PAPER • OPEN ACCESS

A differentiable quantum phase estimation algorithm

View the article online for updates and enhancements.

Quantum Science and Technology

PAPER

A differentiable quantum phase estimation algorithm

Davide Castaldo[1,2,*] ![ORCID], **Soran Jahangiri**[2] ![ORCID], **Agostino Migliore**[1] ![ORCID], **Juan Miguel Arrazola**[2] ![ORCID] and **Stefano Corni**[1,3,4] ![ORCID]

1. Dipartimento di Scienze Chimiche, Università degli Studi di Padova, I-35131 Padova, Italy
2. Xanadu, Toronto, ON M5G2C8, Canada
3. Padua Quantum Technologies Research Center, Università degli Studi di Padova, Padova, Itlay
4. Istituto Nanoscienze—CNR, via Campi 213/A, 41125 Modena, Italy
* Author to whom any correspondence should be addressed.

E-mail: davide.castaldo@studenti.unipd.it

Abstract

The simulation of electronic properties is a pivotal issue in modern electronic structure theory, driving significant efforts over the past decades to develop protocols for computing energy derivatives. In this work, we address this problem by developing a strategy to integrate the quantum phase estimation algorithm within a fully differentiable framework. This is accomplished by devising a smooth estimator able to tackle arbitrary initial states. We provide analytical expressions to characterize the statistics and algorithmic cost of this estimator. Furthermore, we provide numerical evidence that the estimation accuracy is retained when an arbitrary state is considered and that it exceeds the one of standard majority rule. We explicitly use this procedure to estimate chemically relevant quantities, demonstrating our approach through ground-state and triplet excited state geometry optimization with simulations involving up to 19 qubits. This work paves the way for new quantum algorithms that combine interference methods and quantum differentiable programming.

1. Introduction

Advances in big data information processing [1] and sophisticated multiscale modeling [2, 3] required by cutting-edge research determine an ever-increasing demand for large-scale computations [4]. This urgent requirement is already driving a major shift in our approach to computation. As the on-chip power dissipation of available semiconductor technologies limits the further development of standard processing architectures, we are likely to see a slowdown of Moore's law. Along the same lines, the diversification of computing paradigms, both from an architectural and algorithmic point of view, could be beneficial for a more sustainable growth in the use of energy and materials [5].

To some extent, such a transition has already taken place in the context of electronic structure theory. In fact, very active research lines are improving the performance and costs of simulating molecular Hamiltonians by developing implementations on graphical processing units [6–8] and, more recently, tensor processing units [9]. These contributions concern improvements in accuracy and time-to-solution to obtain not only molecular energies but also other molecular properties which, from the chemist's point of view, can be even more important when it comes to simulating real-life experiments. It is well known that many molecular properties can be expressed in terms of energy derivatives [10] and can be used for a variety of applications ranging from electronic or magnetic responses to the determination of minimum energy pathways, to the understanding of reaction mechanisms and drug-protein interactions for pharmaceutical purposes and materials discovery [11, 12]. For these reasons, over the years, much effort has been devoted to developing efficient methods to work out analytical expressions for gradients and higher order derivatives in many fields of quantum chemistry [13–15].

In the field of quantum computing, the advancement of these methods in a fault-tolerant setting is still lagging behind. Past works aimed at reducing the cost of quantum simulations for ground and excited state

energies [16–19]. Only recently works aimed at extracting physically relevant quantities other than single point energies started to appear [20], but very little has been done to study the occurrence of a parametric dependence within the phase estimation circuit.

Here we fill this gap with the development of the first example of an automatically differentiable quantum phase estimation (QPE) algorithm geared to compute energy derivatives for molecular Hamiltonians. Therefore, by merging interference methods and differentiable programming, our contribution unlocks, at once, the quantum simulation of a plethora of molecular properties such as molecular structures or electric and magnetic dipoles.

Automatic differentiation (AD), a clever combination of the chain rule and dynamic programming, is the workhorse of deep learning techniques [21] and in recent years it has also proven to be very valuable when applied to quantum chemistry [22–24] and, more generally, to scientific computing [25–28]. With the advent of noisy intermediate scale quantum devices, variational algorithms [29], in which the circuit is trained to optimize the parameters that define the action of gates tasked with a specific goal, have spurred the development of extensions of AD tailored for quantum computing. Much progress has been made to build AD techniques in quantum computing that take into account programs with [30] or without [31, 32] control flow, leading to the implementation of several differentiable quantum frameworks [33–35].

Here, we focus on the differentiation through the QPE circuit as described by Nielsen and Chuang's textbook [36], developing a technique for post-processing the measurements that allows for integration into automatically differentiable software. This result is relevant *per se*, as it may provide insights into how to reduce the cost of QPE, which, although asymptotically efficient, is still impractical for demonstrating clear and scalable quantum advantage [37–40].

This study is organized as follows. Section 2 is devoted to summarize the QPE algorithm and sketch the proposed differentiable quantum pipeline. In section 3 we discuss the main contribution of this work: building on a previous study of Cruz *et al* [41], we develop a smooth estimator for the QPE for use with arbitrary input states of the system register. This estimator allows for seamless integration of the QPE algorithm within a fully differentiable framework. We discuss its properties and derive error bounds to estimate how the smoothing procedure affects the computational cost of eigenvalue and energy gradient evaluation (see section 4). Section 4.2 provides a summary of the main options devised so far to compute energy derivatives in a Fault-Tolerant setting with a focus on the results of a recent study of O'Brien *et al* [42]. In section 5 we apply our method to two example systems, H_3^+ and CH_2O, performing geometry optimization of their ground singlet and first triplet states. The conclusions include some remarks on potential roadblocks related to the computational cost of the differentiable routine and possible ways to circumvent them.

2. Preliminaries

We begin with a brief review of the fundamental components of the QPE algorithm in its standard formulation [43]. We will not go into the details of the different simulation techniques used within the QPE circuit as they are beyond the scope of this study. Additionally, the second crucial element of this work involves quantum differentiable programming. Accordingly, we will highlight the main features needed for our development and elucidate why a smooth eigenvalue estimator is crucial in this particular context.
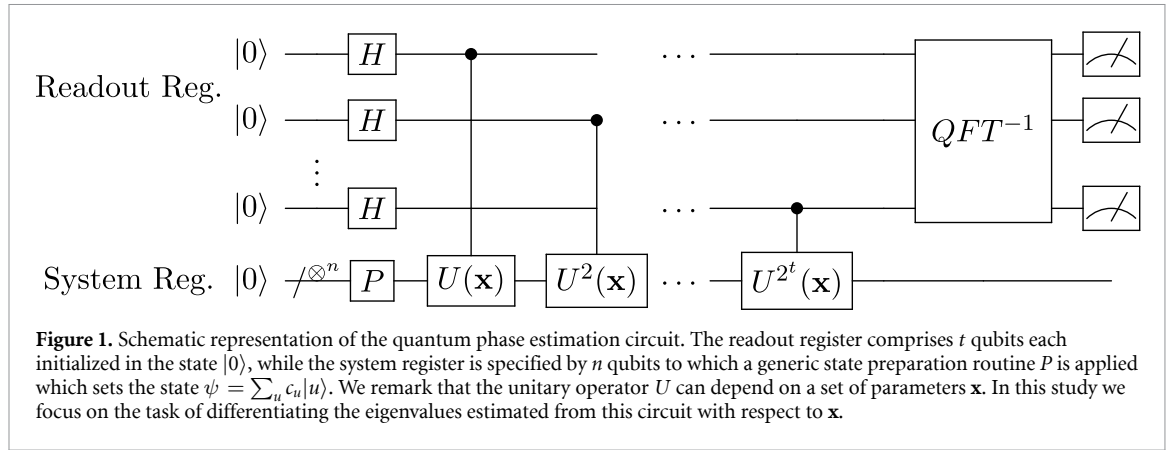
2.1. QPE

The QPE algorithm [43] aims to find (some of) the eigenvalues of an arbitrary unitary transformation U with precision ϵ. In this work we will focus on the problem of Hamiltonian simulation, i.e. for us the eigenvalues of U will be directly related to the energies $\{E_i \,|\, i = 1, 2, \ldots R\}$ of a quantum system encoded in a qubit register (called as system register) of n qubits. This is done by choosing U such that it shares the same eigenstates $\{|i\rangle\}$ with the Hamiltonian H_{sys} of the system and that the eigenvalues of U and H_{sys} are related by an invertible function $\phi_i = f(E_i)$ as shown in [36, 43].

The QPE circuit (figure 1) is implemented using two qubit registers: (i) the system register of n qubits encoding the wavefunction of the simulated system and (ii) the readout register of t qubits measured at the end of the circuit execution to estimate phases ϕ_i.

We can notice from the circuit that the algorithm consists of four main steps. First, the system register is prepared in state $|\psi\rangle = \sum_u c_u |u\rangle$ and the readout register in state $H^{\otimes t}|0\rangle = \frac{1}{2^{\frac{t}{2}}} \bigotimes_{k=1}^{t}(|0_k\rangle + |1_k\rangle)$. After this stage the QPU is in the state

$$|\Psi\rangle = \frac{1}{2^{\frac{t}{2}}} \sum_u c_u \bigotimes_{k=1}^{t} (|0_k\rangle + |1_k\rangle) |u\rangle. \tag{1}$$

Figure 1. Schematic representation of the quantum phase estimation circuit. The readout register comprises t qubits each initialized in the state $|0\rangle$, while the system register is specified by n qubits to which a generic state preparation routine P is applied which sets the state $\psi = \sum_u c_u |u\rangle$. We remark that the unitary operator U can depend on a set of parameters $\mathbf{x}$. In this study we focus on the task of differentiating the eigenvalues estimated from this circuit with respect to $\mathbf{x}$.

