}$ ), though many block encodings are not this efficient. For example, in hydrogen chains we expect  $\|\frac{dH(\mathbf{R})}{dR_i}\|$  to be roughly independent of the system size  $N_a$ . This is matched by a double-factorized encoding, but a sparse block encoding has  $\lambda_F \sim N_a^2$  (Fig. 4). On the other hand, the finite difference algorithm requires that we take the maximum  $\|\frac{dH(\mathbf{R})}{dR_i}\|$  across the range of points used for finite difference, which significantly grows this cost. Assuming that an efficient block encoding is chosen so that  $\lambda_F \sim \|\frac{dH(\mathbf{R})}{dR_i}\|$ , we have  $\tau_{\text{FD},2} \lesssim N_a \tau_{\text{OEA}}$ . This implies that when the cost of state re-preparation is bad enough that the recycling algorithm becomes preferred, the overlap estimation algorithm will likely become favorable again.

We finally compare  $\tau_{\text{FD},3}$  to  $\tau_{\text{OEA}}$ . Ignoring logarithmic factors, the comparison here is between  $\epsilon^{-1}\lambda_F$  and  $(\min_i |a_i|)^{-1}$ . This is highly dependent on the initial state preparation, of which only few studies have been made (e.g., [114]). In lieu of exact statements that can be made here, we note that the above comparison suggests

$$\tau_{\text{FD},1} < \tau_{\text{OEA}} < \tau_{\text{FD},2}. \quad (186)$$

Then, investigating Eq. (185), whenever  $\tau_{\text{FD},3} < \tau_{\text{FD},1}$  the finite difference algorithm will likely be asymptotically faster,

while whenever  $\tau_{\text{FD},3} > \tau_{\text{FD},1}$  the overlap estimation algorithm may have an advantage. This can be summarized in the following statement: Whenever repeat state preparation becomes the dominant cost for estimating energies at any point of the finite difference method, the overlap estimation algorithm likely becomes a more efficient algorithm. This is because as the overlap estimation algorithm can efficiently recycle states with unit probability, while the finite difference method has to pay a significant cost in its choice of step size  $dR$  to guarantee the same. For applications in chemistry it is typical to assume that state preparation is not the dominant cost of the calculation (due to the relative size of the target error  $\epsilon$  to the gap  $\gamma$ ). This suggests that for typical use cases the finite difference method may indeed be an optimal choice.

### C. Simultaneous estimation of force operators using a gradient-based expectation value estimation

Previously in this work we have considered using a quantum computer as an energy-estimation subroutine to evaluate forces via higher-order finite difference methods. One might wonder whether a speedup can be gained from performing the difference estimation on the quantum device itself. This was originally considered in [41], and later significantly improved by [104], which proposed a quantum algorithm for evaluating the gradient of a function encoded as an expectation value. However, in order to apply this algorithm to estimate the gradient, the coherent evaluation of the function at a superposition of points is required. The cost of applying this approach directly to estimate the energy gradients as a function of nuclear coordinates appears prohibitive, especially for second-quantized Hamiltonians employing basis sets that are, themselves, dependent on the positions of the nuclei. Recently however, in [56] some of the authors of this paper proposed a strategy for measuring a collection of expectation values that applies the gradient algorithm of [104] to a simple and easy to implement auxiliary function

$$f(\boldsymbol{\theta}) = -\frac{1}{2}\text{Im} \left[ \langle \psi_f | \prod_{j=1}^M e^{-2i\theta_j O_j} | \psi_f \rangle \right] + \frac{1}{2}, \quad (187)$$

where the  $|\boldsymbol{\theta}\rangle$  encode  $M$  binary numbers representing coordinates in a  $M$ -dimensional box centered on the origin. This function can be encoded in a parameterized quantum circuit by performing a Hadamard test for the imaginary component

of the unitary

$$U_{\theta} = \sum_{\theta} \left[ |\theta\rangle\langle\theta| \otimes \prod_{j=1}^M e^{-2i\theta_j O_j} \right]. \quad (188)$$

Substituting  $O_j := \frac{1}{\lambda_{F_j}} \frac{dH}{dR_j}$  into Eq. (187), and assuming that  $|\psi\rangle$  is the ground state of  $H$  (at fixed  $\mathbf{R}$ ) yields a function  $f(\theta)$  that is different from the energy —  $f(\theta) \neq E(\mathbf{R} + \theta)$  — but that satisfies

$$\left. \frac{df}{d\theta_i} \right|_{\theta=0} = \left. \frac{1}{\lambda_{F_i}} \frac{dE}{dR_i} \right|_{\mathbf{R}}. \quad (189)$$

Crucially, implementing a quantumly-controlled version of this auxiliary function is more straightforward than a quantumly-controlled energy as a function of the nuclear coordinates, essentially because the auxiliary function is only required to reproduce the correct gradient at a single point. Implementing this as an oracle in the gradient-based approach of [104] allows for multiple expectation values to be simultaneously estimated at the Heisenberg limit with some degree of parallelization. In this section we will estimate the asymptotic Toffoli cost of implementing this oracle in the gradient-based algorithm, and compare it to the methods developed in previous sections.

We begin by reviewing the essential features of the gradient-based expectation value estimation algorithm. We do not present a complete review of the quantum algorithm for the gradient presented in [104], nor the extension of this to estimate expectation values given in [56]. Instead, we highlight those details that are necessary to describe the modifications required for our purposes and refer the reader to [104] and [56] for a comprehensive description. The gradient-based expectation value estimation algorithm targets the evaluation of the expectation values of  $M$  operators  $\{O_1, O_2, \dots, O_M\}$  (with  $\|O_j\| \leq 1$  for all  $j$ ) with respect to the state  $|\psi\rangle$  to within a precision  $\epsilon_{\text{grad}}$  in the infinity-norm. It does this by invoking the gradient estimation algorithm of [104] as a subroutine, which requires constructing a quantum oracle of function  $f$ . This is done so in terms of a probability oracle  $U_p$ ,

$$U_p : |x\rangle|0\rangle \rightarrow |x\rangle \otimes (|\sqrt{f(x)}\rangle|\psi_{\text{good}}(x)\rangle|1\rangle + \sqrt{1-f(x)}|\psi_{\text{bad}}(x)\rangle|0\rangle), \quad (190)$$

where  $|x\rangle$  denotes a collection of registers encoding binary approximations to the inputs to  $f$  and  $|\psi_{\text{good}}(x)\rangle$  and  $|\psi_{\text{bad}}(x)\rangle$  are arbitrary normalized states. In our case,  $x = \theta$ , which requires a number of ancilla qubits scaling as  $\tilde{O}(M \log(\epsilon_{\text{grad}}^{-1}))$  for storage.

One of the main results in [56] is the construction of a probability oracle  $U_p$  for the function  $f$  using a single query to the state preparation unitary  $U_{\psi}$ . This probability oracle acts on the input register (the register containing the input state  $|x\rangle$  above), the system register (the Hilbert space containing  $|\psi\rangle$ ), and one additional ancilla qubit. The result of [104] allows us to solve our estimation using  $\tilde{O}(\sqrt{M}/\epsilon_{\text{grad}})$  calls to the resulting phase oracle  $U_p$ . It was further shown in [56] that this may be achieved while upperbounding the maximum time evolution required to implement a call to the probability oracle ( $\|\theta\|_{\infty}$ ) in the following way,  $\|\theta\|_{\infty} \leq t_{\text{max}} \in \tilde{O}(\log(\epsilon_{\text{grad}}^{-1})/\sqrt{M})$ . The failure probability of the algorithm

of [104] is bounded only by  $1/3$ , but this can be reduced to  $\delta$  using the Chernoff bound at a cost that is logarithmic in  $\delta$ .

There are several details to address in order to optimally apply this gradient-based approach to the estimation of forces, and to estimate the asymptotic cost of the final result. Firstly, it is desirable to take advantage of the fact that we can implement the reflection operator  $I - 2|\psi\rangle\langle\psi|$  more efficiently than  $U_{\psi}$  (by a factor  $a_0$ ). We will modify the parallel expectation value algorithm to remove all but one of the calls to the state preparation unitary  $U_{\psi}$ , replacing this component of the algorithm with a comparable number of reflections about the ground state. This allows us to obtain a benefit similar to the one we found for the serial approach to overlap estimation presented in Sec. VB (see specifically Sec. VB 2 for the implementation of the reflection operator). Secondly, we need to account for the cost of implementing the initial state preparation step and the controlled time evolution by the force operators in terms of the number of calls to the appropriate block-encodings. Thirdly, we need to translate the error bounds from [56] to our case, accounting for the normalization of the block-encoded force operators, the various sources of error in the state preparation, reflection, and time evolution subroutines, and the fact that we desire a error in the 2-norm rather than the infinity norm.

### 1. Replacement of the intermediate state preparations with ground-state reflections

We now aim to explain how we can replace all but one of the calls to the state preparation unitary  $U_{\psi}$  and its conjugate by the more affordable reflection about the ground state, thereby avoiding most of the dependence on the overlap of the initial state with the ground state ( $a_0$ ). To do so we require some intermediary details of the gradient-based expectation value algorithm, which we recall from [104] and [56]. The quantum algorithm for the gradient makes use of the probability oracle of Eq. (190) by using the walk operator from amplitude amplification to convert the probability of a good state into a phase,

$$G_U = (2\Pi_1 - I)U_p^{\dagger}(2\Pi_2 - I)U_p, \quad (191)$$

where  $2\Pi_1 - I$  is a reflection about the zero state of the system register and the ancilla qubits of the probability oracle, and  $(2\Pi_2 - I)$  is a reflection about the zero state of the specific ancilla qubit that flags the good and bad states in the probability oracle. The eigenvalues of  $G_U$  are of the form  $\exp(\pm i \sin^{-1}(\sqrt{f(x)}))$ , where  $f(x)$  is the probability of observing a good outcome as given by Eq. (190).

The gradient algorithm of [104] uses a quantum singular value transformation to convert from  $G_U$  to a block encoding of the form  $(|0\rangle \otimes I)\text{QSVT}(G_U)(|0\rangle \otimes I) = C$  wherein  $C$  (within some target error  $\epsilon$ ) has eigenvalues  $e^{-irf(x)}$  within the eigenspace of  $G_U$  that is in the support of  $|x\rangle$ . The block-encoded operation can be viewed as a fractional query to the phase oracle, and the block-encoding procedure can be implemented using either linear combination of unitaries methods or the more ancilla efficient methods of [111] (which use controlled queries to  $G_U$  interspersed with single qubit rotations).

By querying such fractional phase oracles at a superposition of carefully chosen points, the gradient algorithm uses phase kickback and the quantum Fourier transform to obtain a high-order finite difference approximation to the gradient. For our purposes, the crucial detail is that the only interaction that the gradient algorithm has with the system register is through the action of (one of several) walk operators (or their conjugates) that take the same form as the one from Eq. (191) (with some varying number of ancilla qubits or additional single-qubit gates acting on the ancilla). In the rest of this section, we will show how one may shift all calls to the preparation unitary  $U_\psi$  within a circuit-based implementation of  $G_U$  to the edges of the unitary, such that  $G_U$  takes the form

$$G_U = U_\psi^\dagger G'_U U_\psi. \quad (192)$$

The reader may confirm that all other subroutines  $G_A$  used in the gradient estimation algorithm of [104] take the same form, making it possible to write

$$G_A = U_\psi^\dagger G'_A U_\psi, \quad (193)$$

where  $G'_A$  may be implemented without calls to  $U_\psi$ . This implies that we may concatenate any combination of these subroutines and eliminate the product  $U_\psi U_\psi^\dagger$  that forms, i.e.,

$$\prod_A G_A = U_\psi^\dagger \prod_A G'_A U_\psi. \quad (194)$$

This will reduce the total number of preparations required to 1 (the final call to  $U_\psi^\dagger$  may be discarded as we do not use the system register afterwards).

The queries to  $U_\psi$  used in each iteration of  $G_U$  can be similarly removed using a basis transformation (at the price of requiring two queries to transform in and out of the basis of  $|\psi\rangle$ ). Specifically, we consider the walk operator as a product of two components,  $2\Pi_1 - I$  and  $U_p^\dagger(2\Pi_2 - I)U_p$ . We may define the projector

$$\Pi'_1 = U_\psi \Pi_1 U_\psi^\dagger, \quad (195)$$

so that  $2\Pi'_1 - I$  is a reflection operator about the  $|\psi\rangle$  state of the system register together with the  $|0\rangle$  state of the ancilla qubits associated with  $\Pi_1$ . Then, the reader may confirm that the probability oracle  $U_p$  constructed by [56] takes the form

$$U_p = U'_p U_\psi, \quad (196)$$

where  $U'_p$  contains no calls to  $U_\psi$ , but instead queries the force unitary  $U_\theta$ . Now we can rewrite our expression for the reflections by adding the following resolutions of the identity to the two components of  $G_U$ :

$$2\Pi_1 - I = U_\psi^\dagger U_\psi (2\Pi_1 - I) U_\psi^\dagger U_\psi = U_\psi^\dagger (2\Pi'_1 - I) U_\psi, \quad (197)$$

$$\begin{aligned} U_p^\dagger (2\Pi_2 - I) U_p &= U_\psi^\dagger U_\psi U_p^\dagger (2\Pi_2 - I) U_p U_\psi^\dagger U_\psi \\ &= U_\psi^\dagger U_p'^\dagger (2\Pi_2 - I) U_p' U_\psi. \end{aligned} \quad (198)$$

This yields our new form of  $G_U$ ,

$$G_U = U_\psi^\dagger (2\Pi'_1 - I) U_p'^\dagger (2\Pi_2 - I) U_p' U_\psi, \quad (199)$$

and in the above notation

$$G'_U = (2\Pi'_1 - I) U_p'^\dagger (2\Pi_2 - I) U_p', \quad (200)$$

as required. More generally, the reader may see by a careful inspection of [56] that the gradient-based expectation value estimation algorithm only interacts with the system register via the application of the components  $2\Pi_1 - I$  and  $U_p^\dagger(2\Pi_2 - I)U_p$ , interleaving them with other operators that act as the identity on the system register. Therefore, if we consider the whole sequence of steps that comprise the algorithm, we can see that all of the intermediate state preparation unitary calls can be removed by cancellation with their adjacent conjugates. Furthermore, the system register is discarded at the end of the expectation value algorithm, so the final call to  $U_\psi^\dagger$  can also be removed.

Next with these pieces in place we can discuss the cost of the algorithm. We take the cost here, as in previous sections, to be the total number of non-Clifford operations needed for each of the operations carried out. Noting that the cost of the reflection  $(2\Pi_2 - I)$  is negligible (being a simple Pauli-Z gate), we can calculate a total cost for the gradient-based phase estimation algorithm of

$$T_{\text{simultaneous}} = \tilde{O}(T_P + Q(T_R + T_{\text{time}})), \quad (201)$$

where  $Q$  denotes the number of repetitions of  $G_U$  and any related unitary, and  $T_P$ ,  $T_R$ , and  $T_{\text{time}}$  denote the cost of implementing  $U_\psi$ ,  $(2\Pi'_1 - I)$ , and  $U_p'$  respectively; we will estimate these in the next sections.