We recall that here the label u runs over the eigenstates of the Hamiltonian and t is the number of qubits in the readout register.

Then the unknown phases are encoded in the quantum circuit by performing a sequence of controlled evolutions of the system register conditioned on the states of the readout register. This brings the QPU into the state

$$|\Psi\rangle = \frac{1}{2^{\frac{t}{2}}} \sum_u c_u \bigotimes_{k=1}^{t} \left(|0_k\rangle + e^{i 2\pi \, 2^{k-1} \phi_u} |1_k\rangle \right) |u\rangle. \tag{2}$$

This state is achieved because the unitary evolution of the system register gives $U|\psi\rangle = \sum_u c_u e^{i2\pi\phi_u}|u\rangle$.

After all phases are encoded in the readout register, their values are transferred from the corresponding amplitudes to the computational basis spanned by the readout register by applying an inverse quantum Fourier transform [44]. This step is crucial for the whole algorithm. It is the step where the digitalization of the encoded phase occurs. We can think of the computational basis spanned by the readout register as a grid of 2^t evenly spaced points that we can use to represent discretely the Fourier transform of the amplitudes encoded in the state of equation (2). Now the state of the quantum computer (QC) reads:

$$|\Psi\rangle = \sum_u c_u |u\rangle \sum_{b_{1,u}=0}^{1} \sum_{b_{2,u}=0}^{1} \cdots \sum_{b_{t,u}=0}^{1} \left(\sum_{k}^{2^t} \frac{e^{i2\pi(\phi_u - 0.\phi_u)(k-1)}}{2^t} \right) |b_{1,u}\rangle |b_{2,u}\rangle \ldots |b_{t,u}\rangle \tag{3}$$

where the binary fraction $0.\phi_u = 0.b_{1,u}b_{2,u}\ldots b_{t,u}$ is a shorthand for $0.b_{1,u}b_{2,u}\ldots b_{t,u} = \sum_k^t \frac{b_{k,u}}{2^k}$. In the following, with $0.\phi_u$ denotes both this binary fraction and the corresponding bitstring $b_{1,u}b_{2,u}\ldots b_{t,u}$.

Finally, measuring the state of equation (3) expanded on the computational basis, we obtain the parent distribution $P(0.\phi) = \sum_u P_{\phi_u}(0.\phi)$, where each P_{ϕ_u} is the probability of measuring the bitstring $0.\phi$ produced by the eigenphase ϕ_u:

$$P_{\phi_u}(0.\phi) = \begin{cases} 1 & \Delta\phi = \phi_u - 0.\phi = 0 \\ \frac{g(0.\phi)}{2^{2t}} = \frac{1}{2^{2t}} \frac{\sin^2\left(2^t \pi \, (\phi_u - 0.\phi)\right)}{\sin^2(\pi(\phi_u - 0.\phi))} & \Delta\phi = \phi_u - 0.\phi \neq 0. \end{cases} \tag{4}$$

We will examine in detail the shape of the probability distribution in section 3, as this is the starting point to generalize the circular estimator developed in [41]. At this moment we limit ourselves to highlighting that the result of this type of functional form is to have a superposition of peaks that is narrower as the grid with which we operate the discretization is (e.g. the more readout qubits there are) becomes denser. Each of these peaks corresponds to an eigenstate $|u\rangle$ with non-zero overlap with the initial state $|\psi\rangle$ and is centered on the string that best approximates phase ϕ_u. As a consequence, the usual strategy for estimating the eigenvalue of the unitary U associated to a given eigenstate $|u\rangle$ is to prepare a state with high overlap with $|u\rangle$ and estimate the corresponding phase ϕ_u as $\phi_u \approx \text{argmax}(P(0.\phi)) = 0.\phi_u$. This is often referred to as majority rule estimate. Its accuracy is upper bounded by the number of discretization points used to represent $P(0.\phi)$; in particular the error associated with the estimate is $|\Delta\phi| = |\phi_u - 0.\phi_u| \leqslant \frac{1}{2^{t+1}}$.

We now move on to discuss quantum gradients and differentiability as they provide us with motivation to go beyond the majority rule estimator in section 3.
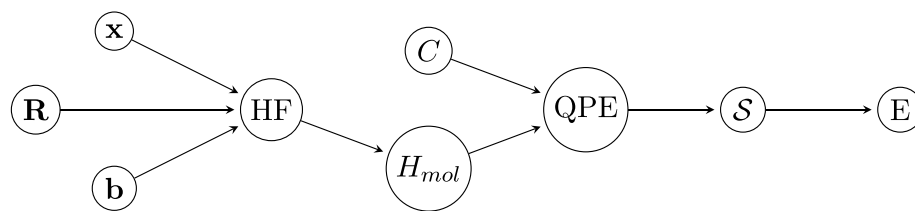
Figure 2. Coarse representation of the hybrid computational graph traversed to compute the molecular energy (and its derivatives) with the QPE algorithm. Each node represents deeply involved functions and, in practice, is written much more accurately. For example, we can consider the quantum circuit of figure 1 as a specific representation of the QPE node. The **x**, **R** and **b** refer to the inputs required by the molecular calculation: external fields, atomic positions and basis set coefficients, respectively. Subsequently the HF node computes a mean-field solution of the electronic structure problem necessary to obtain a set of molecular orbitals. Using this one-electron basis we perform the mapping of the molecular Hamiltonian onto the quantum computer, obtaining H_{mol}. Note how, in this representation, the initial state C fed into the QPE circuit is explicitly an independent (differentiable) variable. For an interpretation of this AD pipeline in terms of electronic structure theory quantities, please refer to appendix A. This work is primarily devoted to developing a postprocessing functional $\mathcal{S}$ so that it provides a smooth estimator of the molecular energy E.

2.2. Differentiating through the QPE circuit

Differentiable programming is a programming paradigm that leverages the chain rule to obtain the derivative of a computer program with respect to an input parameter. At the basis of this computational paradigm is AD, a method for calculating exact numerical derivatives which underlies many machine learning techniques such as the renowned backpropagation algorithm [45]. The basic idea behind AD is that the computer, regardless of the complexity of the function being computed, executes very simple operations that can all be tracked down and stored in memory. Such a sequence of operations is usually called evaluation trace and can be represented on a computational graph (figure 2).

The example in the figure illustrates a coarse-grained representation of the computational graph to calculate the energies of molecular electronic states. In section 5, we present results of energy derivation with respect to atomic positions (**R**), as is required for the optimization of molecular geometries. Here we explicitly set apart the input variables to convey the message that other possible choices lead to natural extensions of the present work. For instance, derivatives with respect to parameters on which the basis set depends (**b**) or external perturbations (**x**) are potential avenues to explore. In particular, the latter can enable the calculation of molecular properties relevant to spectroscopy.

A differentiable program is able to compute the derivative of the evaluation trace by traversing the computational graph through all its paths and gathering the partial derivative associated with each node. Specifically, during this operation, referred to as Jacobian accumulation, the program calculates the derivative of the function using the chain rule, which is composed of the derivatives of all the nodes in the computational graph (i.e. the fundamental operations executed by the computer to calculate a function f). It is crucial to underline for our purposes that, to maintain the differentiability of the evaluation trace (and therefore the viability of the algorithm), all individual steps must themselves be differentiable.

We refer to quantum differentiable programming when computational graphs include expressions executed on both classical and quantum hardware. It is important to highlight the difference between a classical AD of a software that aims to simulate quantum systems and a differentiable quantum program. As we will discuss later in more detail, the latter implies the development of techniques to differentiate expectation values. Here, we develop a fully differentiable QPE algorithm and apply it to quantum chemistry, which is one of the most promising fields of application for this algorithm. From this standpoint, the PennyLane software framework represents a natural choice, as it allows for end-to-end differentiability thanks to the differentiable Hartree–Fock solver integrated within its library [46] and allows one to include quantum functions in the computational graph.

In general, quantum functions accept input instructions for the quantum circuit and output results of its execution (i.e. measurements). For our particular purpose we can explicitly write the QPE parent distribution as the result of a parameterized expectation value:

$$P(0.\phi, \mathbf{x}) = \langle 0^{\otimes^t} | \langle \psi | \text{QFT}^\dagger \mathcal{U}_t^\dagger(\mathbf{x}) \, \text{QFT} | 0.\phi \rangle \langle 0.\phi | \text{QFT}^\dagger \mathcal{U}_t(\mathbf{x}) \, \text{QFT} | \psi \rangle | 0^{\otimes^t} \rangle \tag{5}$$

where $\mathcal{U}_t$ is the sequence of controlled evolutions used to encode the phase on the readout register $\mathcal{U}_t(\mathbf{x}) = \prod_k^t (|0_k\rangle\langle 0_k| \otimes \mathbf{1}^{\otimes^n} + |1_k\rangle\langle 1_k| \otimes U^{2^k}(\mathbf{x}))$, and the initial layer of Hadamard gates has been identified as a QFT on the readout register in the product state $|0\rangle^{\otimes^t}$. Further, we used the notation introduced in equation (3) to denote the projector on the bitstring state of the readout register corresponding to the binary fraction $0.\phi$.

To perform Jacobian accumulation of hybrid computational graphs it is necessary to develop techniques to compute gradients of quantum functions. Extensive literature focuses on calculating the derivatives of expectation values of parameterized quantum circuits [47–49]. The most relevant for our purposes are the works of [50–54], as they provide general parameter-shift rules to compute derivatives of gates like $U(\theta) = e^{i\theta H}$ with arbitrary H. In section 4 we will make use of their results to evaluate the number of calls needed to compute gradients by differentiating through a QPE circuit. We remark that we will make use of these techniques as we are proposing a pipeline to be executed on a quantum hardware.

Finally, now that we have discussed the fundamentals of differentiable programming, we can note that if we want to build a program that estimates the phase out of the parent distribution induced by a QPE circuit, the classical postprocessing that evaluates an estimator for the phase out of $P(0.\phi)$ must also be differentiable. By definition, the majority rule estimator discussed in the previous section does not satisfy this requirement. In the next section we therefore develop a smooth estimator for QPE. Our results ensure that the gradient can be calculated accurately. Should this pipeline, including a differentiable QPE, be adopted in contexts such as machine learning, where the number of parameters is very high, a possible option could be to use a straight-through estimator [55]. In this case, the gradient of the $\mathcal{S}$ node is replaced with the identity during the Jacobian accumulation. An analysis of the physical meaning of our estimator (see appendix A) when applied to a molecular Hamiltonian shows that this approach is not feasible in the field of quantum chemistry.

As anticipated in the introduction, we propose an extension of the work of Cruz *et al* [41], producing an estimator we refer to as the generalized circular estimator (GCE).

3. GCE

In this section we present a general smooth estimator for the QPE algorithm, characterize its statistical properties and discuss its application within a differentiable programming framework meant for evaluating derivatives of the molecular Hamiltonian.

Our idea is to combine the Gumbel-Softmax trick [56], used in machine learning to efficiently differentiate across categorical distributions, and the idea of Cruz *et al* [41] of estimating the phase from a circular average [57] of the parent distribution (equation (4)). We begin by recalling the results of [41] which are valid when the input state corresponds to an eigenstate $|u\rangle$ of the unitary U.

The authors propose to use the mean phase direction μ of the first trigonometric moment of the final QPE measurement distribution as an estimator of the energy. In particular, the first trigonometric moment θ reads

$$\theta = \sum_{0.\phi} P_{\phi_u}(0.\phi) \, e^{i2\pi \, 0.\phi} = |\theta| e^{i2\pi \mu}. \tag{6}$$

We recall that P_{ϕ_u} is defined by equation (4).

As pointed out in [41], a nice way to grasp the meaning of θ is to consider it equivalent to a vector sum in the complex plane, where each vector has a direction specified by $e^{i2\pi \, 0.\phi}$ and a modulus given by $P_{\phi_u}(0.\phi)$. This representation explains another important feature of θ that also emerges from our generalization, i.e. its accuracy exceeds that of the majority rule estimator. This stems from the fact that each string over- or under-represents the true phase ϕ_u because of the discretization in binary fractions, and therefore averaging over many strings weighted by their probabilities allows one to get a more accurate estimate due to error cancellation effects.

Subsequently, the mean phase direction μ of θ is given by:

$$\mu = \arg(\theta). \tag{7}$$

Plugging equation (4) into the definition of θ, the authors get an explicit expression for μ, θ and $|\theta|$

$$\theta = \frac{2^t - 1}{2^t} e^{i2\pi \phi_u} + \frac{1}{2^t} e^{-i\left(2^t - 1\right)2\pi \phi_u} \tag{8}$$

$$\mu = \frac{1}{2\pi} \arctan\left(\frac{(2^t - 1)\sin(2\pi \phi_u) - \sin\left((2^t - 1)2\pi \phi_u\right)}{(2^t - 1)\cos(2\pi \phi_u) + \cos\left((2^t - 1)2\pi \phi_u\right)} \right) \tag{9}$$

$$|\theta| = \sqrt{4^{-t}\left(4^t - 2^{t+1} + 2 + 2\left(2^t - 1\right)\cos\left(2^{t+1}\pi \phi\right)\right)}. \tag{10}$$

Both from equations (8) and (9) we can see that for large values of t the estimator approaches the true value of ϕ_u. Further, looking at equation (10) we can see that in the same regime (i.e. large t) the modulus of the first trigonometric moment approaches $|\theta| = 1$. Furthermore, this value is approached with exponentially

Figure 3. Schematic representation of the postprocessing used to compute the GCE. The lineshapes of the distributions do not reproduce quantitatively equation (4). (1) Different colors indicate bitstring samplings induced by different eigenstates. Note that, depending on the associated eigenvalues the degree of overlap of two peaks may differ. (2) The effect of the tempered Softmax (equation (12)) is to spot the position of the most sampled peak by suppressing all other contributions. (3) Based on the distribution in (2) we are able to build a smooth boxcar function around the peak of interest. (4) Finally, we can estimate the GCE through the convolution of the parent distribution and the boxcar filter.

increasing rapidity. As we shall see later, from an algorithmic point of view this behavior means that the additional sampling cost resulting from the definition of this estimator becomes negligible as we increase the number of readout qubits.

All the above statements about the circular average estimator hold if the input state of our QPE algorithm is an eigenstate of the Hamiltonian. We now extend them to an arbitrary input state.

First of all, it is important to understand why we cannot use equation (8) directly when $|\psi\rangle$ is not an eigenstate of U. As recalled in section 2.1, the probabilities $P(0.\phi)$ result from the contributions of all eigenstates with non-zero overlap with the initial state.

Thus, if we straightforwardly applied the definition of first trigonometric moment to this new distribution we would obtain:
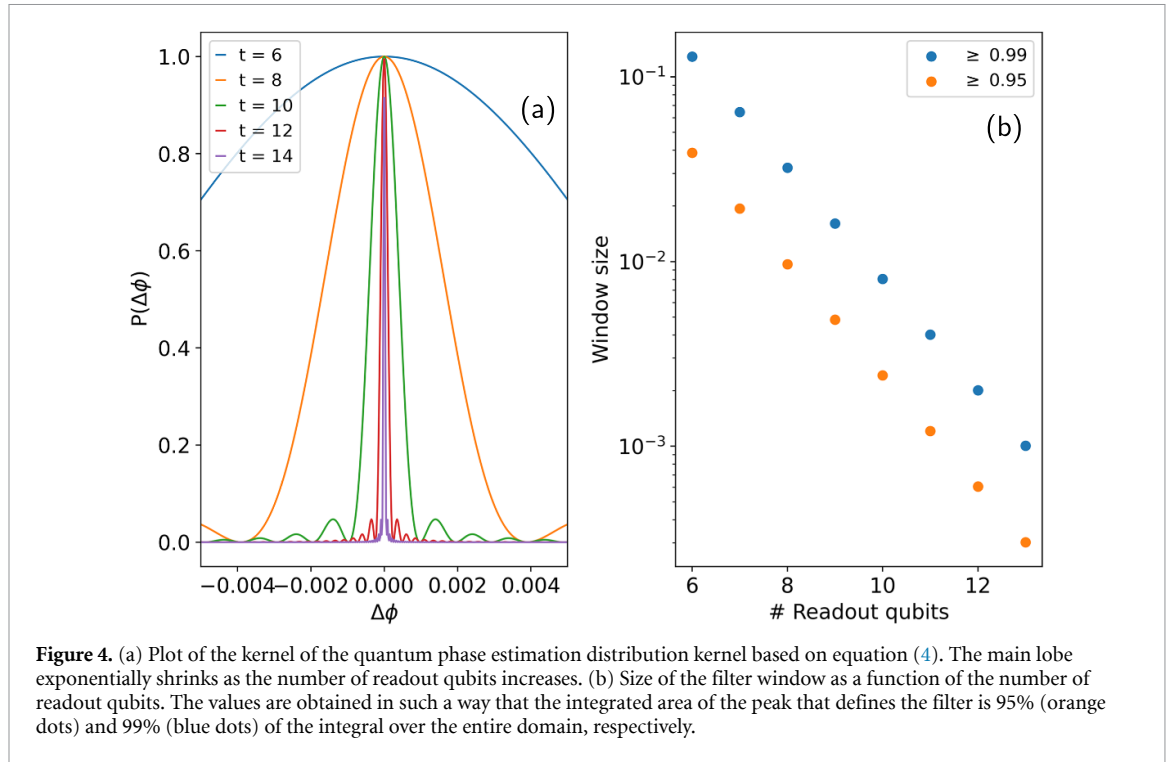
$$\theta' = \sum_u |c_u|^2 \sum_{0.\phi} P_{\phi_u}(0.\phi) \, e^{i2\pi \, 0.\phi} = \sum_u |c_u|^2 \theta_u = \rho' e^{i2\pi \mu'}. \tag{11}$$

The issue with this expression is that we are now taking a circular average that takes into account the contributions due to different eigenstates. Interestingly, we could show that the last expression leads to an estimator for $\langle\psi|H|\psi\rangle$ as $\langle H\rangle = \arg(\theta')$. We will not delve into this point, as it is beyond the scope of this study. Nonetheless we point out that comparing the cost of this estimator with existing algorithms for evaluating expectation values in a fault-tolerant setting [20] could be a matter of interest in itself.

Equation (11) is important not only because it provides us with an estimator for an expectation value of a generic $|\psi\rangle$ state, but also because it gives us a clue as to what needs to be done to generalize equation (6). Indeed, if we were able to filter all the spurious contributions due to the presence of the states $|u'\rangle \neq |u\rangle$ out of the parent distribution, we could use again the original definition of θ. The strategy we adopt here is as follows: (i) we develop a method to obtain a new distribution similar to the initial distribution except for the contribution of all the $|u'\rangle$s not of interest; (ii) we apply the original definition of θ to this new distribution to obtain a GCE $\tilde{\theta}$. These steps are summarized in figure 3.