## 2. Asymptotic costs of reflection, force evolution, and preparation

We now provide asymptotic upper bounds on the scaling of  $T_P$ ,  $T_R$ ,  $T_{\text{time}}$  defined in the previous section. The cost  $T_R$  of reflecting about our ground state  $|\psi\rangle$  depends on the precision to which we need to reflect since a projective measurement on the ground state is not typically available. To calculate the cost of the reflection, let  $\epsilon_R$  be a desired bound on the total contribution to the error from implementing the reflection operators, and let  $R_{\text{prep}}$  denote the number of calls to the reflection operator used by the initial state preparation circuit. We can achieve the desired overall bound on the error by implementing the  $Q/2 - 1 + R_{\text{prep}}$  reflection operators each to within a precision  $\frac{\epsilon_R}{Q/2 - 1 + R_{\text{prep}}}$  in the spectral norm. The Toffoli count for block encoding the Hamiltonian (required by the construction of the reflection operator) and implementing the reflections both scale logarithmically with the desired precision. Assuming that  $Q$  and  $R_{\text{prep}}$  both scale at most polynomially with the system size and any other relevant parameters of the problem (we shall later determine that they satisfy this assumption), we can implement each reflection about  $|\psi\rangle$  for a cost  $T_R$  which scales identically (up to logarithmic factors) to the case of serial amplitude estimation [see Eq. (180)],

$$T_R \in \tilde{O}(T_H \lambda_H \gamma^{-1}), \quad (202)$$

where  $\gamma$  is a lower bound on the eigenvalue gap between the ground state and the rest of the spectrum.

We now address the cost of the initial state preparation  $T_P$ . We can consider the cost for this step given in Eq. (173) of Sec. VB2, where it was addressed in the context of serial expectation value estimation with the overlap estimation algorithm. We can also make the same simplifying assumption later used in Sec. VB3, namely that  $T_R \gg T_\phi + N$ . Here  $T_\phi$  is the Toffoli count for implementing the state preparation



unitary  $U_\phi$ , and will often be negligible in cases where an elementary ansatz such as a Hartree-Fock state is used for state preparation.

The contribution to the overall error from imperfectly implementing the reflection operators is already accounted for in the section above, provided that the initial state preparation step requires a number of calls to the reflection operator  $I - |\psi\rangle\langle\psi|$  that grows at most polynomially with the problem parameters. This conditions hold, as discussed in Sec. VB 2. The only remaining source of error is the error from amplitude estimation, which only appears logarithmically in the cost of the state preparation step. Making these simplifications and substitutions, and neglecting all logarithmic factors, we find that (similarly to the initial preparation in Sec. VB 2) the initial state preparation step requires a number of Toffoli gates scaling as

$$T_P = \tilde{O}(a_0^{-1}T_R). \quad (203)$$

The cost  $T_{\text{time}}$  of implementing  $U_p$  is dominated by the cost of implementing the subroutine  $U_\theta$  [Eq. (188)]. Interestingly, it suffices here to take a naive implementation of  $U_\theta$  to demonstrate a speedup; i.e., we aim to implement the  $e^{2i\theta_j O_j}$  in series for each  $O_j = \frac{1}{\lambda_{F_i}} \frac{dH}{dR_i}$ . As in Sec. IV A, for the  $i$ th force operator, we let  $T_{F,i}$  denote the cost of implementing the circuit for the block encoding and we let  $\lambda_{F_i}$  denote the associated rescaling factor. For simplicity, we actually consider the case where the operators are all block encoded with the same rescaling factor  $\lambda_F$ , where  $\lambda_{F_i} \leq \lambda_F$  for all  $i$ . These block encodings are also block encodings of the operators  $\frac{1}{\lambda_F} \frac{dH}{dR_i}$ , and we have  $\|\frac{1}{\lambda_F} \frac{dH}{dR_i}\| \leq 1$  for all  $i$ . As discussed in [56], for each of the  $Q$  calls to  $U_p$ , we perform time evolution by each of the observables for  $\mathcal{O}(\log(\epsilon_{\text{grad}}^{-1}))$  different lengths of time, with a maximum length of time for any such evolution  $\|\theta\|_\infty \leq t_{\text{max}} = \tilde{O}(\log(\epsilon_{\text{grad}}^{-1})/\sqrt{3N_a})$ . We can use optimal Hamiltonian simulation algorithms to accomplish each of these evolutions to within an accuracy  $\epsilon'$  using  $\mathcal{O}(t_{\text{max}} + \log(1/\epsilon'))$  calls to the block encoding for the appropriate force operator [95,111]. Let  $T_F$  be an upper bound on the cost of implementing any of the block-encoded force operators to within a precision  $\epsilon''$ .

Now we wish to determine the asymptotic scaling of the Toffoli count for performing the time evolution steps while contributing an error of at most  $\epsilon_{\text{time}}$  to the final output of the gradient-based estimation algorithm. The complexity of optimal Hamiltonian simulation depends logarithmically on  $\epsilon'$  and we make the further assumption that the complexity of implementing the block-encoded force operator depends logarithmically on  $\epsilon''$ . The total number of time evolution steps performed, and the total number of calls to the block encoding for each force operator, both scale linearly with  $QN_a$  if we neglect logarithmic factors. Therefore, if we wish to bound the total error that arises from implementing time evolution by the force operators to  $\epsilon_{\text{time}}$  it suffices to set  $\epsilon'$  and  $\epsilon''$  polynomially smaller than  $\epsilon_{\text{time}}$ . When we neglect the logarithmic factors, we then see that we can implement all of the time evolution steps required by a single query to  $U_p$  using a number of Toffoli gates which that as

$$T_{\text{time}} = \tilde{O}(\sqrt{N_a}T_F). \quad (204)$$

### 3. Total cost for gradient-based estimation

Now we are ready to determine the overall cost for the force estimation using the gradient-based approach to measurement. We aim to measure the  $3N_a$  force operators to within a precision  $\epsilon$  in the 2-norm. We can achieve this by setting our desired precision in the infinity norm to  $\epsilon/\sqrt{3N_a}$ . For simplicity, we assume that each of the force operators is block encoded with a rescaling factor  $\lambda_{F_i} \leq \lambda_F$ . Then we satisfy the normalization conditions of the gradient-based estimation algorithm by considering the task of estimating of the  $3N_a$  operators  $O_i = \frac{1}{\lambda_{F_i}} \frac{dH}{dR_i}$  to within a precision  $\frac{\epsilon}{\lambda_{F_i} \sqrt{3N_a}}$ . We need to consider four different sources of error. Let  $\epsilon_{\text{grad}}$  denote the error from the gradient-based estimation algorithm itself. Let  $\epsilon_R$  denote the error from the reflection operator about the ground state that we shall use as a subroutine. Let  $\epsilon_P$  denote the error from the initial ground-state preparations step. Finally, let  $\epsilon_{\text{time}}$  denote the total error contributed by our implementations of the various time evolution operators for each of the forces. In order to achieve our desired error in the 2-norm, we require that each of these components of the overall error be bounded above by  $\frac{\epsilon}{4\lambda_F \sqrt{3N_a}}$ . (The cost of  $\epsilon_R$ ,  $\epsilon_P$ , and  $\epsilon_{\text{time}}$  are already accounted for in our choice of  $T_R$ ,  $T_P$ , and  $T_{\text{time}}$  respectively.) Taking the above bound  $\epsilon_{\text{grad}} \leq \frac{\epsilon}{4\lambda_F \sqrt{3N_a}}$ , and substituting this into the above analysis, we find that we need

$$Q = \tilde{O}(N_a \lambda_F / \epsilon) \quad (205)$$

queries to the probability oracle  $U'_p$ .

The gradient-based estimation algorithm requires a number of Toffoli gates, which is dominated by the cost of the initial state preparation, the  $Q$  queries to  $U'_p$ , and the associated reflection operators. Substituting Eqs. (205), (202), (203), and (204) into Eq. (201), we find that

$$T_{\text{simultaneous}} = \tilde{O}(a_0^{-1}T_H \lambda_H \gamma^{-1} + N_a \lambda_F \epsilon^{-1}(T_H \lambda_H \gamma^{-1} + \sqrt{N_a}T_F)). \quad (206)$$

As in the case of serial expectation value estimation, we expect that the cost of the initial state preparation will be negligible (as it does not depend on  $\epsilon^{-1}$ ) in practice and we can therefore drop the first term. We also assume that  $T_F = \mathcal{O}(T_H)$ , i.e., that the cost of implementing the block-encoded derivative operators scales no worse than the cost of implementing the block-encoded Hamiltonian. Making these simplifications, we have

$$T_{\text{simultaneous}} = \tilde{O}(T_H N_a \lambda_F \epsilon^{-1}(\lambda_H \gamma^{-1} + \sqrt{N_a})). \quad (207)$$

The norm of the Hamiltonian should scale at least linearly with each added nuclei, therefore we expect that  $N_a = \mathcal{O}(\lambda_H)$ . We can use this to perform a final simplification, yielding

$$T_{\text{simultaneous}} = \tilde{O}(T_H \lambda_H \gamma^{-1} N_a \lambda_F \epsilon^{-1}). \quad (208)$$

This improves strictly over the bound for the overlap estimation algorithm [Eq. (181)] by an asymptotic factor of  $N_a^{1/2}$ . This can be understood as a direct square root speedup from the parallelization of the estimation of multiple gradients. It was proven in [56] that, for certain cases, gradient-based estimation achieves the optimal (up to logarithmic factors) scaling in the number of calls to the state preparation unitary  $U_\psi$ . This

TABLE VI. Scaling of the gradient-based expectation value algorithm for a system with  $N_a$  atoms and  $N$  orbitals and  $\eta$  electrons within the Born-Oppenheimer approximation assuming the results of Eq. (208), and using an error  $\epsilon$  as measured by the 2-norm of the gradient vector. Scalings are calculated by multiplying the results of Table IV by the relevant  $\lambda_F N_a^{-1/2}$  quantity. We assume  $\gamma \in \Theta(1)$ .

System	Analytic scaling	Empirical scaling
First quantized plane waves	$\tilde{\mathcal{O}}(N^{4/3} N_a^{10/3} (\mathbb{E}( Z_i ))^{4/3} \epsilon^{-1})$	
Second quantized plane waves	$\tilde{\mathcal{O}}(N^3 N_a^2 \epsilon^{-1})$	
Hydrogen chains (sparse)		$\tilde{\mathcal{O}}(N_a^{3.61} \epsilon^{-1})$
Hydrogen chains (double factorization)		$\tilde{\mathcal{O}}(N_a^{4.00} \epsilon^{-1})$
Water clusters (sparse)		$\tilde{\mathcal{O}}(N_a^{3.80} \epsilon^{-1})$
Water clusters (double factorization)		$\tilde{\mathcal{O}}(N_a^{4.24} \epsilon^{-1})$

suggests that further improvements in our case may be difficult and would necessarily come from exploiting additional structure not present in the general case. The comparison to the finite difference method (Sec. VA) is slightly more complex: Following the comparison made in Sec. VB3, the relevant comparison is between  $\tau_{\text{FD},1} = \epsilon^{-1} c \log^{5/2}(\epsilon)$  and

$$\tau_{\text{GBEV}} := N_a^{-1/2} \epsilon^{-1} \gamma^{-1} \lambda_F = \tau_{\text{OEA}} N_a^{-1/2}. \quad (209)$$

It was argued in Sec. VB3 that there exist systems for which  $\gamma^{-1}$  and  $\lambda_F$  are both constant (or scale logarithmically) in the system size (the latter based on the numerical analysis in Sec. IVA). In this case, the gradient-based expectation value estimation algorithm would be asymptotically faster by a factor  $N_a^{1/2}$ , even assuming the higher-order gradient bound  $c$  is constant. However, for systems with a particularly small gap, or for which the block encoding of the force operator is large, the finite-difference algorithm may be optimal. We give asymptotic scalings of the systems that we studied in this paper in Table VI, following the analysis used to generate Sec. IV.

## VI. CONCLUSIONS

In this paper we performed an in-depth study of the cost of estimating forces using current NISQ and future fault-tolerant devices. We optimized strategies for estimating forces on a NISQ device, and determined that for some state-of-the-art methods, estimating the entire force vector to a fixed error (as measured by the 2-norm of the error vector) may be roughly equal or a lower cost than estimating the energy of the system. This suggests that force vectors would effectively come “for free” in a NISQ electronic structure calculation. As NISQ methods tend to be variational and require calculating the electronic energy repeatedly to minimize it with respect to variational parameters, it does not immediately follow that force estimation is easier than energy estimation for a system. However, for methods that are not variational (e.g., [82]), or methods using ansatzes with precalculated parameters [115], these methods would allow for quicker estimation of forces than the energies themselves.

We further present, optimize, and cost several algorithms in this paper for calculating forces within a fault-tolerant quantum computing framework. Specifically, we propose three methods, one based on finite difference algorithms for computing gradients, and two based on estimating expectation values of force operators (that differ in their means of ex-

pectation value estimation). We find that in situations where the expected cost of state preparation is low, the finite difference algorithms provide asymptotically better scaling. In contrast, the gradient operator approach provides better scaling if the cost of state preparation is high. In the best case scenario, the scaling of these algorithms for first quantized plane waves are a factor of  $N_a^{3/2}$  worse than the cost of the best first quantized algorithms for estimating energies. The cost estimates for block-encoding forces using atomic-centered basis orbitals in second quantization is more complex because it requires a numerical study. However, similar to the above first-quantized examples, we find that block-encoding force operators in an atomic-centered basis does not translate into a relative reduction in cost for estimating forces compared to estimating energies for the electronic structure problem. The reason behind both results is that all methods studied have a multiplicative cost of Hamiltonian simulation, either directly or as part of a reflection around the ground state. This reflection is required to target the ground state of an electronic structure problem, so we do not see an easy method to circumvent this problem and make direct force estimation on a quantum device easier than energy estimation. This stands in contrast to the above point for NISQ algorithms.

The limitation that the time to estimate forces must be greater than the time to estimate energies narrows the list of applications of this paper within fault tolerance (where the main value of quantum computers is expected to be realized). Our methods look promising for applications that do not require so many repeated force or derivative calculations, e.g., geometry optimization and dipole estimation appear more practical following this paper. However, although we estimate that the cost of reaching chemical accuracy for forces (6.4 mHa/Å error per force component) is comparable to that of energies (1.6 mHa), the bottleneck of practical MD simulations is that they typically require on the order of  $10^6 - 10^9$  unique force calculations [79]. Wall-clock time estimates for single point energy calculations on fault-tolerant devices in the beyond-classical regime are typically on the order of multiple hours; repeating this millions of times to perform a MD simulation is not practical. This suggests that to enhance MD simulations with quantum computers, new approaches need to be found.