To accomplish the first step we assume that our initial state has the highest overlap $|c_u|^2$ with the eigenstate of interest $|u\rangle$. This is a standard assumption, since we are usually interested in minimizing the computational cost, which for the majority rule estimator scales as $\mathcal{O}(\frac{1}{|c_u|^2})$.

After that, we can identify the region of the parent distribution induced by $|u\rangle$ using a tempered Softmax function S_T. This function amplifies differences in the original distribution depending on a parameter T

Figure 4. (a) Plot of the kernel of the quantum phase estimation distribution kernel based on equation (4). The main lobe exponentially shrinks as the number of readout qubits increases. (b) Size of the filter window as a function of the number of readout qubits. The values are obtained in such a way that the integrated area of the peak that defines the filter is 95% (orange dots) and 99% (blue dots) of the integral over the entire domain, respectively.

(often referred to as a temperature due to its equivalence to the Boltzmann distribution) according to the following rule:

$$P'(0.\phi) = S_T(P(0.\phi)) = \frac{e^{-\frac{P(0.\phi)}{T}}}{\sum_{0.\phi} e^{-\frac{P(0.\phi)}{T}}}. \tag{12}$$

The shape of $P'(0.\phi)$ depends on the chosen temperature T and the initial state $|\psi\rangle$ in the same way as the Boltzmann distribution in statistical mechanics [58]. In particular, when the temperature is very low we tend to get a one-hot distribution [59] peaked on the most sampled bitstring of the distribution, i.e. $0.\phi_u$. Subsequently, we can extract the $0.\phi_u$ value by weight-averaging the $P'(0.\phi)$ distribution with the binary fraction represented on the computational basis, $0.\phi_u = \sum_{0.\phi} 0.\phi P'(0.\phi)$.
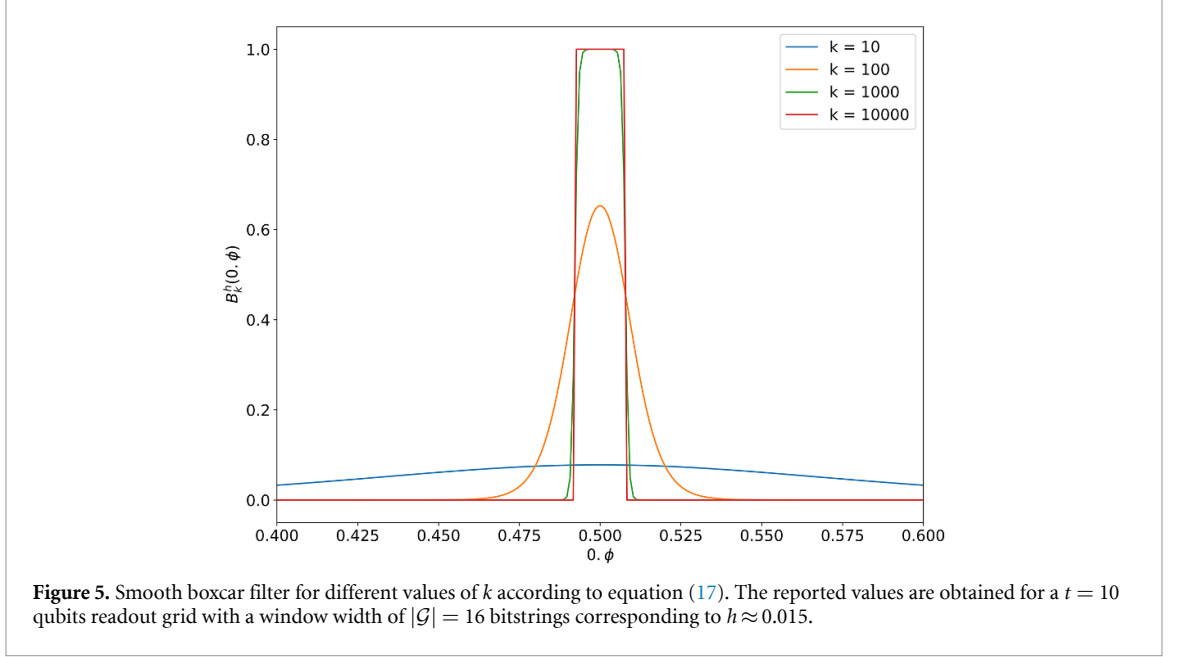
Once the above is done, we want to identify a subset of bitstrings around $0.\phi_u$ to be used for circular averaging exactly as in the case of equation (6). The reason why we assume we can decouple the contributions of different eigenvalues in populating the same string lies in the definition of $P_{\phi_u}(0.\phi)$. As shown in figure 4(a), where we plot equation (4) for different values of t, $P_{0.\phi_u}$ is a function that shrinks exponentially with the number of readout qubits. The full width at half maximum (g_{FWHM}) of the central maximum depends on the readout qubits as [60]

$$g_{\text{FWHM}} = \frac{1}{2^t}. \tag{13}$$

This feature is essential for our filtering procedure as it allows us to identify a subset of binary strings $0.\phi$ for each eigenstate $|u\rangle$ such that the following condition holds:

$$P(0.\phi) \approx |c_u|^2 P_{\phi_u}(0.\phi). \tag{14}$$

The meaning of this statement is that when two eigenphases are far apart the contribution of one of the two to the sampling probability of the given bitstring will be negligible. In other words, given a sufficient number of readout qubits we can always find a bitstring $0.\phi$ for which $P(0.\phi) - \frac{|c_u|^2}{2^t} P_{\phi_u}(0.\phi) < \gamma$, where γ is an arbitrarily small number.

Figure 5. Smooth boxcar filter for different values of k according to equation (17). The reported values are obtained for a $t = 10$ qubits readout grid with a window width of $|\mathcal{G}| = 16$ bitstrings corresponding to $h \approx 0.015$.

These arguments allow us to propose an extension of θ to arbitrary input states, which we label $\tilde{\theta}$, as:

$$\tilde{\theta} = |c_u|^2 \sum_{0.\phi \in \mathcal{G}} P_{\phi_u}(0.\phi)\, e^{i2\pi 0.\phi}. \tag{15}$$

Here we have introduced the set of relevant bitstrings $\mathcal{G} = \{0.\phi_u - h, 0.\phi_u + h\}$, where the width parameter h determines the number of bitstrings included in the circular average.

In practice, when we run the QPE circuit, we can only access measured probabilities. To get a distribution similar to equation (15) we apply a smooth boxcar filter centered at $0.\phi_u$:

$$\tilde{\theta} = \sum_{0.\phi} B_k^h(0.\phi)\, P(0.\phi)\, e^{i2\pi 0.\phi} \tag{16}$$

where $B_k^h(0.\phi)$ is given by:

$$B_k^h(0.\phi) = \frac{1}{2}\left\{ \tanh\left[k\left(0.\phi - (\phi_u - h)\right)\right] - \tanh\left[k\left(0.\phi - (0.\phi_u + h)\right)\right]\right\} \tag{17}$$

and k is a tunable parameter which accounts for the steepness of the window function. As we can see (figure 5), the value of k must be chosen appropriately to ensure that the assumptions underlying equation (15) are valid.

Finally, we obtain the following expressions for the first trigonometric moment $\tilde{\theta}$ and its mean phase direction $\tilde{\mu}$ (which we recall is our estimator for the target eigenvalue):

$$\tilde{\theta} = |c_u|^2 \sum_{0.\phi \in \mathcal{G}} P_{\phi_u}(0.\phi)\, e^{i2\pi 0.\phi} \approx \sum_{0.\phi} B_k^h(0.\phi)\, P(0.\phi)\, e^{i2\pi 0.\phi} = |c_u|^2 \rho\, e^{i2\pi \tilde{\mu}} \tag{18}$$

$$\tilde{\mu} = \arg\left(\tilde{\theta}\right) = \frac{1}{2\pi |\tilde{\theta}|} \arctan\left(\frac{\operatorname{Im}\left(\tilde{\theta}\right)}{\operatorname{Re}\left(\tilde{\theta}\right)} \right). \tag{19}$$

At this point we need to understand how to choose a suitable value for h. This should be done by balancing two factors: (i) including as many bitstring as possible, (ii) avoiding spurious contributions from the distributions of other eigenstates'. A possible heuristic to obtain a reasonable value for h is to consider a window size such that the ratio $\frac{I_h(g)}{I(g)}$ between the integrated area of $P(\Delta\phi)$ over the window spanned by h and the integral over the complete domain $I(g)$ is at least 0.95 (see figure 4(b)). Such a strategy may prioritize the first constraint over excluding the distributions of other eigenstates from the circular average. A formal analysis is required to better understand this point.

In the next section we will provide analytical expressions for $\tilde{\mu}$ that will allow us to provide a rationale for this heuristic procedure and demonstrate the accuracy of the GCE. Lastly, we will also be able to quantitatively account for the sampling cost associated with the measurement distribution.

3.1. Statistical analysis of the GCE

In this section we derive analytic expressions for the GCE that allow us to highlight its numerical and statistical properties.

By substituting equation (4) in the definition of θ, in the limit of a large number of readout qubits, the resolution of the grid with which we estimate the phase ϕ becomes infinitesimal and we can write

$$\tilde{\theta} = \frac{|c_u|^2}{2^t} \int_{0.\phi_u - h}^{0.\phi_u + h} \frac{\sin^2\left(2^t \pi \left(\phi - x\right)\right)}{\sin^2\left(\pi \left(\phi - x\right)\right)} e^{i2\pi x} dx \tag{20}$$

the variable $0.\phi$ was changed to the continuous integration variable x and $0 \leqslant h \leqslant \frac{1}{2}$.

Compared with the original formulation of Cruz *et al* here we only obtain the first term of equation (8) as we do not carry out the summation but integrate over a continuous variable. The effect of this continuum approximation on the accuracy of our analysis is negligible provided that we consider about $t \geqslant 10$ qubits.

After some algebra we rewrite equation (20) as

$$\tilde{\theta} = \frac{|c_u|^2 e^{i2\pi\phi}}{\pi \, 2^t} \int_{\pi(\Delta\phi - h)}^{\pi(\Delta\phi + h)} \sum_{n,n'=0}^{2^t - 1} e^{2i\left(n + n' - 2^t\right)u} du \tag{21}$$

where we used Euler equations to manipulate the integral function and changed the integration variable to $u = \pi(\phi - x)$. Furthermore, we introduced $\Delta\phi = \phi - 0.\phi_u$ as a shorthand notation for the mismatch between the real eigenphase and the center of the filtering box.

To solve equation (21) we treat separately the terms with $n + n' = 2^t$ and $n + n' \neq 2^t$:

$$
\begin{aligned}
\tilde{\theta} &= \frac{|c_u|^2 e^{i2\pi\phi}}{\pi \, 2^t} \left[(2^t - 1) \int_{\pi(\Delta\phi - h)}^{\pi(\Delta\phi + h)} du + \int_{\pi(\Delta\phi - h)}^{\pi(\Delta\phi + h)} \sum_{n,n'=0,\,n+n'\neq 2^t}^{2^t - 1} e^{2i\left(n + n' - 2^t\right)u} du \right] \\
&= \frac{|c_u|^2 e^{i2\pi\phi}}{\pi \, 2^t} \left| \left[(2^t - 1)\,u - i \sum_{n,n'=0,\,n+n'\neq 2^t}^{2^t - 1} \frac{e^{2i\left(n + n' - 2^t\right)u}}{2\left(n + n' - 2^t\right)} \right] \right|_{\pi(\Delta\phi - h)}^{\pi(\Delta\phi + h)} \\
&= \frac{|c_u|^2 e^{i2\pi\phi}}{\pi \, 2^t} \left[2\pi (2^t - 1)h - i \sum_{n,n'=0,\,n+n'\neq 2^t}^{2^t - 1} \frac{e^{2i\left(n + n' - 2^t\right)\pi(\Delta\phi + h)} - e^{2i\left(n + n' - 2^t\right)\pi(\Delta\phi - h)}}{2\left(n + n' - 2^t\right)} \right].
\end{aligned}
\tag{22}
$$

We are interested in studying the limiting cases of this formula in which (i) the box spans all possible values of the phase (i.e. $h = \frac{1}{2}$) and (ii) the box is very narrow and of width comparable to the grid spacing (i.e. $h = \frac{|\mathcal{G}|}{2^t}$ where $|\mathcal{G}|$ is the number of bitstrings included in $\mathcal{G}$). The first case is interesting because it allows us to recover the original picture of Cruz *et al* [41]. The second case is useful as it represents the typical regime of use for the GCE.
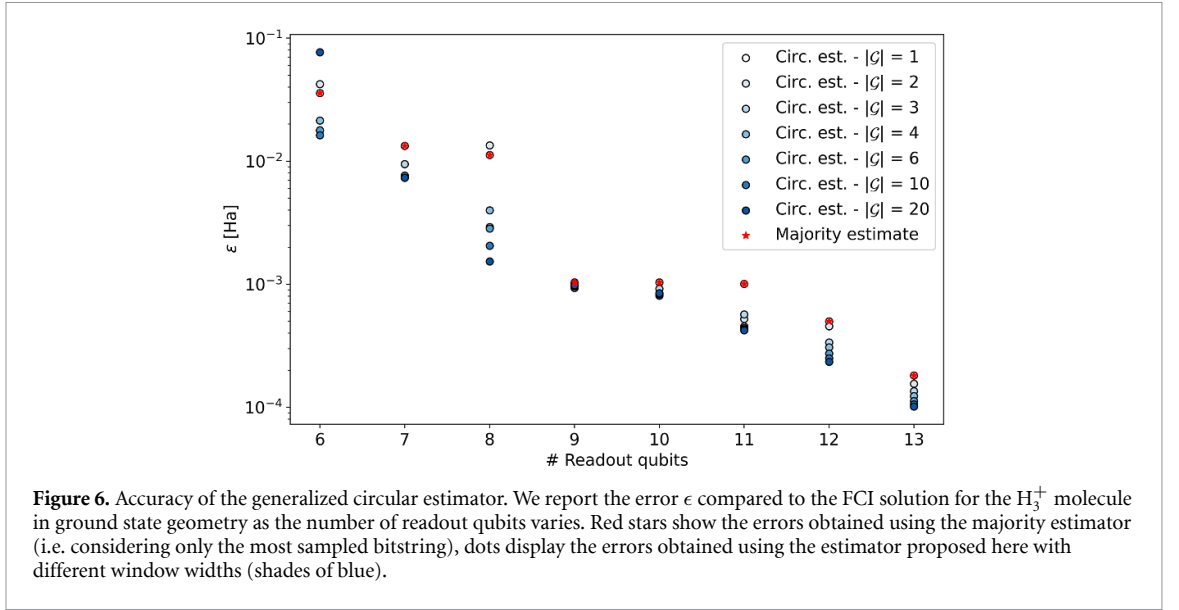
Starting with the case in which $h = \frac{1}{2}$, it is straightforward to show that $\tilde{\theta} = \frac{|c_u|^2 e^{i2\pi\phi}}{2^t}(2^t - 1)$, and therefore

$$|\tilde{\theta}| = |c_u|^2 \frac{2^t - 1}{2^t} . \tag{23}$$

This is exactly the result of [41] when $|c_u|^2 = 1$ (i.e. the input state is an eigenstate) in the limit of large t.

Let us now move to the case in which $h = \frac{|\mathcal{G}|}{2^t}$. We expand the complex exponential functions in the summation up to the second order in close proximity to $\Delta\phi = 0$ and $h = 0$, which holds when the number of readout qubits is reasonably high:

$$
\begin{aligned}
\tilde{\theta} &\approx \frac{|c_u|^2 e^{i2\pi\phi}}{\pi \, 2^t} \left[2\pi (2^t - 1)h - \frac{i}{2}\left(i4\pi h \sum_{n,n'=0}^{2^t - 1} 1 - 8\pi^2 \Delta\phi h \sum_{n,n'=0}^{2^t - 1} \left(n + n' - 2^t\right) \right) \right] \\
&= \frac{|c_u|^2 e^{i2\pi\phi}}{\pi \, 2^t} \left[2\pi (2^t - 1)h + 2\pi h 2^{2t} - i4\pi^2 \Delta\phi h 2^{2t} \right] \\
&= \frac{|c_u|^2 e^{i2\pi\phi}}{\pi \, 2^t} \left[2\pi (2^t - 1)\frac{|\mathcal{G}|}{2^t} + 2\pi |\mathcal{G}| 2^t - i4\pi^2 2^t \Delta\phi |\mathcal{G}| \right] \\
&= \frac{2|c_u|^2 |\mathcal{G}| |B|}{2^t} e^{i(2\pi\phi - \alpha)} \\
&\approx 2|\mathcal{G}| |c_u|^2 e^{i2\pi\phi}.
\end{aligned}
\tag{24}
$$

Figure 6. Accuracy of the generalized circular estimator. We report the error ϵ compared to the FCI solution for the H_3^+ molecule in ground state geometry as the number of readout qubits varies. Red stars show the errors obtained using the majority estimator (i.e. considering only the most sampled bitstring), dots display the errors obtained using the estimator proposed here with different window widths (shades of blue).

To pass from the second to the third line we used the fact that $h = \frac{|\mathcal{G}|}{2^t}$. In the fourth line we introduced B as a shorthand for

$$B = |B|e^{-i\alpha} = (1 - i2\pi\Delta\phi)\,2^t + \frac{2^t - 1}{2^t}\,. \tag{25}$$