We suggest some directions that we believe may be fruitful for future research. A clear target for further research in force optimization is to improve the bounds on

the variance that we obtained in the NISQ section (e.g., by implementing locally-biased classical shadow techniques [93]), although we do not believe that this will be a sufficient improvement to achieve beyond-classical molecular dynamics simulations in the NISQ era. In a similar vein, it would be of interest to determine the error tolerances required for other first-order derivative quantities, e.g., those given in Table II. It is also of critical importance to study how generalizable error tolerances are for forces; it is unclear whether our result of 6.4 mHa/Å applies to systems other than water. Alternatively, instead of performing a semi-classical simulation where one extracts force information from a quantum device to perform a molecular dynamical step, it should be possible to perform a fully-quantum simulation—which is performed entirely on the device. This is relevant because one of the major drivers of complexity is the  $\tilde{O}(1/\epsilon)$  cost of computing the gradient. A more practical approach would entail either simulating the differential equation governing the dynamics on the quantum computer itself (while retaining a classical description of the nuclear motion within the Born-Oppenheimer approximation), or by a fully-quantum simulation that treats the nuclei and electrons on a similar footing such as a first quantized simulation. Alternatively, one can consider taking a small number of measurements from a quantum device and fitting a semiempirical model to the results. This could be done using a semiclassical molecular mechanical model, for example machine-learning techniques that take a few high-accuracy and many low-accuracy data points to generate high-accuracy force fields [116]. Equivalently, one could consider reusing points from finite difference calculations, or sparsely sampling gradients to reduce these constant factor overheads overhead.

### ACKNOWLEDGMENTS

The authors thank Jarrod McClean, Jordi Tura, Kianna Wan, and Clemens Utschig-Utschig for helpful discussions. N.W. worked on this project under a research grant from Google Quantum AI and was partially supported by the “Embedding Quantum Computing into Many-body Frameworks for Strongly Correlated Molecular and Materials Systems” project, which is funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (BES), the Division of Chemical Sciences, Geosciences, and Biosciences. D.W.B. worked on this project under a sponsored research agreement with Google Quantum AI. D.W.B. is also supported by Australian Research Council Discovery Projects No. DP190102633 and No. DP210101367.

### APPENDIX A: DEPENDENCE OF THE MOLECULAR ORBITAL COEFFICIENTS ON THE ATOMIC ORBITAL OVERLAP COEFFICIENTS

Because only the MO coefficients  $\mathbf{C}$  carry the dependence on  $\mathbf{S}$ , to derive explicit expressions for these terms, we will need to derive expressions for  $\partial\mathbf{C}/\partial\mathbf{S}$ . Herein, we will assume that the MO coefficients are real-valued. We begin by writing a general parametrization [117] of the molecular

orbitals

$$\mathbf{C}(\mathbf{C}^0, \mathbf{S}, \Theta) = \mathbf{C}^0 \mathbf{M}(\mathbf{S}) \mathbf{U}(\Theta) \quad (\text{A1})$$

where  $\mathbf{C}^0$  are the (unperturbed) reference orbitals—here taken to be from the solution of the Hartree-Fock equations,  $\mathbf{M}$  ensures the orthogonalization of the metric  $\mathbf{S}$ , and the Hartree-Fock energy is minimized by the unitary matrix  $\mathbf{U}$ , which is a function of orbital rotation parameters  $\Theta$ . Because  $\mathbf{U}$  is unitary, it may take the form

$$\mathbf{U}(\Theta) = \exp[\mathbf{T}(\Theta)] \quad (\text{A2})$$

provided that  $\mathbf{T}$  is antisymmetric. The elements of  $\mathbf{T}(\Theta)$  are given by

$$\mathbf{T}(\Theta)_{mn} = \Theta_{ia}^* \delta_{mi} \delta_{na} - \Theta_{ai} \delta_{ma} \delta_{ni}, \quad (\text{A3})$$

which, again, is antisymmetric and therefore ensures unitarity of  $\mathbf{U}$ . Therefore, we can write the expression for  $\mathbf{U}$  more clearly as

$$\mathbf{U}(\Theta) = \exp \begin{bmatrix} \mathbf{0} & \Theta^\dagger \\ -\Theta & \mathbf{0} \end{bmatrix}. \quad (\text{A4})$$

$\Theta$  is a  $(\text{virt} \times \text{occ})$ -dimensional matrix where the elements  $\Theta_{ai}$  are the orbital rotation angles that minimize the Hartree-Fock energy. When canonical Hartree-Fock orbitals are used as the reference  $\mathbf{C}^0$ , the rotation angles  $\Theta_{ai}$  are zero—and  $\mathbf{U}(\Theta)$  is the identity matrix—by virtue of the variational optimization of the orbitals.

An expression for  $\mathbf{M}$  can be determined from the Hartree-Fock orthogonality condition

$$\mathbf{C}^\dagger \mathbf{S} \mathbf{C} = \mathbf{U}^\dagger \mathbf{M}^\dagger \mathbf{C}^{\dagger 0} \mathbf{S} \mathbf{C}^0 \mathbf{M} \mathbf{U} = \mathbf{I}. \quad (\text{A5})$$

$\mathbf{U}$  is unitary and may be removed by left and right multiplication by  $\mathbf{U}$  and  $\mathbf{U}^\dagger$ , respectively. Requiring that  $\mathbf{M}$  is both symmetric and invertible allows us to end up with the expression

$$\mathbf{M} = (\mathbf{C}^{0\dagger} \mathbf{S} \mathbf{C}^0)^{-1/2}, \quad (\text{A6})$$

which is by no means a unique solution for  $\mathbf{M}$ . Putting the previous results together, the final parametrization of the molecular orbitals yields the expression

$$\mathbf{C} = \mathbf{C}^0 (\mathbf{C}^{0\dagger} \mathbf{S} \mathbf{C}^0)^{-1/2} \exp \begin{bmatrix} \mathbf{0} & \Theta^\dagger \\ -\Theta & \mathbf{0} \end{bmatrix} \quad (\text{A7})$$

where the elements  $C_{\mu p}$  are given as

$$C_{\mu p} = \sum_{qr} C_{\mu q}^0 \left[ \sum_{\lambda\sigma} C_{\lambda q}^0 S_{\lambda\sigma} C_{\sigma r}^0 \right]^{-1/2} U_{rp}. \quad (\text{A8})$$

For a reference set of orbitals  $\mathbf{C}^0$ , the parametrization of the orbitals depends on  $\mathbf{S}$  and  $\Theta$ . We will want to consider partial derivatives of the orbitals with respect to these quantities. First we will consider derivatives of the orbitals with respect to the atomic orbital overlap matrix  $\mathbf{S}$

$$\frac{\partial C_{\mu p}}{\partial S_{\lambda\sigma}} = -\frac{1}{2} \sum_q C_{\mu q}^0 C_{\lambda r}^0 C_{\sigma r}^0 \left[ \sum_{\lambda\sigma} C_{\lambda q}^0 S_{\lambda\sigma} C_{\sigma r}^0 \right]^{-3/2} U_{rp}. \quad (\text{A9})$$

Without loss of generality, we can consider the case where the MO coefficients are small perturbations of the canonical

Hartree-Fock orbitals. Therefore we consider the limit where  $\mathbf{C}^0 \rightarrow \mathbf{C}$  and  $\mathbf{\Theta} \rightarrow \mathbf{0}$ , such that  $\mathbf{U} \rightarrow \mathbf{I}$ . This allows us to set  $\mathbf{M} = \mathbf{I}$  and  $\mathbf{U} = \mathbf{I}$ , respectively, when evaluating expressions after differentiation. Therefore, the above expression simplifies to

$$\left. \frac{\partial C_{\mu p}}{\partial S_{\lambda \sigma}} \right|_{\mathbf{C}^0=\mathbf{C}, \mathbf{\Theta}=\mathbf{0}} = -\frac{1}{2} \sum_q C_{\mu q} C_{\lambda r} C_{\sigma r} \delta_{qr} \delta_{rp} \quad (\text{A10})$$

$$= -\frac{1}{2} \sum_q C_{\mu q} C_{\lambda q} C_{\sigma p} \quad (\text{A11})$$

such that

$$\frac{\partial C_{\mu p}}{\partial S_{\lambda \sigma}} \frac{dS_{\lambda \sigma}}{dx} = -\frac{1}{2} \sum_q C_{\mu q} C_{\lambda q} C_{\sigma p} \frac{dS_{\lambda \sigma}}{dx} \quad (\text{A12})$$

with these, we are ready for handling general derivatives of *ab initio* energies involving the overlap, as in Eq. (18).

For the one body terms

$$\begin{aligned} \sum_{\lambda \sigma} \frac{\partial}{\partial S_{\lambda \sigma}} \left[ \sum_{qm} \gamma_{qm} h_{qm} \right] \frac{dS_{\lambda \sigma}}{dx} &= \sum_{\lambda \sigma} \frac{\partial}{\partial S_{\lambda \sigma}} \left[ \sum_{qm} \gamma_{qm} \sum_{\mu\nu} C_{\mu q} C_{\nu m} h_{\mu\nu} \right] \frac{dS_{\lambda \sigma}}{dx} \\ &= \sum_{qm} \gamma_{qm} \sum_{\mu\nu\lambda\sigma} \left[ \frac{\partial C_{\mu q}}{\partial S_{\lambda \sigma}} C_{\nu m} h_{\mu\nu} + C_{\mu q} \frac{\partial C_{\nu m}}{\partial S_{\lambda \sigma}} h_{\mu\nu} \right] \frac{dS_{\lambda \sigma}}{dx} \\ &= \sum_{qm} \gamma_{qm} \sum_{\mu\nu\lambda\sigma} \left[ -\frac{1}{2} \sum_p C_{\mu p} C_{\lambda p} C_{\sigma q} C_{\nu m} h_{\mu\nu} - \frac{1}{2} \sum_p C_{\mu q} C_{\nu p} C_{\lambda p} C_{\sigma m} h_{\mu\nu} \right] \frac{dS_{\lambda \sigma}}{dx} \\ &= -\frac{1}{2} \sum_{pqm} \gamma_{qm} h_{pm} \frac{dS_{pq}}{dx} - \frac{1}{2} \sum_{pqm} \gamma_{qm} h_{qp} \frac{dS_{pm}}{dx} \\ &= -\sum_{pqm} \gamma_{qm} h_{pm} \frac{dS_{pq}}{dx}. \end{aligned} \quad (\text{A13})$$

The two body terms are slightly more complicated. For the two body terms, we can rearrange to isolate the terms depending on the overlap

$$\sum_{\alpha\beta} \frac{\partial}{\partial S_{\alpha\beta}} \left[ \sum_{pqrs} \Gamma_{pqrs} g_{pqrs} \right] \frac{dS_{\alpha\beta}}{dx} = \sum_{\alpha\beta} \sum_{pqrs} \Gamma_{pqrs} \left[ \frac{\partial}{\partial S_{\alpha\beta}} \sum_{\mu\nu\lambda\sigma} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s} g_{\mu\nu\lambda\sigma} \right] \frac{dS_{\alpha\beta}}{dx}, \quad (\text{A14})$$

next, we can focus on evaluating the terms in brackets above,

$$\begin{aligned} \frac{\partial}{\partial S_{\alpha\beta}} \sum_{\mu\nu\lambda\sigma} \frac{\partial}{\partial S_{\alpha\beta}} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s} g_{\mu\nu\lambda\sigma} &= \sum_{\mu\nu\lambda\sigma} \frac{\partial C_{\mu p}}{\partial S_{\alpha\beta}} C_{\nu q} C_{\lambda r} C_{\sigma s} g_{\mu\nu\lambda\sigma} + \sum_{\mu\nu\lambda\sigma} C_{\mu p} \frac{\partial C_{\nu q}}{\partial S_{\alpha\beta}} C_{\lambda r} C_{\sigma s} g_{\mu\nu\lambda\sigma} \\ &\quad + \sum_{\mu\nu\lambda\sigma} C_{\mu p} C_{\nu q} \frac{\partial C_{\lambda r}}{\partial S_{\alpha\beta}} C_{\sigma s} g_{\mu\nu\lambda\sigma} + \sum_{\mu\nu\lambda\sigma} C_{\mu p} C_{\nu q} C_{\lambda r} \frac{\partial C_{\sigma s}}{\partial S_{\alpha\beta}} g_{\mu\nu\lambda\sigma} \\ &= -\frac{1}{2} \sum_{\mu t} C_{\mu t} C_{\alpha t} C_{\beta p} g_{\mu qrs} - \frac{1}{2} \sum_{\nu t} C_{\nu t} C_{\alpha t} C_{\beta q} g_{\mu pvr} \\ &\quad - \frac{1}{2} \sum_{\lambda t} C_{\lambda t} C_{\alpha t} C_{\beta r} g_{\mu pqt} - \frac{1}{2} \sum_{\sigma t} C_{\sigma t} C_{\alpha t} C_{\beta s} g_{\mu pqr} \\ &= -\frac{1}{2} \sum_t C_{\alpha t} C_{\beta p} g_{tqrs} - \frac{1}{2} \sum_t C_{\alpha t} C_{\beta q} g_{ptrs} \\ &\quad - \frac{1}{2} \sum_t C_{\alpha t} C_{\beta r} g_{pqts} - \frac{1}{2} \sum_t C_{\alpha t} C_{\beta s} g_{pqrt}, \end{aligned} \quad (\text{A15})$$

so, upon plugging back in to (A14),

$$\sum_{\alpha\beta} \sum_{pqrs} \Gamma_{pqrs} \left[ \frac{\partial}{\partial S_{\alpha\beta}} \sum_{\mu\nu\lambda\sigma} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\sigma s} g_{\mu\nu\lambda\sigma} \right] \frac{dS_{\alpha\beta}}{dx}$$

$$\begin{aligned}
&= -\frac{1}{2} \sum_{\alpha\beta} \left[ \sum_{pqrst} \Gamma_{pqrs} C_{\alpha t} C_{\beta p} g_{tqrs} + \sum_{pqrst} \Gamma_{pqrs} C_{\alpha t} C_{\beta q} g_{ptrs} + \sum_{pqrst} \Gamma_{pqrs} C_{\alpha t} C_{\beta r} g_{pqts} + \sum_{pqrst} \Gamma_{pqrs} C_{\alpha t} C_{\beta s} g_{pqrt} \right] \frac{dS_{\alpha\beta}}{dx} \\
&= -2 \sum_{pq} \frac{dS_{pq}}{dx} \sum_{rst} \Gamma_{qrst} g_{prst}
\end{aligned}$$

where in the last line we used the permutational symmetries of the real two-electron integrals and the fact that indices  $\{p, q, r, s, t\}$  are dummy indices. Putting it all together

$$\frac{\partial E}{\partial \mathbf{S}} \frac{d\mathbf{S}}{dx} = - \sum_{pqm} \gamma_{qm} h_{pm} \frac{dS_{pq}}{dx} - 2 \sum_{pq} \frac{dS_{pq}}{dx} \sum_{rst} g_{prst} \Gamma_{qrst}, \quad (\text{A16})$$

which, reindexing and collecting terms by RDM, gives the final expression

$$\begin{aligned}
\frac{dE}{dx} &= \sum_{pq} \gamma_{pq} \left[ \frac{dh_{pq}}{dx} - \sum_m h_{mq} \frac{dS_{mp}}{dx} \right] \\
&+ \sum_{pqrs} \Gamma_{pqrs} \left[ \frac{dg_{pqrs}}{dx} - 2 \sum_t g_{tqrs} \frac{dS_{tp}}{dx} \right]. \quad (\text{A17})
\end{aligned}$$

## APPENDIX B: ADDITIONAL CALCULATIONS FOR NISQ FORCE ESTIMATION

In this Appendix we provide some additional calculations and numerics relevant to the estimation of forces on NISQ quantum computers that were not considered in the main text.