Finally, the last step in equation (24) is carried out by considering that $\Delta\phi$ is at most $\frac{1}{2^t}$ and keeping only the leading term in 2^t. The additional terms originate as an artifact of the smoothing procedure, but it is easy to show that for large t values $|B| = \mathcal{O}(2^t)$ whence $\alpha \equiv \arcsin\left(\frac{2\pi}{|B|}\right) \approx 0$.
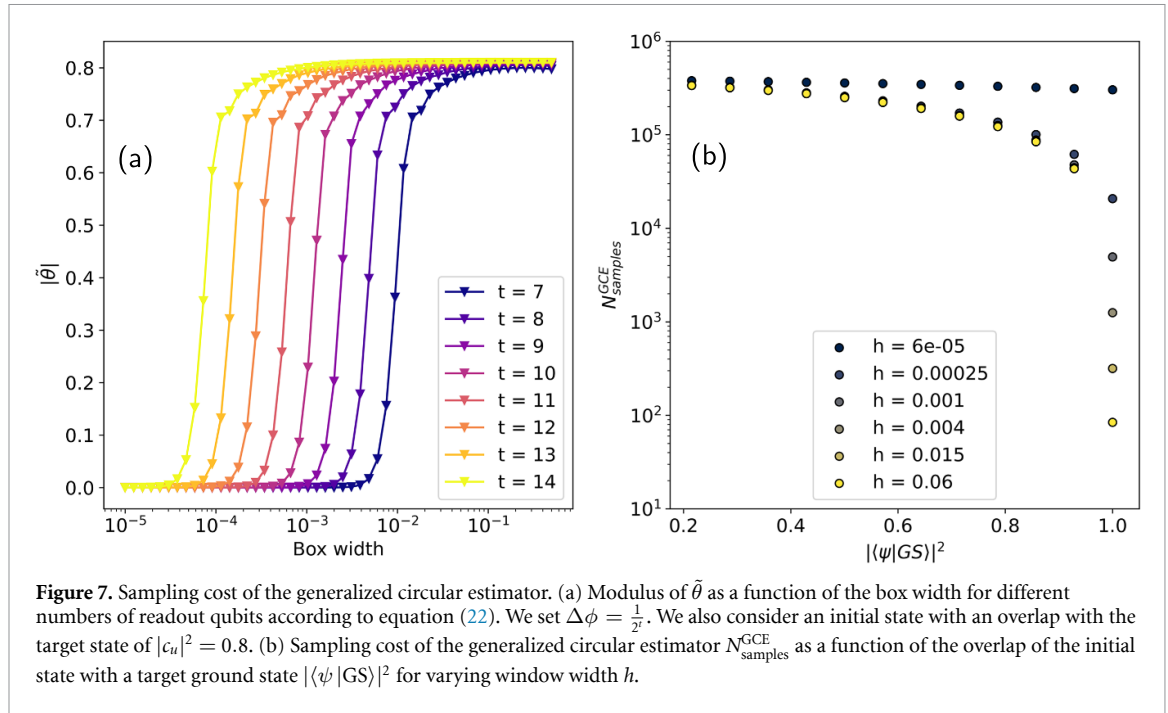
The expression in equation (24), obtained expanding equation (22) around small values of $\Delta\phi$ and h, usefully shows that the GCE introduces an exponentially vanishing bias into our estimator. In practice, considering the complete expression, the GCE accuracy can exceed that of the standard majority estimator. This result was already found for the case of an exact input eigenstate [41]; here we show that it holds for an arbitrary input state provided that the measurement distribution is reconstructed via a sufficient number of samples. For example, in figure 6 we report the error ϵ made by the GCE in estimating the energy of an electronic state as a function of the number of readout qubits (dots). We compare this error to that made using the standard majority rule (stars). Furthermore, we study the effect of the number of bitstrings included in the circular average, i.e. $|\mathcal{G}|$ (different shades of blue). First of all, we note that the GCE performs better than the majority rule estimate for all values of t. More precisely the error of the latter scales as $\epsilon_{\text{MR}} = \frac{1}{2^{t+1}}$ while the circularly averaged estimated error scales as $\epsilon_{\text{GC}} = \frac{1}{2^{t+2}}$.

Regarding the dependence of the error on the window size, two effects are noticeable: (i) for relatively small values of t, there are some values of h for which we make a larger error than expected using the standard estimator; (ii) contrary to expectations, for low t values it is not true that a wider window necessarily gives a more accurate result. The first effect is related to the point discussed at the end of the last paragraph, that including too many bitstrings when peaks corresponding to different eigenstates are not sufficiently separated can introduce spurious errors. The second effect stems from the fact that the improved accuracy of the GCE is due to error cancellation effects. With a lower number of qubits, this error cancellation may undergo significant variation due to discretization effects on the averaged distribution $P(0.\phi)$.

We now conclude this section by answering the following question: how many samples are necessary to reconstruct the parent distribution with the accuracy shown in figure 6?

To this end, we compute the variance of the mean phase direction $\tilde{\mu}$ and use Chebyshev's inequality:

$$N_{\text{samples}}^{\text{GCE}} \geqslant \frac{\text{Var}\,(\tilde{\mu})}{\epsilon^2}\,. \tag{26}$$

Figure 7. Sampling cost of the generalized circular estimator. (a) Modulus of $\tilde{\theta}$ as a function of the box width for different numbers of readout qubits according to equation (22). We set $\Delta\phi = \frac{1}{2^t}$. We also consider an initial state with an overlap with the target state of $|c_u|^2 = 0.8$. (b) Sampling cost of the generalized circular estimator $N^{\text{GCE}}_{\text{samples}}$ as a function of the overlap of the initial state with a target ground state $|\langle\psi|\text{GS}\rangle|^2$ for varying window width h.

The variance of the mean phase direction is obtained by applying the error propagation rule to $\tilde{\theta}$ (equation (27)), assuming the latter is a proper random variable [61] and using the definition of variance from circular statistics [57] (equation (28)):

$$\text{Var}\left(\tilde{\mu}\right) = \frac{1}{2}\left[\left|\frac{\partial\tilde{\mu}}{\partial\text{Re}\left\{\tilde{\theta}\right\}}\right|^2 \text{Var}\left(\tilde{\theta}\right) + \left|\frac{\partial\tilde{\mu}}{\partial\text{Im}\left\{\tilde{\theta}\right\}}\right|^2 \text{Var}\left(\tilde{\theta}\right)\right] \tag{27}$$

$$\text{Var}\left(\tilde{\theta}\right) = 1 - |\tilde{\theta}|. \tag{28}$$

In figure 7 we study the cost of obtaining an estimate for the mean phase direction with an error of $\epsilon = 1$ mHa. In particular, in figure 7(a) we plot $|\tilde{\theta}|$ with varying number of readout qubits and box width h. We can see that the effect of increasing the number of readout qubits is to recover the complete information of the peak (that is, the limit value of the quantity in equation (23) is reached) with a smaller window. As can be seen from equations (26)–(28), this directly impacts the cost of the routine. Indeed, as we approach the situation of equation (23), we tend to lower the estimation cost, and to do so carefully (i.e. without including contributions of other peaks in the average) we need to increase the grid resolution. Similar considerations can be drawn from figure 7(b) where we see the effect of the initial state overlap on the cost while varying the window width at fixed number of readout qubits ($t = 13$). Numerical analysis is presented in appendix B.

Overall we see that the increased accuracy of $\tilde{\mu}$ has a considerable computational cost, especially if the input state has a small overlap with the eigenstate of interest. Further investigation is required to reduce such a computational overhead and thus enhance the scalability of the approach with respect to the size of the system. The improvement would be particularly relevant to the treatment of strongly correlated systems, where, even for modest system size, the initial state is generally expected to have a relatively small overlap with the target state [38].

It is worth commenting on the possible impact of this routine when considering the simulation of other molecular degrees of freedom, i.e. vibrations and rotations. A possible approach would then be to use the n-mode method developed by Christiansen [62] and later applied in the context of variational quantum algorithms [63]. In this context we would be interested in obtaining quantum-refined energies for systems characterized by strong anharmonicities. We envisage that to use our estimator without incurring in large computational overhead one could use as input states wavefunctions obtained with vDMRG methods [64] obtained at low bond dimensions. This approach could be extended to resolve the rotational energy levels within a vibrational peak. In order to do so, following Yurchenko *et al* [65], one should include rotovibrational couplings into the system Hamiltonian and expand the qubit register to account for the rotational wavefunctions which can be represented in the rigid-rotor basis.

Finally, we recall that beyond an improved accuracy this approach allows to obtain gradients of the estimated eigenvalues. This is a completely different task and in the next section we go through the costing of this routine and compare with existing methods in literature.

4. Cost of gradient estimation

Building on the previous analysis, in this section we analyze the cost of evaluating the gradient with the quantum differentiable program sketched in figure 2 compared to state-of-the-art methods for evaluating energy gradients for molecular Hamiltonians on a fault-tolerant QC.

4.1. Computational cost analysis
The overall cost for the gradient evaluation with the differentiable QPE is given by:

$$\tilde{C} = CNk \tag{29}$$

where C is the cost of a single QPE circuit evaluation, N is the number of calls to the routine needed to evaluate $\tilde{\mu}$ with an ϵ-good estimate and k is the number of calls required for the gradient evaluation. In this section we estimate $\tilde{C}$ breaking down the different contributions.

Regarding molecular systems, we note that the scaling of the computational cost of the routine with system size (meant as the number of electrons or spin-orbitals) is included in the quantity C and is not considered here. As for N, by definition we have $N = N_{\text{samples}}^{\text{GCE}}$ as given by equation (26).

Thus, all that remains is to estimate k. The number of calls to the QPE subroutine to estimate the gradient with a parameter-shift rule, following Wierichs *et al* [53], is given by $k = RM$ where M is the number of Cartesian coordinates (or more generally the number of variables) with respect to which the differentiation is performed and R is a factor counting either the number of parameter-dependent gates (decomposition based parameter-shift rule) or the number of unique positive differences in the parameterized circuit spectrum (Fourier based parameter-shift rule). Depending on the structure of the Hamiltonian the number of independent circuit executions required by the parameter-shift rule can vary substantially. Since the eigenvalue distribution for a molecular Hamiltonian does not have an ordered structure, in the cost analysis it is safer to consider in the cost analysis a decomposition based strategy for the quantum gradient. If we denote G the number of gates where the parameters appear after the compilation of the simulation routine implemented in the quantum circuit, we obtain:

$$k \approx \mathcal{O}\left(GtM\right) \tag{30}$$

where t is the number of readout qubits and therefore of calls to the quantum simulation routine within the phase estimation circuit.

The last step needed to estimate the overall cost $\tilde{C}$ is to obtain the number of shots required for the gradient estimation. We remark that this last step is necessary since we are considering an actual quantum differentiable program, i.e. the QPE node is executed on a quantum device. As we focus on the quantum chemistry framework, we label $\mathbf{R}$ (to indicate nuclei positions) the parameters with respect to which we take the gradient. Incidentally, it is worth emphasizing once again that the scope of application of energy derivatives and AD for quantum chemistry is much broader than just geometry optimization.

Within the (either decomposition-based or Fourier-based) parameter-shift framework we can write the force acting on the nucleus at the position R_j using the GCE as:

$$\frac{\partial \tilde{\mu}}{\partial R_j} = \sum_{l}^{k} y_l \tilde{\mu}(\mathbf{R}_l) \tag{31}$$

where k is the number of function evaluation required by the parameter-shift rule and $\mathbf{y}$ is a coefficient vector defined by the particular flavour of gradient rule used.

The optimal budget shot allocation to obtain $\frac{\partial \tilde{\mu}}{\partial R_j}$ within a tolerable error ϵ is:

$$N_{\partial_j \tilde{\mu}} = \frac{\text{Var}\left(\tilde{\mu}\right) ||\mathbf{y}||_1^2}{\epsilon^2}. \tag{32}$$

Finally, the number of queries to a circuit that implements the evolution operator U for gradient estimation with a fully differentiable QPE can be written as:

$$\tilde{C} = \mathcal{O}\left(GtMN_{\partial_j \tilde{\mu}}\right). \tag{33}$$

Interestingly, lowering the number of readout qubits affects simultaneously: (i) t, (ii) ϵ and (iii) $N_{\partial_j \tilde{\mu}}$ as we expect that the 1-norm of the parameter-shift rule's coefficients decreases as their number decreases. Methodologies such as those developed in [66] may be extremely useful in this context.

4.2. Related literature on gradient estimation

The goal of this section is to summarize previous related studies on the evaluation of energy derivatives within the context of quantum computation. We refer to the study of O'Brien *et al* [42] for a more detailed report on state-of-the-art methods to compute energy gradients in a fault-tolerant setting.

We remark that all our considerations are made for a fully variational wavefunction in the sense that, given a set of *external physical parameters* **x** and a set of *wave function parameters* λ we calculate the energy from the expression

$$E(\mathbf{x}) = E(\mathbf{x}, \lambda^*) \tag{34}$$

where the parameter set λ^* represents the optimal value of λ and where the optimized energy function $E(\mathbf{x}, \lambda^*)$ satisfies the variational conditions for all values of the external parameters **x**. This statement also holds for the quantities extracted from QPE since, within the discretization error due to the finite number of readout qubits, we sample from the exact eigenbasis of the molecular Hamiltonian.

4.2.1. Hellmann–Feynman approach

This first approach is based on the Hellmann–Feynman theorem [67], which expresses the derivative of the total energy with respect to a parameter (a nuclear coordinate in this case) as the expectation value of the derivative of the Hamiltonian operator with respect to that same parameter:

$$\frac{\mathrm{d}E}{\mathrm{d}\mathbf{x}} = \langle \Psi | \frac{\mathrm{d}H}{\mathrm{d}\mathbf{x}} | \Psi \rangle. \tag{35}$$

Equation (35) does not provide a recipe for directly computing energy gradients but only tells us the quantity that we need to calculate. Among the various strategies proposed in the literature, a possible strategy is to use the overlap estimation algorithm which allows one to compute equation (35) at the Heisenberg limit (i.e. with a cost scaling as $\mathcal{O}(\epsilon^{-1})$). This method provides an estimate of the energy gradient by directly calculating the expectation value with the algorithm proposed in [68]. However, since we are dealing with a real expectation value we only need to call the amplitude estimation algorithm twice instead of three as originally sketched in [68].

It is important to notice that this method requires using a block-encoding routine for the energy derivative operator. Using tensor hyper-contraction the authors of [42] were able to show that, despite a larger prefactor, the asymptotic cost of block-encoding $\frac{\mathrm{d}H}{\mathrm{d}\mathbf{x}}$ equals the asymptotic cost of block-encoding the molecular Hamiltonian.

4.2.2. Direct gradient-based quantum estimation algorithm

Another ad hoc approach developed for computing gradients of functions [42] can be used as an alternative to the previous method. Building upon the algorithm developed in [69] and refined in [70], O'Brien *et al* [42] demonstrate that nearly optimal scaling can be achieved for calculating the forces of molecular Hamiltonians.

Without going into details the main idea is that using a single query to a state preparation routine U_ψ (which implies that we know how to prepare the exact ground state of the molecule) and querying $\mathcal{O}(\frac{\sqrt{M}}{\epsilon})$ (where M is the number of variables to differentiate) times the oracle

$$U_p : |x\rangle |0\rangle \to |x\rangle \otimes \left(\sqrt{f(x)} |\psi_1(x)\rangle |1\rangle + \sqrt{1-f(x)} |\psi_0(x)\rangle |0\rangle \right) \tag{36}$$

it is possible to simultaneously calculate a given $f(\mathrm{x})$ at different points and compute a finite difference approximation of the function directly on the QC.

This algorithm, which uses different qubit registers, makes several calls to routines that implement operators like $U = \mathrm{e}^{-\mathrm{i}t\frac{\mathrm{d}H}{\mathrm{d}R_k}}$, to which the considerations in the previous paragraph on block-encoding apply.

4.2.3. Finite difference approach

Finally, another possibility is to consider the QPE algorithm itself as a black-box which the function to be sampled at a given point in the parameter space. From this standpoint, we can use this algorithm to compute gradients via classical finite difference approximations.

Following [42], we write the finite difference formula of degree $2m$ as:

$$\frac{\mathrm{d}E^{(m)}}{\mathrm{d}\mathbf{x}} = \sum_{l=-m}^{l=m} b_l^{(m)} E(\mathbf{x} + l\delta\mathbf{x}\cdot\mathbf{v})$$

$$b_l^{(m)} = \begin{cases} 0 & l = 0 \\ \frac{(-1)^{l-1}}{\delta\mathbf{x}l} \frac{\binom{m}{|l|}}{\binom{m+|l|}{|l|}} & l \neq 0. \end{cases} \tag{37}$$

While details on the computational cost of this algorithm can be found in [42], here we want to highlight the differences that arise from tackling equation (37) with the ME and GCE. Applying error propagation to the expression for the energy derivative in equation (37) we obtain:

$$\left| \frac{\mathrm{d}E}{\mathrm{d}\mathbf{x}} - \frac{\mathrm{d}E^{(m)}}{\mathrm{d}\mathbf{x}} \right| = \sum_{l=-m, l\neq 0}^{l=m} \epsilon_{\mathrm{PE}} |b_l|^2 + \epsilon_{\mathrm{FD}}. \tag{38}$$

Two sources of error appear in the last equation: (i) the error due to the phase estimation procedure ϵ_{PE} and (ii) the error due to the finite difference approximation ϵ_{FD}. At first, we focus on the error due to the phase estimation procedure. When using the ME estimator, $\epsilon_{\mathrm{PE}} = \epsilon_{\mathrm{ME}} = \frac{1}{2^{t+1}}$. Using the GC estimator instead, it is $\epsilon_{\mathrm{PE}} = \epsilon_{\mathrm{GC}} = \frac{1}{2^{t+2}} + \sqrt{\frac{\mathrm{Var}^{h,t}(\mu)}{N}}$. From the last two error expressions, we deduce that the gradient approximation obtained with the GC estimator is more accurate if

$$\sqrt{\frac{\mathrm{Var}^{h,t}(\mu)}{N}} \leqslant \frac{1}{2^{t+2}}. \tag{39}$$

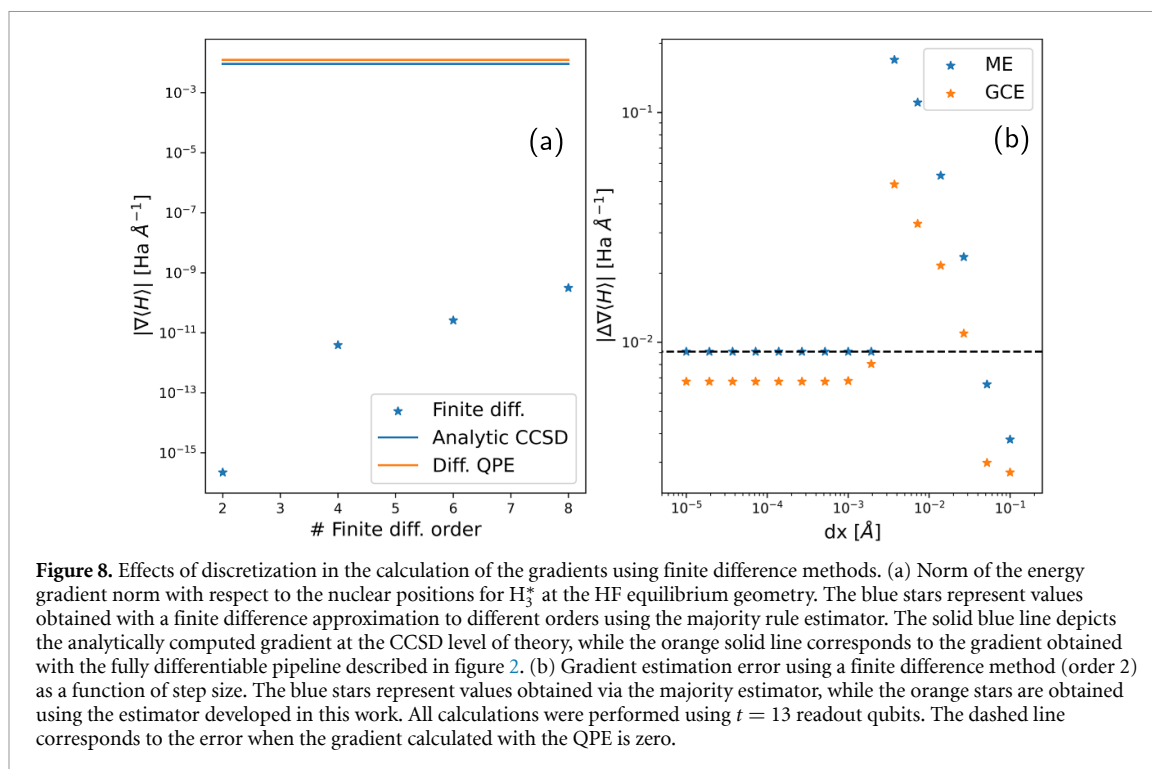At this point it is important to emphasize that, when using the ME, the step size $\delta\mathbf{x}$ used must satisfy

$$|\phi(\mathbf{x} + l\delta\mathbf{x}\cdot\mathbf{v}) - \phi(\mathbf{x} + l'\delta\mathbf{x}\cdot\mathbf{v})| > \frac{1}{2^t} \ \forall l, l'. \tag{40}$$