### 1. Importance sampling to optimize 1-norm estimation

In the main text we studied the effect of importance sampling to minimize the 2-norm of the error in the force vector. In this section we repeat this analysis to minimize the 1-norm instead. Given that we have a  $D$ -component vector, our aim is to minimize the total number of shots needed to ensure that the estimate of the gradient is  $\epsilon$ -close to the true value of the gradient with respect to the vector 1-norm. That is to say, we wish to minimize the expectation value over data collected of the 1-norm of the difference between the true gradient and this operation, which we bound using Jensen's inequality as

$$\begin{aligned}
\mathbb{E} \left( \left\| \frac{\widetilde{dE}}{d\mathbf{R}} - \frac{dE}{d\mathbf{R}} \right\|_1 \right) &= \mathbb{E} \left( \sum_i \sqrt{\left( \frac{\widetilde{dE}}{dR_i} - \frac{dE}{dR_i} \right)^2} \right) \\
&\leq \left( \sum_i \mathbb{E} \left( \sqrt{\left( \frac{\widetilde{dE}}{dR_i} - \frac{dE}{dR_i} \right)^2} \right) \right). \quad (\text{B1})
\end{aligned}$$

Let us assume that our estimator for the component of the gradient is unbiased with variance  $\sigma_{F,i}^2$ ,

$$\mathbb{E} \left( \left\| \frac{\widetilde{dE}}{d\mathbf{R}} - \frac{dE}{d\mathbf{R}} \right\|_1 \right) \leq \sum_i \sigma_{F,i}. \quad (\text{B2})$$

Next let us assume that each shot of  $\frac{\widetilde{dE}}{d\mathbf{R}}$  is estimated by using importance sampling to stochastically sample terms from the gradient operator. Let us assume that the variance yielded from the importance sampling distribution for component  $i$  is at most  $|\alpha_i|^2$ .

Now assume that the operator computed has an absolute value of at most  $|\alpha_i|$  when computing the gradient. This implies, if we use the sample mean as the estimator of  $\sigma_{F,i}$  using  $M_i$  shots then the standard deviation satisfies

$$\sigma_{F,i} = \frac{|\alpha_i|}{\sqrt{M_i}}, \quad (\text{B3})$$

with  $M$  being our target for  $\sum_i M_i$ . The first problem that we wish to tackle is the problem of distributing our  $M$  measurements over these different possibility in a way that minimizes the error. This fortunately, can be expressed as a simple constrained optimization problem with Lagrange multiplier  $\lambda$  wherein the optimal solution is found when the derivative of the Lagrangian

$$\mathcal{L} = \sum_i \frac{|\alpha_i|}{\sqrt{M_i}} + \lambda \left( \sum_i M_i - M \right), \quad (\text{B4})$$

with respect to all parameters:  $M_1, \dots, M_D$  and  $\lambda$ , is equal to 0. Setting this condition gives us  $D$  nontrivial equations of the form

$$M_k = \left( \frac{|\alpha_k|}{2\lambda} \right)^{2/3} \quad \forall k. \quad (\text{B5})$$

The Lagrange multiplier is, as yet, unspecified. We can constrain it by demanding that the expected error in the 1-norm is bounded above by  $\epsilon$ . Specifically, we have that

$$\sum_k \frac{|\alpha_k|}{\sqrt{M_k}} = (2\lambda)^{1/3} \left( \sum_k |\alpha_k|^{2/3} \right) = \epsilon. \quad (\text{B6})$$

Therefore we can choose the Lagrange multiplier to satisfy

$$\lambda = \frac{\epsilon^3}{2 \left( \sum_k |\alpha_k|^{2/3} \right)^3} \Rightarrow M_k = \frac{|\alpha_k|^{2/3} \left( \sum_k |\alpha_k|^{2/3} \right)^2}{\epsilon^2}. \quad (\text{B7})$$

This further implies that the total number of shots, as a function of the individual standard deviations of the individual terms is

$$M = \frac{\left( \sum_k |\alpha_k|^{2/3} \right)^3}{\epsilon^2}. \quad (\text{B8})$$

Next, let us assume that for each shot of a particular component of the force vector we use an importance sampling procedure wherein for each of the  $Q_{ij}$  terms that need to be measured to estimate the gradient. We then choose our

probability such that

$$Pr_{ij} = \frac{\|Q_{ij}\|}{\sum_j \|Q_{ij}\|} \quad (\text{B9})$$

and further consider the operator  $Y_{ij} = Q_{ij}/Pr_{ij}$  choose to measure, for fixed  $i$ ,  $Q_{ij}$  with probability  $Pr_{ij}$ . Thus  $\sum_j \langle \psi | Q_{ij} | \psi \rangle = \sum_j \langle \psi | Y_{ij} | \psi \rangle Pr_{ij}$ . The population variance of this estimate is then

$$\begin{aligned} \mathbb{E}(Y_{ij}^2) - \mathbb{E}(Y_{ij})^2 &= \sum_j \frac{(\langle \psi | Q_{ij} | \psi \rangle)^2}{Pr_{ij}} - \left( \sum_j \langle \psi | Q_{ij} | \psi \rangle \right)^2 \\ &\leq \left( \frac{\sum_k \|Q_{ik}\| \langle \psi | Q_{ij} | \psi \rangle}{\|Q_{ij}\|} \right)^2 \\ &\leq \left( \sum_k \|Q_{ik}\| \right)^2. \end{aligned} \quad (\text{B10})$$

Thus the overall number of measurements the  $Q_{ij}$  needed to reduce the variance of the one-norm of the gradient to epsilon is at most

$$M \leq \frac{(\sum_i (\sum_j \|Q_{ij}\|)^{2/3})^3}{\epsilon^2}. \quad (\text{B11})$$

## 2. Localization of molecular orbitals

Canonical molecular orbitals (CMOs) from a Hartree-Fock calculation are only one of many possibilities to express the Hamiltonian and the force operators - any other orthonormal molecular orbitals might be used too. A prominent example for another possible choice are localized molecular orbitals (LMOs). In contrast to CMOs, which typically have support over the entire molecule, LMOs are restricted to a certain spatial region of the molecule [118–120]. Given a set of CMOs, one finds LMOs by applying a unitary transformation  $U$ ,

$$\tilde{\phi}_q(r) = \sum_p \phi_p(r) U_{pq}.$$

The unitary  $U$  is altered with the goal to optimize the locality of the molecular orbitals. Different schemes with different cost functions exists, such as Edmiston-Ruedenberg (ER) [118], Pipek-Mezey (PM) [119], or Foster-Boys (FB) [120]. In Ref. [98], it was observed that LMOs in comparison to CMOs have a positive, lowering, effect on the 1-norm of the qubit Hamiltonian. As the force operators force operators include terms with  $\mathcal{O}(1/r^2)$  while the Coulomb potential is  $\mathcal{O}(1/r)$ , we expect that using LMOs will also lower the 1-norms of the force operators. As different localization schemes have similar effects on the 1-norm of the Hamiltonian [98], we restrict ourselves to the ER localization scheme in the following. Using ER, we aim to maximize

$$\mathcal{L}_{\text{ER}} = \sum_p \iint |\tilde{\phi}_p(r_1)|^2 \frac{1}{r_{12}} |\tilde{\phi}_p(r_2)|^2 dr_1 dr_2 \quad (\text{B12})$$

within the occupied/virtual molecular orbitals in order to find localized molecular orbitals. We then use the new set of LMOs to express the Hamiltonian and the force operators. We note that the fermionic shadow tomography scheme and the basis rotation grouping strategy, see Sec. III, are not affected by any

localization. The basis rotation grouping scheme is based on a diagonalization, which is independent from the basis used, while the variance of the estimator of the fermionic shadow tomography depends on the 2-norm, see Eq. (68), which is preserved under any unitary transformation.

## APPENDIX C: ADDITIONAL DETAILS ON FORCE ESTIMATION BY NUMERICAL DIFFERENTIATION

### 1. Comparison with first-order finite differences

The first-order finite difference formula is

$$\frac{dE}{dR_i} = \frac{E(R + \frac{dR \cdot v_i}{2}) - E(R - \frac{dR \cdot v_i}{2})}{dR} + \epsilon_{fd} \quad (\text{C1})$$

as in the general case, the total error for this will be the result of the error due to the quantum phase estimation and the discretization error due to the first-order finite difference approximation.

By propagating the statistical error, due to quantum phase estimation, where we assume each QPE error is upper bounded by Eq. (C13), we can write

$$\epsilon_{\text{PE}}^{(m=1)} \lesssim \sqrt{2} \frac{\lambda_H \pi}{2T dR}, \quad (\text{C2})$$

where  $T$  is the number of oracle calls used in each QPE evaluation of the energy and  $\lambda_H \lesssim \|H\|_1$ . On the other hand, if we assume that the energy is a well behaved function, then  $\epsilon_{fd}$  is

$$\epsilon_{fd} = \mathfrak{k} \sum_{i=1}^{\infty} \frac{d^{2i} E(R)}{dR^{2i}} \frac{dR^{2i}}{2i!} \approx \mathfrak{k} \mathfrak{a} \times dR^2, \quad (\text{C3})$$

with  $\mathfrak{k}$  a constant factor with the dimension of an energy over a length cubed, and  $\mathfrak{a}$  a dimensionless constant, which upper bounds the derivatives at  $R$ . So the total error is

$$\epsilon_{1-\text{tot}} \approx \sqrt{2} \frac{\lambda_H \pi}{2T dR} + \mathfrak{k} \mathfrak{a} dR^2. \quad (\text{C4})$$

This has the same  $\tilde{O}$  scaling as the upper bound on the error in Eq. (112) for  $m = 1$ . In fact, for  $m = 1 \rightarrow \Gamma = \sqrt{2}$  Eq. (112) is

$$\epsilon = \epsilon_{\text{PE}} + \epsilon_{fd}^{(m=1)} < \frac{\pi \lambda_H}{2T} \frac{6^{3/2}}{dR} + 2\epsilon(8c)^3 dR^2. \quad (\text{C5})$$

The two have the same scalings in terms of  $dR$ ,  $\lambda_H$ , and  $T$ , and they are, in fact, identical up to constant factors in the second term.

### 2. The error contribution from quantum phase estimation and the Holevo variance

The error in phase estimation protocols is typically quantified by the Holevo variance of a phase  $\phi$  [105], which is given by

$$V_H(\phi) = \frac{1}{|\mathbb{E}(e^{-i(\phi - \phi_{\text{true}})})|^2} - 1, \quad (\text{C6})$$

where  $\phi$  is the estimated eigenphase and  $\phi_{\text{true}}$  is the true value of the eigenphase. Specifically, it can be shown that there exists a phase estimation protocol that applies the fundamental

unitary  $T_l$  times and yields an estimate such that

$$\mathbb{E}(\phi - \phi_{\text{true}}) = 0, \text{ and } V_H(\phi) = \tan^2\left(\frac{\pi}{2T_l + 1}\right) \approx \left(\frac{\pi}{2T_l}\right)^2, \quad (\text{C7})$$

where the first expression means that  $(\phi - \phi_{\text{true}})$  is a random variable with zero mean [67,105]. The purpose behind this estimate is to address the problem that the variance of a phase depends on the branch chosen for the phase as the Holevo variance is independent of the branch chosen. Because of these differences, the properties that one would expect of a variance do not necessarily always hold here. Specifically, we

$$\mathbb{V}_H(\phi_j) = \frac{1}{\left(1 - \frac{1}{2}\mathbb{E}(\phi_j - \phi_{j,\text{true}})^2\right)\left(1 - \frac{1}{2}\mathbb{E}(\phi_j - \phi_{j,\text{true}})^2\right) + O(\mathcal{K}_j)} - 1 = \mathbb{V}(\phi_j) + O(\mathcal{K}_j). \quad (\text{C9})$$

Thus we have from the additivity of variance that for independent random variables  $\phi_i \in [-\pi, \pi)$  the variance of the sum of the random variables is

$$\mathbb{V}\left(\sum_{j=1}^m \phi_j\right) = \sum_{j=1}^m \mathbb{V}(\phi_j) \leq \sum_{j=1}^m V_H(\phi_j) + O\left(m \max_j \mathcal{K}_j\right). \quad (\text{C10})$$

We therefore have that for such estimates, in the limit where the fourth and higher moments of the angular distribution are negligible then the statistical errors from phase estimation add in quadrature.

We obtain then that the variance of the estimates yielded by phase estimation can be written as

$$\mathbb{V}(\phi_l) = \frac{\pi^2}{4T_l^2} + O\left(\max_l \mathcal{K}_l\right). \quad (\text{C11})$$

If we apply this to the problem of finding the eigenvalue of the energy with the quantum phase estimation, we have that the phases returned from applying phase estimation on the qubitization walk oracle are  $\pm \cos^{-1}(E_j/\lambda_H)$ , which is to say that the energy estimates needed for the gradient evaluation are from Eq. (106) propagating the IID variances of the  $2m$  estimations we obtain the total mean square error  $\epsilon_{\text{PE}}$ ,

$$\epsilon_{\text{PE}}^2 \lesssim \pi^2 dR^{-2} \lambda_H^2 \sum_l |a_l^{(m)}|^2 (2T_l)^{-2}. \quad (\text{C12})$$

It is equivalent to take Eq. (23) from [67],

$$\epsilon_{\text{PE}}(l)^2 \lesssim \lambda_H^2 \left( \frac{\pi^2}{4T_l^2} + (\epsilon_{\text{prep}} + \pi^2 \epsilon_{\text{QFT}})^2 \right), \quad (\text{C13})$$

neglect the contribution to the errors  $\epsilon_{\text{prep}}$  and  $\epsilon_{\text{QFT}}$  and then propagate assuming the covariances are zero to obtain

would like to apply the additive property of the variance to ensure that the variance of the sum of the energy estimated at each point is at most the sum of the variances. We have that if for any  $\phi_j$  the fourth-order momentum of the unbiased random variable  $(\phi_j - \phi_{j,\text{true}})$ ,

$$\mathcal{K}_j := \mathbb{E}((\phi_j - \phi_{j,\text{true}})^4) \gg \sum_{k=3}^{\infty} \frac{\mathbb{E}(\phi_j - \phi_{j,\text{true}})^{2k}}{2k!}, \quad (\text{C8})$$

Then using the fact that the phase estimation protocol used is unbiased we have from Taylor's theorem that

Eq. (C12).  $\epsilon_{\text{prep}}$  and  $\epsilon_{\text{QFT}}$  are expected to be considerably smaller than the  $1/T_l$  contribution, although they cannot be neglected in general, their consideration would require to find numerical solutions for estimating the scaling of the error.

### 3. Robustness of phase estimation success probability

The aim of this section is to provide a result that can be used in order to argue how the success probability for phase estimation. This result is especially important for the finite difference algorithm discussed above because this result will show that if the eigenvalue gap is sufficiently large then with high probability all the values used for the state preparation can be taken from the same ground state. If this is not true, then in the worst case scenario the ground-state preparation will need to be applied for each perturbed Hamiltonian evaluated. The result is stated below in the following Lemma, and its proof is a direct consequence of matrix calculus and elementary norm inequalities.

*Lemma 1.* Let  $H : \mathbb{R} \mapsto \mathbb{C}^{2^n \times 2^n}$  be a parameterized family of Hamiltonians and let  $|\psi_i(\mathbf{R})\rangle$  be the  $i$ th eigenstate of  $H(\mathbf{R})$  evaluated at parameters  $\mathbf{R} \in \mathbb{R}^{3N_a}$  and let  $E_i(\mathbf{R})$  be the corresponding eigenvalue. If  $|\tilde{\psi}_0(\mathbf{R})\rangle$  is known such that  $\| |\tilde{\psi}_0(\mathbf{R})\rangle - |\psi_0(\mathbf{R})\rangle \| \leq \delta$  and  $\Delta \in \mathbb{R}^{3N_a}$  such that  $\|\Delta\| < (E_1(\mathbf{R}) - E_0(\mathbf{R})) / (4 \max_{\mathbf{R},i} \|\frac{dH(\mathbf{R}')}{dR_i}\|)$  then

$$|\langle \tilde{\psi}_0(\mathbf{R}) | \psi_0(\mathbf{R} + \Delta) \rangle| \geq 1 - \delta - \frac{2\|\Delta\| \max_{\mathbf{R},i} \|\frac{dH(\mathbf{R}')}{dR_i}\|}{E_1(\mathbf{R}) - E_0(\mathbf{R})}.$$

Here, the maximum is taken over all points in a ball of radius  $\|\Delta\|$  around  $\mathbf{R}$  and over all derivative components  $\frac{d}{dR_i}$ .

*Proof.* First we have from the fundamental theorem of calculus, the assumption that  $E_j(\mathbf{R} + s) > E_0(\mathbf{R} + s)$  that

$$\begin{aligned} |\psi(\mathbf{R} + \Delta)\rangle &= |\psi(\mathbf{R})\rangle + \int_0^1 \frac{d|\psi(\mathbf{R} + s\Delta)\rangle}{ds} ds \\ &= |\psi(\mathbf{R})\rangle + \int_0^1 \sum_{j \neq 0} \frac{|\psi_j(\mathbf{R} + s\Delta)\rangle \langle \psi_j(\mathbf{R} + s\Delta) | \frac{dH(\mathbf{R} + s\Delta)}{ds} | \psi(\mathbf{R} + s\Delta)\rangle}{E_j(\mathbf{R} + s\Delta) - E_0(\mathbf{R} + s\Delta)} ds. \end{aligned} \quad (\text{C14})$$

Next we have from the Cauchy-Schwarz inequality

$$\begin{aligned}
 |\langle \psi(\mathbf{R}) | \psi(\mathbf{R} + \Delta) \rangle| &\geq 1 - \left| \int_0^1 \sum_{j \neq 0} \frac{\langle \psi(\mathbf{R}) | \psi_j(\mathbf{R} + s\Delta) \rangle \langle \psi_j(\mathbf{R} + s\Delta) | \frac{dH(\mathbf{R}+s\Delta)}{ds} | \psi(\mathbf{R} + s\Delta) \rangle}{E_j(\mathbf{R} + s\Delta) - E_0(\mathbf{R} + s\Delta)} ds \right| \\
 &\geq 1 - \int_0^1 \left| \sum_{j \neq 0} \frac{\langle \psi(\mathbf{R}) | \psi_j(\mathbf{R} + s\Delta) \rangle \langle \psi_j(\mathbf{R} + s\Delta) | \frac{dH(\mathbf{R}+s\Delta)}{ds} | \psi(\mathbf{R} + s\Delta) \rangle}{E_j(\mathbf{R} + s\Delta) - E_0(\mathbf{R} + s\Delta)} \right| ds \\
 &\geq 1 - \int_0^1 \sqrt{\left| \sum_{j \neq 0} \frac{\langle \psi(\mathbf{R} + s\Delta) | \frac{dH(\mathbf{R}+s\Delta)}{ds} | \psi_j(\mathbf{R} + s\Delta) \rangle \langle \psi_j(\mathbf{R} + s\Delta) | \frac{dH(\mathbf{R}+s\Delta)}{ds} | \psi(\mathbf{R} + s\Delta) \rangle}{(E_1(\mathbf{R} + s\Delta) - E_0(\mathbf{R} + s\Delta))^2} \right|} ds \\
 &\geq 1 - \int_0^1 \frac{\| \frac{dH(\mathbf{R}+s\Delta)}{ds} \|}{E_1(\mathbf{R} + s\Delta) - E_0(\mathbf{R} + s\Delta)} ds \\
 &\geq 1 - \frac{\| \Delta \| \max_s \| \frac{dH(\mathbf{R}+s\Delta)}{ds} \|}{\min_s |E_1(\mathbf{R} + s\Delta) - E_0(\mathbf{R} + s\Delta)|} \geq 1 - \frac{\| \Delta \| \max_s \| \frac{dH(\mathbf{R}+s\Delta)}{ds} \|}{E_1(\mathbf{R}) - E_0(\mathbf{R}) - 2\| \Delta \| \max_s \| \frac{dH(\mathbf{R}+s\Delta)}{ds} \|} \\
 &\geq 1 - \frac{2\| \Delta \| \max_s \| \frac{dH(\mathbf{R}+s\Delta)}{ds} \|}{E_1(\mathbf{R}) - E_0(\mathbf{R})} \geq 1 - \frac{2\| \Delta \| \max_{\mathbf{R},i} \| \frac{dH(\mathbf{R})}{dR_i} \|}{E_1(\mathbf{R}) - E_0(\mathbf{R})} \tag{C15}
 \end{aligned}$$

where the last inequality only follows under the assumption that  $\max_s \| \frac{d}{ds} H(\mathbf{R} + s\Delta) \| \leq (E_1(\mathbf{R}) - E_0(\mathbf{R}))/4\Delta$ .

Finally if  $\| |\tilde{\psi}_0(\mathbf{R} + \Delta) \rangle - |\psi_0(\mathbf{R} + \Delta) \rangle \| \leq \delta$  then there exists  $\delta' \leq \delta$  and  $|\phi\rangle$  such that  $|\tilde{\psi}_0(\mathbf{R} + \Delta) \rangle = |\psi_0(\mathbf{R} + \Delta) \rangle + \delta'|\phi\rangle$ . We therefore have from the reverse triangle inequality that under the assumptions of the lemma

$$|\langle \tilde{\psi}_0(\mathbf{R} + \Delta) | \psi_0(\mathbf{R}) \rangle| \geq 1 - \delta - \frac{2\| \Delta \| \max_s \| \frac{dH(\mathbf{R}+s\Delta)}{ds} \|}{E_1(\mathbf{R}) - E_0(\mathbf{R})} \geq 1 - \delta - \frac{2\| \Delta \| \max_{\mathbf{R},i} \| \frac{dH(\mathbf{R})}{dR_i} \|}{E_1(\mathbf{R}) - E_0(\mathbf{R})}. \tag{C16}$$

■

This shows that provided the eigenvalue gap is significant relative to the norm of the derivative operator then the reduction in success probability for phase estimation is minimal. This means that in practice, small shifts in the Hamiltonian will not meaningfully impact the overlaps provided these gap assumptions are met.

*Corollary 2.* Assume that the displacements  $\Delta_1 \in \mathbb{R}^{3N_a}$  and  $\Delta_2 \in \mathbb{R}^{3N_a}$  individually satisfy the assumptions of the previous lemma we then have that

$$\begin{aligned}
 &|\langle \tilde{\psi}_0(x + \Delta_1) | \tilde{\psi}_0(x + \Delta_2) \rangle| \\
 &\geq 1 - 3\delta - \frac{6 \max\{\| \Delta_1 \|, \| \Delta_2 \| \} \max_{\mathbf{R},i} \| \frac{dH(\mathbf{R})}{dR_i} \|}{E_1(x) - E_0(x)}.
 \end{aligned}$$

*Proof.* Under the assumptions of the previous theorem we have that there exist  $\delta_1$  and  $\delta_2$  that are bounded above in absolute value by  $\delta + \frac{4 \max\{\| \Delta_1 \|, \| \Delta_2 \| \} \max_{\mathbf{R},i} \| \frac{dH(\mathbf{R})}{dR_i} \|}{E_1(x) - E_0(x)}$  and  $|\phi_1\rangle, |\phi_2\rangle$  such that

$$\begin{aligned}
 &|\langle \tilde{\psi}_0(x + \Delta_1) | \tilde{\psi}_0(x + \Delta_2) \rangle| \\
 &= |1 + \delta_2 \langle \tilde{\psi}_0(x + \Delta_1) | \phi_2 \rangle + \delta_1 \langle \phi_1 | \tilde{\psi}_0(x + \Delta_2) \rangle \\
 &\quad + \delta_1 \delta_2 \langle \tilde{\psi}_0(x + \Delta_1) | \tilde{\psi}_0(x + \Delta_2) \rangle| \\
 &\geq 1 - 3 \max\{|\delta_1|, |\delta_2|\}. \tag{C17}
 \end{aligned}$$

The result then immediately follows by the assumed bounds on  $\delta_1, \delta_2$ . ■

#### 4. Optimal T for numerical differentiation of the energy with Lagrange multipliers

In order to find the optimal time to achieve a given error in the evaluation of the gradient with finite differences, we need to minimize  $T = \sum_l T_l$ , with the condition  $\epsilon_{PE}^2 = \pi^2 dR^{-2} \lambda_H^2 \sum_l |a_l^{(m)}|^2 (2T_l)^{-2}$ .

$$\mathcal{L} = \sum_{l=-m}^m T_l + \lambda (\epsilon_{PE}^2 - \tilde{\epsilon}_{PE}^2), \tag{C18}$$

$$\mathcal{L} = \sum_{l=-m}^m T_l + \lambda \left( \pi^2 dR^{-2} \lambda_H^2 \sum_{l=-m}^m |a_l^{(m)}|^2 (2T_l)^{-2} - \tilde{\epsilon}_{PE}^2 \right). \tag{C19}$$

We have then

$$\frac{\partial \mathcal{L}}{\partial T_l} = 1 - 2^{-1} \lambda (\pi^2 dR^{-2} \lambda_H^2 |a_l^{(m)}|^2 T_l^{-3}) = 0. \tag{C20}$$

Solving in  $T_l$  we have

$$T_l = (\lambda 2^{-1} \pi^2 dR^{-2} \lambda_H^2 |a_l^{(m)}|^2)^{\frac{1}{3}}, \tag{C21}$$

substituting this back in Eq. (C12) with equal sign, which is our condition for the Lagrangian, and solving in lambda, we have

$$\epsilon_{PE}^2 = \pi^2 dR^{-2} \lambda_H^2 \sum_{l=m}^m (|a_l^{(m)}|^2 2^{-2} \lambda^{-\frac{2}{3}} 2^{\frac{2}{3}} \pi^{-\frac{4}{3}} dR^{\frac{4}{3}} \lambda_H^{-\frac{4}{3}} |a_l^{(m)}|^{-\frac{4}{3}}), \tag{C22}$$



$$\epsilon_{\text{PE}}^2 = \pi^{\frac{2}{3}} dR^{-\frac{2}{3}} \lambda_{\text{H}}^{\frac{2}{3}} \lambda^{-\frac{2}{3}} 2^{-\frac{4}{3}} \sum_{l=m}^m |a_l^{(m)}|^{\frac{2}{3}}, \quad (\text{C23})$$

$$\lambda = \epsilon_{\text{PE}}^{-3} \pi dR^{-1} \lambda_{\text{H}} 2^{-2} \left( \sum_{l=m}^m |a_l^{(m)}|^{\frac{2}{3}} \right)^{\frac{3}{2}}, \quad (\text{C24})$$

and substituting this in Eq. (C21)

$$T_l = \epsilon_{\text{PE}}^{-1} \pi dR^{-1} \lambda_{\text{H}} 2^{-1} |a_l^{(m)}|^{\frac{2}{3}} \left( \sum_{l=m}^m |a_l^{(m)}|^{\frac{2}{3}} \right)^{\frac{1}{2}}. \quad (\text{C25})$$

Summing over  $l$  and using the shorthand  $\Gamma$  from Eq. (C32),

$$\begin{aligned} T &= \sum_{l=-m}^m T_l = \epsilon_{\text{PE}}^{-1} \pi dR^{-1} \lambda_{\text{H}} 2^{-1} \\ &\quad \times \sum_{l=-m}^m |a_l^{(m)}|^{\frac{2}{3}} \left( \sum_{l=m}^m |a_l^{(m)}|^{\frac{2}{3}} \right)^{\frac{1}{2}} = \quad (\text{C26}) \\ &\epsilon_{\text{PE}}^{-1} \pi dR^{-1} \lambda_{\text{H}} 2^{-1} \Gamma^{\frac{3}{2}} \Gamma^{\frac{1}{3}} \\ &= \epsilon_{\text{PE}}^{-1} \pi dR^{-1} \lambda_{\text{H}} 2^{-1} \Gamma^{\frac{5}{6}} \leq \epsilon_{\text{PE}}^{-1} \pi dR^{-1} \lambda_{\text{H}} 2^{-1} \Gamma, \quad (\text{C27}) \end{aligned}$$

In summary, using the Lagrange multipliers method, solving in  $T_l$  and summing over all  $l$  we obtain that the total time necessary to obtain a target statistical error  $\epsilon_{\text{PE}}$ ,

$$T = \sum_l T_l \leq \pi \epsilon_{\text{PE}}^{-1} dR^{-1} \lambda_{\text{H}} 2^{-1} \left[ \sum_{l=-m}^m |a_l^{(m)}|^{2/3} \right]^{3/2}. \quad (\text{C28})$$

So that we can write

$$T < \epsilon_{\text{PE}}^{-1} \pi dR^{-1} \lambda_{\text{H}} 2^{-1} \Gamma \quad (\text{C29})$$

and inverting Eq. (C29)

$$\epsilon_{\text{PE}} \leq \frac{\pi \lambda_{\text{H}} \Gamma}{2dR T} < \frac{\pi \lambda_{\text{H}} 6^{3/2} m^{1/2}}{2dR T}. \quad (\text{C30})$$

It remains to place a bound on  $\Gamma$ . We have that

$$\begin{aligned} |a_l^{(m)}| &= \frac{1}{|l|} \frac{\binom{m}{|l|}}{\binom{m+|l|}{|l|}} = \frac{1}{|l|} \frac{\binom{m}{|l|}}{\binom{m+1}{|l|}} \left( \frac{\binom{m+1}{|l|}}{\binom{m+|l|}{|l|}} \right) \\ &= \frac{1}{|l|} \left( \frac{m!}{(m-|l|)!} \frac{(m+1-|l|)!}{(m+1)!} \right) \left( \frac{\binom{m+1}{|l|}}{\binom{m+|l|}{|l|}} \right) \\ &\leq \frac{1}{|l|} \binom{m}{m+1} \left( \frac{\binom{m+1}{|l|}}{\binom{m+|l|}{|l|}} \right) \leq \frac{1}{|l|} \binom{m}{m+1}, \quad (\text{C31}) \end{aligned}$$

and taking a continuum approximation yields

$$\begin{aligned} \Gamma^{\frac{2}{3}} &= \sum_{l=-m}^m |a_l^{(m)}|^{\frac{2}{3}} = 2 \sum_{l=1}^m |a_l^{(m)}|^{\frac{2}{3}} \leq 2 \sum_{l=1}^m |l|^{-\frac{2}{3}} \\ &\leq 2 \left( 1 + \int_1^{m-1} |l|^{-\frac{2}{3}} dl \right) = 6(m-1)^{\frac{1}{3}} - 4 < 6m^{\frac{1}{3}}. \quad (\text{C32}) \end{aligned}$$

#### APPENDIX D: MOMENTUM STATE PREPARATION FOR FIRST-QUANTIZED PLANE-WAVE SIMULATIONS

In order to block encode the force operator in first quantization in a plane wave basis, we require the ability to prepare a momentum state proportional to

$$\sum_{v \neq 0} \frac{\sqrt{|v_x|}}{\|v\|} |v\rangle. \quad (\text{D1})$$

The challenge for this preparation is that there are  $\Theta(N)$  values of  $v$  in this state which means that the Toffoli gate complexity scales like  $\Theta(N)$  if we prepare it naively. Instead, we will describe a circuit implementation which has an asymptotic cost of  $\mathcal{O}(\text{polylog}(N, 1/\epsilon))$ , where  $\epsilon$  is the error of the block-encoded operator.

Note that this is different from performing Hamiltonian simulation in first quantization [57,58]. There our goal is to prepare a state proportional to  $\sum_{v \neq 0} \frac{1}{\|v\|} |v\rangle$ , and we achieve that by grouping the set of momenta into nested boxes

$$\begin{aligned} \{v \mid |v_x| < 2^{\mu-1} \wedge |v_y| < 2^{\mu-1} \wedge |v_z| < 2^{\mu-1} \\ \wedge (|v_x| \geq 2^{\mu} \vee |v_y| \geq 2^{\mu} \vee |v_z| \geq 2^{\mu})\}. \quad (\text{D2}) \end{aligned}$$

The number of boxes scales only like  $\mathcal{O}(\log N)$ , so a superposition over all nested boxes can be prepared efficiently. Meanwhile, the momenta within each nested box do not vary significantly in size, which means that the coefficients are nearly uniformly distributed and a superposition over those momenta can also be efficiently prepared. However, we now have the additional weighting of  $\sqrt{|v_x|}$  for block encoding the derivative operator, whose value changes dramatically within each box. As a result, the preparation subroutine would succeed with a very small probability which would need to be amplified with significant overhead.

To address this issue, we consider grouping the momenta into nested boxes for  $(v_x, v_y)$  alone,

$$\begin{aligned} C_{\mu} := \{(v_y, v_z) \mid |v_y| < 2^{\mu-1} \wedge |v_z| < 2^{\mu-1} \\ \wedge (|v_y| \geq 2^{\mu-2} \vee |v_z| \geq 2^{\mu-2})\}. \quad (\text{D3}) \end{aligned}$$

We now describe a quantum circuit that prepares the momentum state in (D1). We start by preparing

$$\frac{1}{\sqrt{2^{n+1}}} \sum_{\mu=2}^n 2^{\mu/2} |\mu\rangle. \quad (\text{D4})$$

The method for preparation of this state was given in Sec. II B of [58], and has Toffoli complexity linear in  $n$  (given a  $|T\rangle$  state that is used catalytically to implement controlled Hadamards). Here we have allowed this state to be subnormalized to account for failure of this preparation. Given  $\mu$ , we aim to prepare  $y, z$  values within box  $C_{\mu}$ . This can be achieved by preparing the state  $|\mu\rangle$  in unary, and performing controlled Hadamards on the  $y$  and  $z$  registers with complexity  $\mathcal{O}(\log N)$ . We encode  $y, z$  as signed integers, and also eliminate the negative zero as a failure with complexity  $\mathcal{O}(\log N)$ . That gives

$$\frac{1}{\sqrt{2^{n+1}}} \sum_{\mu=2}^n 2^{\mu/2} \frac{1}{2^{\mu}} \sum_{v_y, v_z = -(2^{\mu-1}-1)}^{2^{\mu-1}-1} |\mu, v_y, v_z\rangle. \quad (\text{D5})$$

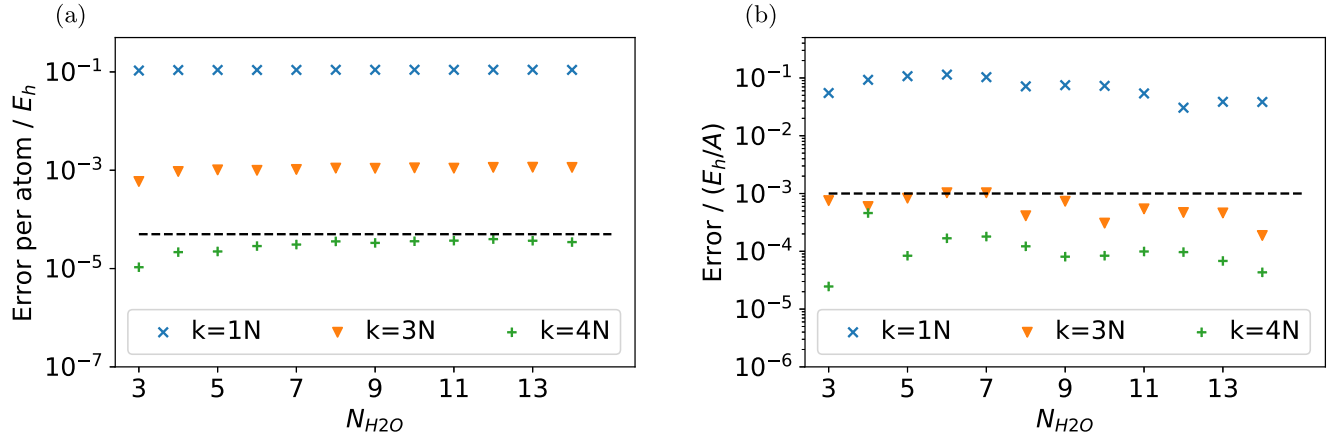


FIG. 7. The error from using Lanczos algorithm to decompose the matricized ERI of the water clusters on the (a) expectation value of Hamiltonian and (b) the force operators with respect to the CISD wave function.

Next, we can eliminate the inner squares in  $v_y, v_z$  using inequality tests [with complexity  $\mathcal{O}(\log N)$ ] to give

$$\frac{1}{\sqrt{2^{2n+1}}} \sum_{\mu=2}^n 2^{-\mu/2} \sum_{\substack{v_y, v_z = -(2^{\mu-1}-1) \\ |v_y|, |v_z| \geq 2^{\mu-2}}} |\mu, v_y, v_z\rangle. \quad (\text{D6})$$

For  $v_x$  we simply create an equal superposition using Hadamards (without any nesting of boxes), and flag the negative zero [complexity  $\mathcal{O}(\log N)$ ], to give

$$\frac{1}{\sqrt{2^{2n+1}}} \sum_{v_x = -(2^{n-1}-1)}^{2^{n-1}-1} \sum_{\mu=2}^n 2^{-\mu/2} \sum_{(v_y, v_z) \in C_\mu} |\mu, v_x, v_y, v_z\rangle. \quad (\text{D7})$$

Next, we aim to show that

$$\frac{\sqrt{|v_x|}}{\sqrt{v_x^2 + v_y^2 + v_z^2}} \leq 2^{-(\mu-1)/2}, \quad (\text{D8})$$

so that we can perform inequality testing for the state preparation. To maximise this expression, we need to choose  $v_y, v_z$  such that  $v_y^2 + v_z^2$  is minimised for a given box  $C_\mu$ . The minimum is obtained for  $|v_y| = 2^{\mu-2}$  and  $|v_z| = 0$  or  $|v_z| = 2^{\mu-2}$  and  $|v_y| = 0$  (the middle of a side of an inner square), in which case  $v_y^2 + v_z^2 = (2^{\mu-2})^2$ . The maximum as a function of  $v_x$  is then for  $v_x = 2^{\mu-2}$ . That gives us an upper bound of

$$\frac{\sqrt{|v_x|}}{\sqrt{v_x^2 + v_y^2 + v_z^2}} \leq \frac{2^{(\mu-2)/2}}{\sqrt{2^{2\mu-4} + 2^{2\mu-4}}} = 2^{-(\mu-1)/2}, \quad (\text{D9})$$

as required. Therefore, we now prepare another superposition over  $M$  values of  $m$  and test the inequality

$$\frac{|v_x|}{v_x^2 + v_y^2 + v_z^2} > 2^{-(\mu-1)} \frac{m}{M}. \quad (\text{D10})$$

In order to avoid divisions, this inequality test can be rewritten as

$$2^{\mu-1} |v_x| M > m(v_x^2 + v_y^2 + v_z^2). \quad (\text{D11})$$

This inequality test can be evaluated using only additions and multiplications.

The Toffoli complexity of the multiplications is the product of the number of bits, and we have  $\mathcal{O}(\log N)$  bits for the components of  $v$ . Using Lemma 3 of [58] we can give the number of bits for  $m$  as  $\mathcal{O}(\log(N/\epsilon))$  for the case of energy estimation. The case of force estimation is sufficiently similar that the scaling is the same. The squares of components of  $v$  will have complexity  $\mathcal{O}((\log N)^2)$ , and the multiplication by  $m$  will have complexity  $\mathcal{O}(\log N \log(N/\epsilon))$ . The additions have linear complexity in the number of bits, so the total complexity is  $\mathcal{O}(\log N \log(N/\epsilon))$  Toffolis for the complete inequality test.

The resulting state, with success of the inequality test, can be written as

$$\frac{1}{\sqrt{2^{2n+1}}} \sum_{v_x = -(2^{n-1}-1)}^{2^{n-1}-1} \sum_{\mu=2}^n 2^{-\mu/2} \times \sum_{(v_y, v_z) \in C_\mu} |\mu, v_x, v_y, v_z\rangle \frac{1}{\sqrt{M}} \sum_{m=0}^{Q-1} |m\rangle, \quad (\text{D12})$$

with

$$Q = \left\lceil \frac{2^{\mu-1} M |v_x|}{\|v\|^2} \right\rceil. \quad (\text{D13})$$

For  $M$  sufficiently large, one can verify that the above state is proportional to the momentum state we want to prepare. Indeed, for a fixed choice of  $\mu$  and  $v$ , the corresponding amplitude is proportional to

$$\frac{1}{\sqrt{2^{2n+1}}} 2^{-\mu/2} \frac{2^{(\mu-1)/2} \sqrt{|v_x|}}{\|v\|} = \frac{1}{2^{n+1}} \frac{\sqrt{|v_x|}}{\|v\|} \quad (\text{D14})$$

up to the error due to discretization. Now, the probability of success has the asymptotic scaling of

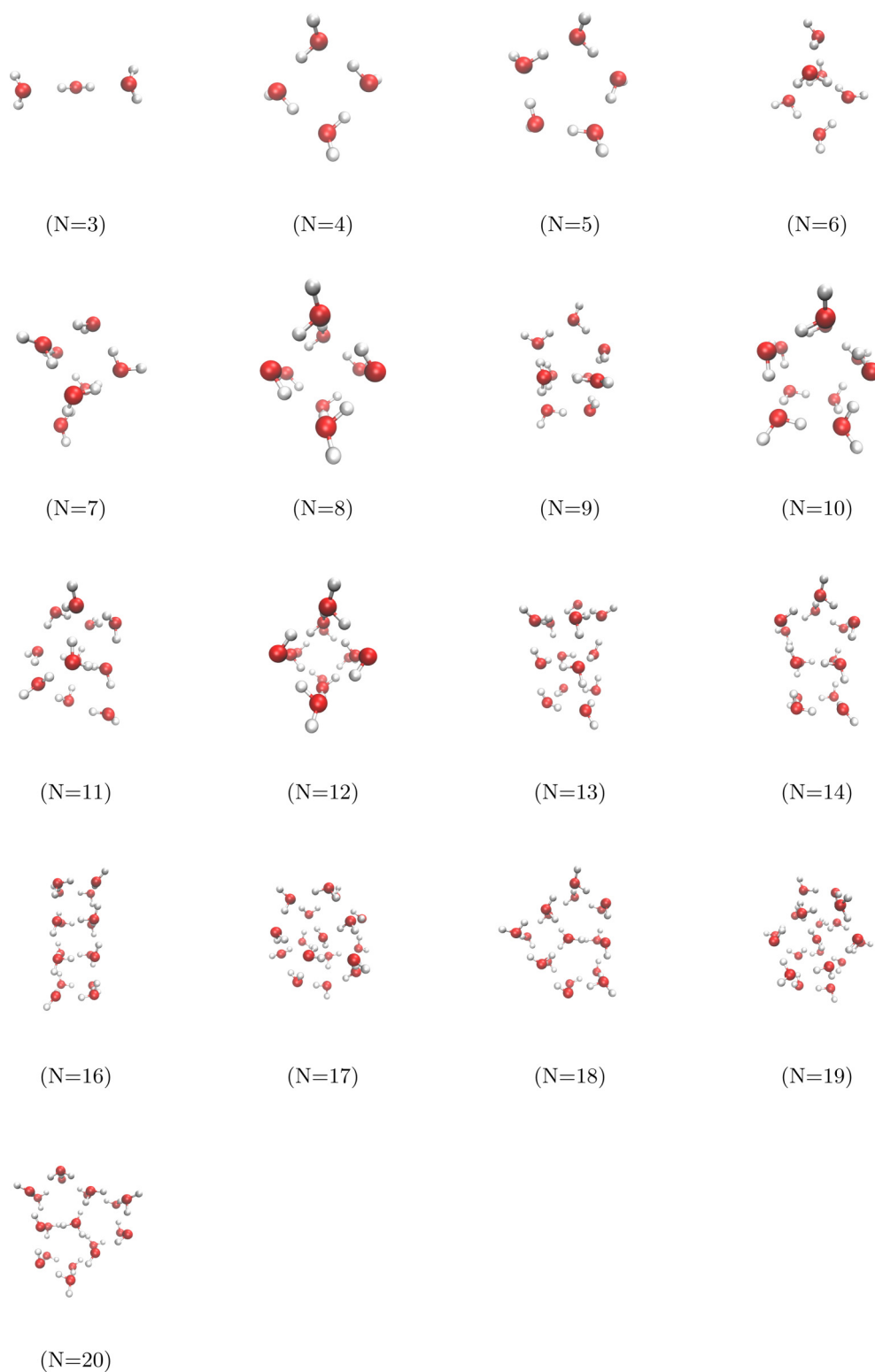


FIG. 8. Structures of the water clusters used in Sec. IV D. Red and white balls represent oxygen and hydrogen respectively. Geometries from [103]. For visualization, the visual molecular dynamics software package (VMD) was used [121].

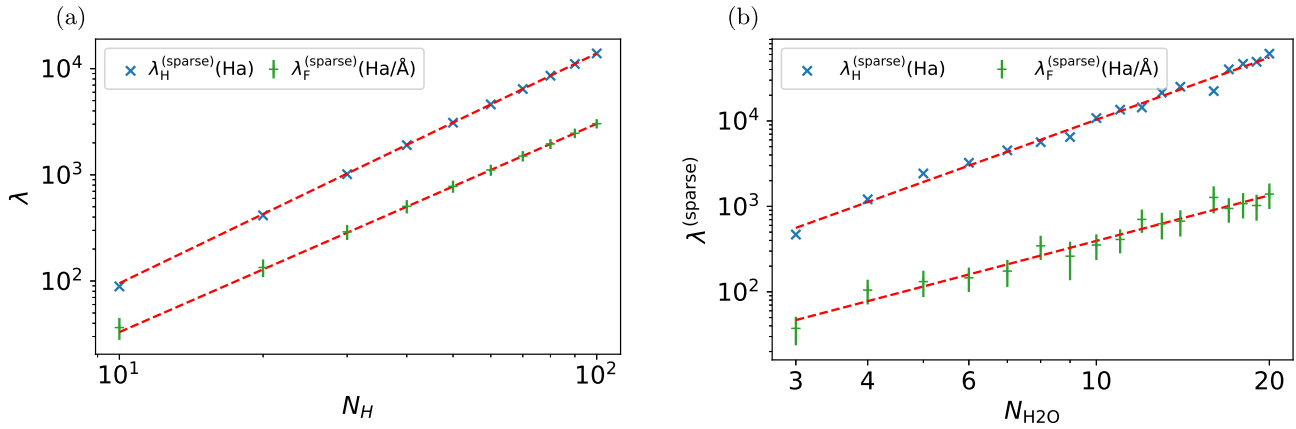


FIG. 9. Same plots as in Figs. 4 and 5(a) respectively, but using CMOs instead of LMOs. The fits suggest scalings of  $\lambda_H^{(\text{sparse})} = \mathcal{O}(N_H^{2.165 \pm 0.001})$  and  $\lambda_F^{(\text{sparse})} = \mathcal{O}(N_H^{1.963 \pm 0.001})$  for the hydrogen chain (a), and  $\lambda_H^{(\text{sparse})} = \mathcal{O}(N_{H_2O}^{2.421 \pm 0.065})$  and  $\lambda_F^{(\text{sparse})} = \mathcal{O}(N_{H_2O}^{1.772 \pm 0.086})$  for the water clusters (b).

$$\begin{aligned}
 \frac{1}{2^{2n+1}} \sum_{v_x, v_y, v_z = -(2^{n-1}-1)}^{2^{n-1}-1} \frac{|v_x|}{\|v\|^2} &\approx \frac{1}{2^{2n+1}} \int_{-2^{n-1}}^{2^{n-1}} dv_z \int_{-2^{n-1}}^{2^{n-1}} dv_y \int_{-2^{n-1}}^{2^{n-1}} dv_x \frac{|v_x|}{v_x^2 + v_y^2 + v_z^2} \\
 &= \frac{1}{2} \int_0^1 dz \int_0^1 dy \int_0^1 dx \frac{x}{x^2 + y^2 + z^2} \\
 &= \frac{1}{4} \int_0^1 dz \int_0^1 dy \log \left( 1 + \frac{1}{y^2 + z^2} \right) \\
 &= \frac{1}{4} \int_0^1 dz \left[ -2z \operatorname{arccot}(z) + 2\sqrt{1+z^2} \operatorname{arccot}\sqrt{1+z^2} + \log \left( 1 + \frac{1}{1+z^2} \right) \right] \\
 &= \frac{1}{4} \left[ -1 - \frac{\pi}{2} + 2\sqrt{2} \arctan \left( \frac{1}{\sqrt{2}} \right) + \log(3/2) \right] \\
 &\quad + \frac{1}{2} \int_0^1 dz \left[ \sqrt{1+z^2} \operatorname{arccot}\sqrt{1+z^2} \right]. \tag{D15}
 \end{aligned}$$

The integral can be computed to arbitrary precision numerically, giving

$$\int_0^1 dz \left[ \sqrt{1+z^2} \operatorname{arccot}\sqrt{1+z^2} \right] \approx 0.8197551759351555. \tag{D16}$$

That then gives the probability of success as approximately

$$\frac{1}{2^{2n+2}} \sum_{v_x, v_y, v_z = -(2^{n-1}-1)}^{2^{n-1}-1} \frac{|v_x|}{\|v\|^2} \approx 0.3037546589794463. \tag{D17}$$

To boost the success probability close to unity, we can use just one step of amplitude amplification because the probability is greater than 1/4. The dominant complexity is involved in the squares and multiplications in the inequality test, and the amplitude amplification only gives a multiplicative factor, for a total complexity of  $\mathcal{O}(\log N \log(N/\epsilon))$  Toffolis.

## APPENDIX E: NUMERICAL STUDIES OF THE FORCE OPERATOR IN MOLECULAR SYSTEMS

### 1. Details on numerical calculations of water clusters

When calculating the Hamiltonian and the force operators of the water clusters, shown in Fig. 8, in STO-3G, we recognize that the matricized two-body coefficients  $M_{(pq),(rs)}$  have degenerated eigenspaces. The resulting eigenvectors are thus only defined up to some gauge. While this is not a practical issue for applying any of the methods, repeated calculations will lead to slightly different lambdas. To avoid this ambiguity, we break the symmetry by shifting the positions of the atoms,  $x_i \rightarrow x_i + \mathcal{N}(0, \sigma)$  with  $\sigma = 10^{-6}$ .

For the largest water cluster we study here, with  $N_{H_2O} = 20$  water molecules, the spatial two-body integrals of the Hamiltonian and the force operators have a dimension of (140,140,140,140) and consequently the matrix  $M_{(pq),(rs)}$  is of dimension (19 600, 19 600). As a direct diagonalization of a matrix of such size is getting computationally expensive, we instead use Scipy's implementation of Lanczos algorithm to find the most important eigenvalues and eigenvectors. To find the error of this approximation, we calculate the expectation value of the Hamiltonian and the force operator with respect

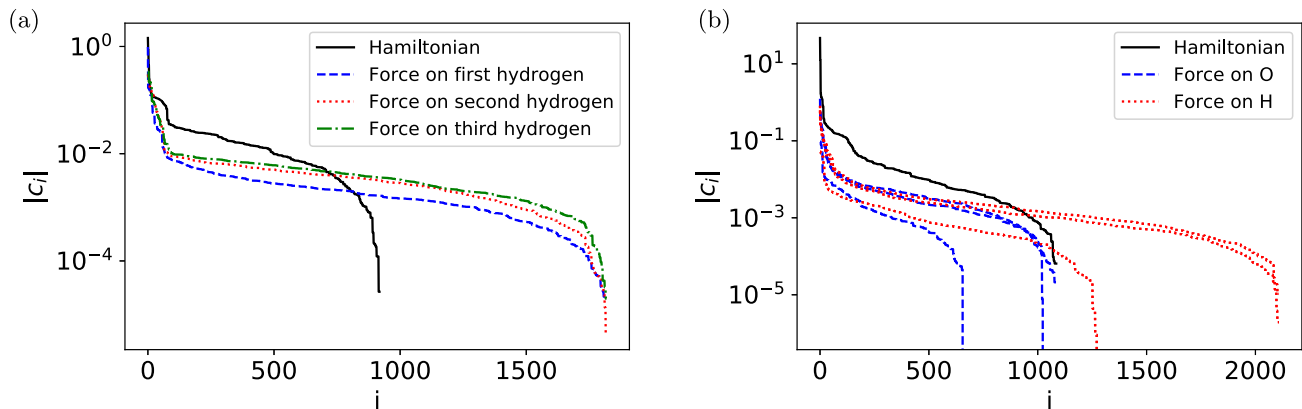


FIG. 10. Distribution of the absolute values of Pauli weights  $|c_i|$  of the Hamiltonian and force operators after a Jordan-Wigner transformation. For the calculations, canonical Hartree-Fock orbitals were used. (a) An one-dimensional chain of 6 hydrogen atoms with spacing of 0.74084 Å. Only forces in direction of the connecting line are shown. (b) For a single water molecule, with the geometry obtained from [122].

to the CISD wave function found under the reconstructed ERIs. In Fig. 7(a), we show the error on the CISD energy per atom  $\epsilon_E^{\text{CISD}} = (E_{\text{full}} - E_{\text{Lanczos}}^k)/(3N_{H_2O})$ . We find that keeping the  $4N$  most important eigenvalues and eigenvectors yields an error below 50  $\mu\text{Ha}$  per atom. In Fig. 7(b), we show the error on the expectation value of the force operator with respect to the CISD wave function. As the force is a vector of dimension  $3 \times 3 \times N_{H_2O}$ , we use  $\epsilon_F^{\text{CISD}} = \|\vec{F}_{\text{full}} - \vec{F}_{\text{Lanczos}}\|$  as error. As visible from the plots, keeping  $4N$  eigenvalues, as for the Hamiltonian, is a reasonable choice for the error on the force vector.

## 2. Structure of individual force operators

Any fermionic operator can be represented after a Jordan-Wigner transformation by an operator defined in qubit space,

$A = c_i P_i$ , where  $P_i$  is a  $n$ -qubit Pauli operator and  $c_i$  its coefficient. In Fig. 10, we show the distribution of the Pauli coefficients  $c_i$  after a Jordan-Wigner transformation of the Hamiltonian and the force operators for a hydrogen chain with 6 hydrogen and a water molecule. For both system, the operators were obtained from canonical Hartree-Fock orbitals in the STO-6G basis.

## 3. Additional calculations using canonical molecular orbitals

In Fig. 9, we provide additional results on  $\lambda^{(\text{sparse})}$  for the Hamiltonian and the force operators using CMOs instead of LMOs. For both systems, we find a worse scaling compared to the scaling obtained by using LMOs.

- [1] S. Lloyd, Universal quantum simulators, *Science* **273**, 1073 (1996).
- [2] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love, and M. Head-Gordon, Simulated quantum computation of molecular energies, *Science* **309**, 1704 (2005).
- [3] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya *et al.*, Quantum chemistry in the age of quantum computing, *Chem. Rev.* **119**, 10856 (2019).
- [4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [5] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong *et al.*, Quantum computational advantage via 60-qubit 24-cycle random circuit sampling, *Science Bulletin* **67**, 240 (2022).
- [6] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov *et al.*, Exponential suppression of bit or phase flip errors with repetitive error correction, *Nature (London)* **595**, 600 (2021).
- [7] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas, M. Newman, M. Li, K. R. Brown, M. Cetina, and C. Monroe, Fault-tolerant control of an error-corrected qubit, *Nature (London)* **598**, 281 (2021).
- [8] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown *et al.*, Realization of Real-Time Fault-Tolerant Quantum Error Correction, *Phys. Rev. X* **11**, 041058 (2021).
- [9] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
- [10] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G.-L. Chan, Low Depth Quantum Simulation of Electronic Structure, *Phys. Rev. X* **8**, 011044 (2018).
- [11] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush, and G. K.-L. Chan, Low rank representations for quantum simulation of electronic structure, *npj Quantum Inf.* **7**, 83 (2021).
- [12] D. W. Berry, C. Gidney, M. Motta, J. R. McClean, and R. Babbush, Qubitization of arbitrary basis quantum chemistry leveraging sparsity and low rank factorization, *Quantum* **3**, 208 (2019).

- [13] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe, and R. Babbush, Even more efficient quantum computations of chemistry through tensor hypercontraction, *PRX Quantum* **2**, 030305 (2021).
- [14] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, Elucidating reaction mechanisms on quantum computers, *Proc. Natl. Acad. Sci. USA* **114**, 7555 (2017).
- [15] V. von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler, and M. Troyer, Quantum computing enhanced computational catalysis, *Phys. Rev. Res.* **3**, 033055 (2021).
- [16] I. H. Kim, Y.-H. Liu, S. Pallister, W. Pol, S. Roberts, and E. Lee, Fault-tolerant resource estimate for quantum chemical simulations: Case study on Li-ion battery electrolyte molecules, *Phys. Rev. Res.* **4**, 023019 (2022).
- [17] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, *New J. Phys.* **18**, 023023 (2016).
- [18] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, *Nat. Commun.* **10**, 3007 (2019).
- [19] W. J. Huggins, J. McClean, N. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers, *npj Quantum Inf.* **7**, 23 (2021).
- [20] X. Bonet-Monroig, R. Babbush, and T. E. O'Brien, Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States, *Phys. Rev. X* **10**, 031064 (2020).
- [21] 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, *Phys. Rep.* **986**, 1 (2022).
- [22] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke *et al.*, Noisy intermediate-scale quantum algorithms, *Rev. Mod. Phys.* **94**, 015004 (2022).
- [23] 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, *Nat. Commun.* **5**, 4213 (2014).
- [24] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus *et al.*, Scalable Quantum Simulation of Molecular Energies, *Phys. Rev. X* **6**, 031007 (2016).
- [25] 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 (London)* **549**, 242 (2017).
- [26] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone *et al.*, Witnessing eigenstates for quantum simulation of Hamiltonian spectra, *Sci. Adv.* **4**, eaap9646 (2018).
- [27] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz, H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator, *Phys. Rev. X* **8**, 031022 (2018).
- [28] F. Arute, K. Arya, R. Babbush, D. Bacon, J. Bardin, R. Barends, S. Boixo, M. Broughton, B. Buckley, D. Buell *et al.*, Hartree-Fock on a superconducting qubit quantum computer, *Science* **369**, 1084 (2020).
- [29] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, *Nature (London)* **603**, 416 (2022).
- [30] H. P. Hratchian and H. B. Schlegel, Chapter 10 - Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces, in *Theory and Applications of Computational Chemistry*, edited by C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Elsevier, Amsterdam, 2005), pp. 195–249.
- [31] Y.-F. Li, *Ab initio* molecular dynamics in heterogeneous catalysis, in *Heterogeneous Catalysts* (John Wiley and Sons, Hoboken, NJ, 2021), Chap. 24, pp. 419–437.
- [32] M. Matsumoto, S. Saito, and I. Ohmine, Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing, *Nature (London)* **416**, 409 (2002).
- [33] O. M. H. Salo-Ahen, I. Alanko, R. Bhadane, A. M. J. J. Bonvin, R. V. Honorato, S. Hossain, A. H. Juffer, A. Kabedev, M. Lahtela-Kakkonen, A. S. Larsen *et al.*, Molecular dynamics simulations in drug discovery and pharmaceutical development, *Processes* **9**, 71 (2021).
- [34] Z. Cournia, B. Allen, and W. Sherman, Relative binding free energy calculations in drug discovery: Recent advances and practical considerations, *J. Chem. Inf. Model.* **57**, 2911 (2017).
- [35] F. D. Malone, R. M. Parrish, A. R. Welden, T. Fox, M. Degroote, E. Kyoseva, N. Moll, R. Santagati, and M. Streif, Towards the simulation of large scale protein–ligand interactions on nisq-era quantum computers, *Chem. Sci.* **13**, 3094 (2022).
- [36] J. J. M. Kirsopp, C. Di Paola, D. Z. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyder, D. Wolf, M. Strahm, and D. Muñoz Ramo, Quantum computational quantification of protein-ligand interactions, *Int. J. Quant. Chem.* **122**, e26975 (2022).
- [37] O. Fleetwood, J. Carlsson, and L. Delemotte, Identification of ligand-specific g protein-coupled receptor states and prediction of downstream efficacy via data-driven modeling, *eLife* **10**, e60715 (2021).
- [38] A. K. Nivedha, C. S. Tautermann, S. Bhattacharya, S. Lee, P. Casarosa, I. Kollak, T. Kiechle, and N. Vaidehi, Identifying functional hotspot residues for biased ligand design in G-protein-coupled receptors, *Mol. Pharmacol.* **93**, 288 (2018).
- [39] A. Nunes-Alves, D. B. Kokh, and R. C. Wade, Recent progress in molecular simulation methods for drug binding kinetics, *Curr. Opin. Struct. Biol.* **64**, 126 (2020).
- [40] I. Kassal and A. Aspuru-Guzik, Quantum algorithm for molecular properties and geometry optimization, *J. Chem. Phys.* **131**, 224102 (2009).
- [41] S. P. Jordan, Fast Quantum Algorithm for Numerical Gradient Estimation, *Phys. Rev. Lett.* **95**, 050501 (2005).
- [42] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo, and L. Visscher, Calculating energy derivatives for quantum chemistry on a quantum computer, *npj Quantum Inf.* **5**, 113 (2019).

- [43] K. Mitarai, Y. O. Nakagawa, and W. Mizukami, Theory of analytical energy derivatives for the variational quantum eigensolver, *Phys. Rev. Res.* **2**, 013129 (2020).
- [44] R. M. Parrish, E. G. Hohenstein, P. L. McMahon, and T. J. Martínez, Hybrid quantum/classical derivative theory: Analytical gradients and excited-state dynamics for the multistate contracted variational quantum eigensolver, [arXiv:1906.08728](https://arxiv.org/abs/1906.08728).
- [45] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, *Phys. Rev. A* **98**, 032309 (2018).
- [46] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, *Phys. Rev. A* **99**, 032331 (2019).
- [47] A. B. Magann, M. D. Grace, H. A. Rabitz, and M. Sarovar, Digital quantum simulation of molecular dynamics and control, *Phys. Rev. Res.* **3**, 023165 (2021).
- [48] D. A. Fedorov, M. J. Otten, S. K. Gray, and Y. Alexeev, *Ab initio* molecular dynamics on quantum computers, *J. Chem. Phys.* **154**, 164103 (2021).
- [49] I. O. Sokolov, P. K. Barkoutsos, L. Moeller, P. Suchsland, G. Mazzola, and I. Tavernelli, Microcanonical and finite-temperature *ab initio* molecular dynamics simulations on quantum computers, *Phys. Rev. Res.* **3**, 013125 (2021).
- [50] K. Omiya, Y. O. Nakagawa, S. Koh, W. Mizukami, Q. Gao, and T. Kobayashi, Analytical energy gradient for state-averaged orbital-optimized variational quantum eigensolvers and its application to a photochemical reaction, *J. Chem. Theory Comput.* **18**, 741 (2022).
- [51] S. Yalouz, E. Koridon, B. Senjean, B. Lasorne, F. Buda, and L. Visscher, Analytical nonadiabatic couplings and gradients within the state-averaged orbital-optimized variational quantum eigensolver, *J. Chem. Theory Comput.* **18**, 776 (2022).
- [52] R. M. Parrish, G.-L. R. Anselmetti, and C. Gogolin, Analytical ground- and excited-state gradients for molecular electronic structure theory from hybrid quantum/classical methods, [arXiv:2110.05040](https://arxiv.org/abs/2110.05040).
- [53] T.-C. Yen, V. Verteletskyi, and A. F. Izmaylov, Measuring all compatible operators in one series of a single-qubit measurements using unitary transformations, *J. Chem. Theory Comput.* **16**, 2400 (2020).
- [54] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* **16**, 1050 (2020).
- [55] E. Knill, G. Ortiz, and R. D. Somma, Optimal quantum measurements of expectation values of observables, *Phys. Rev. A* **75**, 012328 (2007).
- [56] W. J. Huggins, K. Wan, J. McClean, T. E. O'Brien, N. Wiebe, and R. Babbush, Nearly Optimal Quantum Algorithm for Estimating Multiple Expectation Values, *Phys. Rev. Lett.* **129**, 240501 (2022).
- [57] R. Babbush, D. W. Berry, J. R. McClean, and H. Neven, Quantum simulation of chemistry with sublinear scaling in basis size, *npj Quantum Inf.* **5**, 92 (2019).
- [58] Y. Su, D. W. Berry, N. Wiebe, N. Rubin, and R. Babbush, Fault-tolerant quantum simulations of chemistry in first quantization, *PRX Quantum* **2**, 040332 (2021).
- [59] T. U. Helgaker and J. Almlöf, A second-quantization approach to the analytical evaluation of response properties for perturbation-dependent basis sets, *Int. J. Quantum Chem.* **26**, 275 (1984).
- [60] A. Zhao, N. C. Rubin, and A. Miyake, Fermionic Partial Tomography via Classical Shadows, *Phys. Rev. Lett.* **127**, 110504 (2021).
- [61] L. Lin and Y. Tong, Near-optimal ground state preparation, *Quantum* **4**, 372 (2020).
- [62] Y. Yamaguchi and H. F. Schaefer III, Analytic derivative methods in molecular electronic structure theory: A new dimension to quantum chemistry and its applications to spectroscopy, in *Handbook of High-resolution Spectroscopy* (John Wiley & Sons, New Jersey, 2011).
- [63] T. Helgaker and P. Jørgensen, Calculation of geometrical derivatives in molecular electronic structure theory, in *Methods in Computational Molecular Physics*, edited by S. Wilson and G. H. F. Diercksen (Springer US, Boston, 1992), pp. 353–421.
- [64] H. Hellmann, *Einführung in die Quantenchemie*, Deuticke, Leipzig and Wien (1937).
- [65] H. Hellmann, *Kvantovaya Khimiya [Quantum Chemistry (in Russian)]* (ONTI, Moscow and Leningrad, 1937).
- [66] R. P. Feynman, Forces in molecules, *Phys. Rev.* **56**, 340 (1939).
- [67] R. Babbush, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, A. Paler, A. Fowler, and H. Neven, Encoding Electronic Spectra in Quantum Circuits with Linear T Complexity, *Phys. Rev. X* **8**, 041015 (2018).
- [68] I. D. Kivlichan, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, W. Sun, Z. Jiang, N. Rubin, A. Fowler, A. Aspuru-Guzik, H. Neven, and R. Babbush, Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via trotterization, *Quantum* **4**, 296 (2020).
- [69] G. H. Low and N. Wiebe, Hamiltonian simulation in the interaction picture, [arXiv:1805.00675](https://arxiv.org/abs/1805.00675).
- [70] Y. Su, H.-Y. Huang, and E. T. Campbell, Nearly tight Trotterization of interacting electrons, *Quantum* **5**, 495 (2021).
- [71] F. Verstraete, J. I. Cirac, and J. I. Latorre, Quantum circuits for strongly correlated quantum systems, *Phys. Rev. A* **79**, 032316 (2009).
- [72] A. J. Ferris, Fourier Transform for Fermionic Systems and the Spectral Tensor Network, *Phys. Rev. Lett.* **113**, 010401 (2014).
- [73] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji *et al.*, Gaussian 16 Revision C.01 (2016), Gaussian Inc. Wallingford CT, <https://gaussian.com/citation/>.
- [74] L. Gross, F. Mohn, N. Moll, B. Schuler, A. Criado, E. Guitián, D. Peña, A. Gourdon, and G. Meyer, Bond-order discrimination by atomic force microscopy, *Science* **337**, 1326 (2012).
- [75] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* **79**, 926 (1983).
- [76] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus *et al.*, The atomic simulation environment—A python library for working with atoms, *J. Phys.: Condens. Matter* **29**, 273002 (2017).
- [77] Y. Luo, A. Zen, and S. Sorella, *Ab initio* molecular dynamics with noisy forces: Validating the quantum Monte Carlo ap-

- proach with benchmark calculations of molecular vibrational properties, *J. Chem. Phys.* **141**, 194112 (2014).
- [78] O. Marsalek and T. E. Markland, Quantum dynamics and spectroscopy of *ab initio* liquid water: The interplay of nuclear and electronic quantum effects, *J. Phys. Chem. Lett.* **8**, 1545 (2017).
- [79] S. A. Hollingsworth and R. O. Dror, Molecular dynamics simulation for all, *Neuron* **99**, 1129 (2018).
- [80] R. Lazim, D. Suh, and S. Choi, Advances in molecular dynamics simulations and enhanced sampling methods for the study of protein systems, *Int. J. Mol. Sci.* **21**, 6339 (2020).
- [81] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution, *Nat. Phys.* **16**, 205 (2020).
- [82] S. Polla, Y. Herasymenko, and T. E. O'Brien, Quantum digital cooling, *Phys. Rev. A* **104**, 012414 (2021).
- [83] J. Cotler and F. Wilczek, Quantum Overlapping Tomography, *Phys. Rev. Lett.* **124**, 100401 (2020).
- [84] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, *J. Chem. Phys.* **152**, 124114 (2020).
- [85] T.-C. Yen and A. F. Izmaylov, Cartan subalgebra approach to efficient measurements of quantum observables, *PRX Quantum* **2**, 040320 (2021).
- [86] H.-Y. Huang, R. Kueng, and J. Preskill, Efficient Estimation of Pauli Observables by Derandomization, *Phys. Rev. Lett.* **127**, 030503 (2021).
- [87] T.-C. Yen, A. Ganeshra, and A. F. Izmaylov, Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates, [arXiv:2201.01471](https://arxiv.org/abs/2201.01471).
- [88] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley, Efficient quantum measurement of Pauli operators in the presence of finite sampling error, *Quantum* **5**, 385 (2021).
- [89] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K.-L. Chan, and R. Babbush, Quantum Simulation of Electronic Structure with Linear Depth and Connectivity, *Phys. Rev. Lett.* **120**, 110501 (2018).
- [90] N. C. Rubin, J. Lee, and R. Babbush, Compressing many-body fermion operators under unitary constraints, *J. Chem. Theory Comput.* **18**, 1480 (2022).
- [91] N. C. Rubin, R. Babbush, and J. McClean, Application of fermionic marginal constraints to hybrid quantum algorithms, *New J. Phys.* **20**, 053020 (2018).
- [92] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, *Phys. Rev. A* **92**, 042303 (2015).
- [93] S. Bravyi, D. Gosset, and R. Movassagh, Classical algorithms for quantum mean values, *Nat. Phys.* **17**, 337 (2021).
- [94] A. M. Childs and N. Wiebe, Hamiltonian simulation using linear combinations of unitary operations, *Quantum Inf. Comput.* **12**, 901 (2012).
- [95] G. H. Low and I. L. Chuang, Hamiltonian simulation by qubitization, *Quantum* **3**, 163 (2019).
- [96] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Graduate Studies in Mathematics* (American Mathematical Society Providence, RI, 2002), Vol. 47, p. 257.
- [97] C. Gidney, Halving the cost of quantum addition, *Quantum* **2**, 74 (2018).
- [98] E. Koridon, S. Yalouz, B. Senjean, F. Buda, T. E. O'Brien, and L. Visscher, Orbital transformations to reduce the 1-norm of the electronic structure hamiltonian for quantum computing applications, *Phys. Rev. Res.* **3**, 033127 (2021).
- [99] F. N. Keutsch and R. J. Saykally, Water clusters: Untangling the mysteries of the liquid, one molecule at a time, *Proc. Natl. Acad. Sci. USA* **98**, 10533 (2001).
- [100] D. Manna, M. K. Kesharwani, N. Sylvetsky, and J. M. L. Martin, Conventional and explicitly correlated *ab initio* benchmark study on water clusters: Revision of the BEDGB and WATER27 data sets, *J. Chem. Theory Comput.* **13**, 3136 (2017).
- [101] G. Vladilo and A. Hassanali, Hydrogen bonds and life in the universe, *Life* **8**, 1 (2018).
- [102] M. Huš and T. Urbic, Strength of hydrogen bonds of water depends on local environment, *J. Chem. Phys.* **136**, 144305 (2012).
- [103] A. Rakshit, P. Bandyopadhyay, J. P. Heindel, and S. S. Xantheas, Atlas of putative minima and low-lying energy networks of water clusters  $n = 3-25$ , *J. Chem. Phys.* **151**, 214307 (2019).
- [104] A. Gilyén, S. Arunachalam, and N. Wiebe, Optimizing quantum optimization algorithms via faster quantum gradient computation, in *Proceedings of the 2019 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, edited by Timothy M. Chan (Philadelphia, PA, 2017), <https://epubs.siam.org/doi/book/10.1137/1.9781611975482>.
- [105] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman, How to perform the most accurate possible phase measurements, *Phys. Rev. A* **80**, 052114 (2009).
- [106] B. L. Higgins, D. W. Berry, S. D. Bartlett, M. W. Mitchell, H. M. Wiseman, and G. J. Pryde., Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements., *New J. Phys.* **11**, 073023 (2009).
- [107] Y. Ge, J. Tura, and J. I. Cirac, Faster ground state preparation and high-precision ground energy estimation with fewer qubits, *J. Math. Phys.* **60**, 022202 (2019).
- [108] R. Babbush, B. Dominic, and H. Neven, Quantum simulation of the Sachdev-Ye-Kitaev model by asymmetric qubitization, *Phys. Rev. A* **99**, 040301(R) (2019).
- [109] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: beating the standard quantum limit, *Science* **306**, 1330 (2004).
- [110] M. Szegedy, Quantum speed-up of Markov chain based algorithms, in *45th Annual IEEE Symposium on Foundations of Computer Science (IEEE, 2004)*, pp. 32–41.
- [111] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, Quantum Singular Value Transformation and beyond: Exponential Improvements for Quantum Matrix Arithmetics, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019* (Association for Computing Machinery, New York, 2019), pp. 193–204.
- [112] T. J. Yoder, G. H. Low, and I. L. Chuang, Fixed-Point Quantum Search with an Optimal Number of Queries, *Phys. Rev. Lett.* **113**, 210501 (2014).



- [113] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, Quantum amplitude amplification and estimation, *Contemp. Math.* **305**, 53 (2002).
- [114] N. M. Tubman, C. Mejuto-Zaera, J. M. Epstein, D. Hait, D. S. Levine, W. Huggins, Z. Jiang, J. R. McClean, R. Babbush, M. Head-Gordon, and K. B. Whaley, Postponing the orthogonality catastrophe: efficient state preparation for electronic structure simulations on quantum devices, [arXiv:1809.05523](https://arxiv.org/abs/1809.05523).
- [115] J. Romero, R. Babbush, J. McClean, C. Hempel, P. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, *Quantum Sci. Technol.* **4**, 014008 (2017).
- [116] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, Quantum chemical accuracy from density functional approximations via machine learning, *Nat. Commun.* **11**, 5223 (2020).
- [117] D. R. Maurice, Single electron theories of excited states, Ph.D. thesis, University of California, Berkeley, 1998.
- [118] C. Edmiston and K. Ruedenberg, Localized atomic and molecular orbitals, *Rev. Mod. Phys.* **35**, 457 (1963).
- [119] J. Pipek and P. G. Mezey, A fast intrinsic localization procedure applicable for *ab initio* and semiempirical linear combination of atomic orbital wave functions, *J. Chem. Phys.* **90**, 4916 (1989).
- [120] J. Foster and S. Boys, Canonical configurational interaction procedure, *Rev. Mod. Phys.* **32**, 300 (1960).
- [121] W. Humphrey, A. Dalke, and K. Schulten, VMD – Visual molecular dynamics, *J. Mol. Graphics* **14**, 33 (1996).
- [122] R. D. Johnson III, NIST computational chemistry comparison and benchmark database, nist standard reference database number 101, Release 21, August 2020, <http://cccbdb.nist.gov/> (2020).