The last condition derives from the fact that for a signal to be detected with the finite difference approximation not only must the step size be small enough to avoid truncation errors, but it must also be such that function variations at different points can be spotted according to the QPE grid resolution. As shown in figure 8, even high-order approximations to the gradient fail unless both conditions are satisfied. In particular, figure 8(a) shows the gradient norm $|\nabla\langle H\rangle|$ (used as sensitivity metric) computed with a central finite difference approximation at different orders (blue stars) adopting an ME. For comparison, we reported as a solid blue line the analytical result at the CCSD computational level (which is exact for a two-electron system as the one considered in this case) and as a solid orange line the gradient norm resulting from the differentiable QPE. As we can see, even using a degree 8 finite difference approximation with a step size of $\delta x = 10^{-3}$ Å the resolution enabled by a $t = 13$ grid does not allow us to detect any signal. In figure 8(b) we compare the error on the gradient (reported as the difference from the CCSD value) estimated with a degree 2 finite difference method using the GCE (orange stars) and the ME (blue stars). The black dashed line is the exact gradient norm. Points lying on (above) this line are associated with null (or pointing in the wrong direction) gradients: in all these cases the inaccuracy makes the calculation useless. We see that the smooth GCE enables the computation of the gradient with very small step sizes, whereas the ME fails to detect any signal. Upon increasing δx the estimated gradient starts suffering from truncation errors. Eventually, for $\delta x \approx 0.1$ Å the estimated gradients becomes more accurate. However, these step size values have already proven to be unreliable in standard quantum chemistry software implementations.

These results show that using finite differences with the ME may not be feasible on a small scale fault-tolerant computer (e.g. using $t \approx 10$–20), despite the good asymptotic scaling shown in [42].

It is also important to remark a recent contribution by Sugisaki *et al* [71], where the authors proposed a lightweight Bayesian approach to QPE to compute energy differences and applied it to the evaluation of molecular forces using a 2-points finite difference method.

Summarizing the comparative analysis of this section, our approach currently suffers from a finite-shot sampling cost which makes it unfavourable compared to other strategies such as sections 4.2.1 and 4.2.2 which saturate the Heisenberg limit. Nonetheless, our algorithm does not necessitate the preparation of the exact ground state on the QC, a requirement that, on the other hand, is mandatory in both sections 4.2.1 and 4.2.2. This condition could potentially limit the application of these methodologies to regimes in which classical computation proves to be more efficient than quantum computation [37, 38]. Further, the proposed strategy is more flexible when it comes to considering extensions of this algorithm towards higher-order

Figure 8. Effects of discretization in the calculation of the gradients using finite difference methods. (a) Norm of the energy gradient norm with respect to the nuclear positions for H_3^* at the HF equilibrium geometry. The blue stars represent values obtained with a finite difference approximation to different orders using the majority rule estimator. The solid blue line depicts the analytically computed gradient at the CCSD level of theory, while the orange solid line corresponds to the gradient obtained with the fully differentiable pipeline described in figure 2. (b) Gradient estimation error using a finite difference method (order 2) as a function of step size. The blue stars represent values obtained via the majority estimator, while the orange stars are obtained using the estimator developed in this work. All calculations were performed using $t = 13$ readout qubits. The dashed line corresponds to the error when the gradient calculated with the QPE is zero.

derivatives. In fact, the extension to higher order derivatives of the Hellmann–Feynman type of approach or the gradient-based estimation algorithm requires an exponentially higher [72] cost or redesigning the same algorithm, respectively.

Next, considering the Bayesian phase energy difference [71], an accurate analysis regarding the total runtime is still missing. The shallow circuit proposed by the authors combined with the Bayesian optimization could make it a competitive option also considering that this approach would allow the evaluation of derivatives without the knowledge of the exact ground state (or excited state). However, this approach would suffer from the fact that the accuracy of the gradients is limited by the 2-points finite difference approximation, which is known to be too crude for some potential energy surfaces.

The goal of future work will be to restore optimal (Heisenberg) scaling of the approach proposed here by leveraging techniques such as compressed sensing [73] or amplitude amplification [74] tailored to an automatically differentiable framework. Another essential direction of research to align this strategy with other proposals in the literature and possibly improve on them concerns reducing the cost of differentiation itself. Currently, quantum differentiation methods exhibit backpropagation-like scalings for only a specific class of circuits [75].

5. Geometry optimization

Here we report the numerical results of our implementation of the GCE described in section 3 coupled to the automatically differentiable pipeline provided by PennyLane. The latter, thanks to the differentiable quantum chemistry feature [46], ensures the end-to-end differentiability of the procedure and enables its application to the calculation of energy gradients of molecular Hamiltonians, which is of utmost importance in the field of quantum chemistry.

In addition to numerically validating the algorithm proposed in this study, the calculations themselves represent a novelty. To the best of our knowledge, only two other papers (focusing on fault-tolerant algorithms) in the literature report similar calculations: O'Brien *et al* [72] calculated the equilibrium bond length of H_2 with a minimal basis set on a real quantum processor, and, more recently, Sugisaki *et al* [71] developed a protocol to optimize molecular geometries using a Bayesian phase difference algorithm and applied it to different molecules of which the largest is N_2 with a Complete Active Space CI of 6 electrons in 6 orbitals (6e, 6o).

5.1. Computational details
All calculations were performed using a locally modified version of PennyLane [33] v0.24. The code generating all the data of this section can be found at [76].
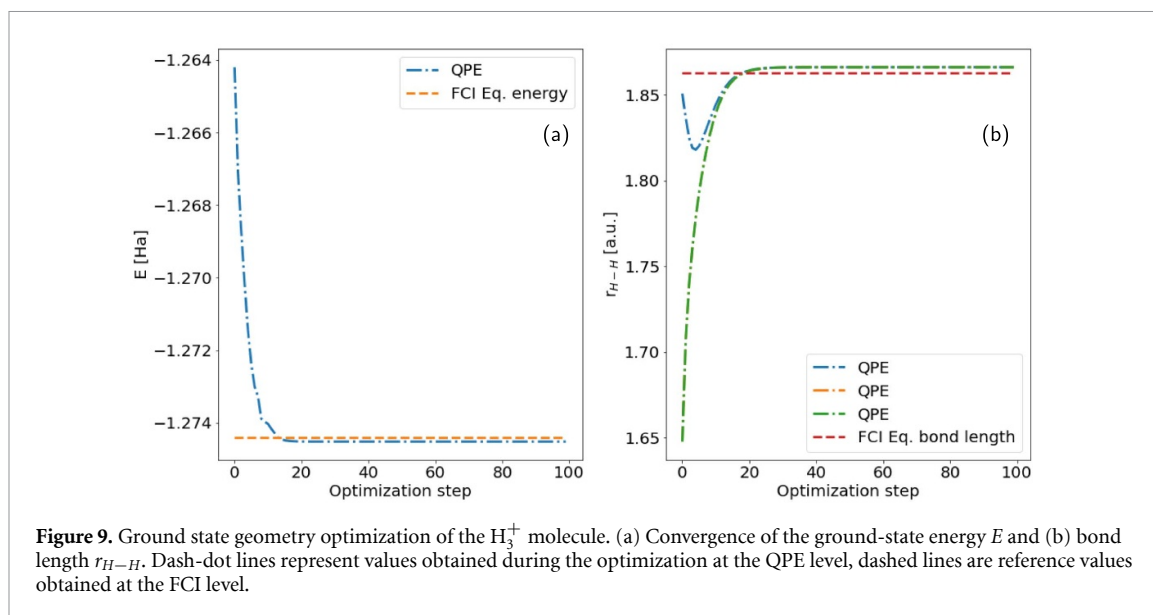
Figure 9. Ground state geometry optimization of the H_3^+ molecule. (a) Convergence of the ground-state energy E and (b) bond length r_{H-H}. Dash-dot lines represent values obtained during the optimization at the QPE level, dashed lines are reference values obtained at the FCI level.

We studied the trihydrogen cation and the formaldehyde. For the first molecule we used a gradient descent optimizer with stepsize of 0.01 Å as implemented in PennyLane, while for the second molecule we used an in-house automatically differentiable implementation of the BFGS optimizer [77].

Reference calculations at the full configuration interaction (FCI) level were performed with the Psi4 [78] code. CASCI calculations on the CH_2O molecule were carried out with the PySCF code [79]. The STO-3G basis set was employed in all calculations. For the H_3^+ molecule, adopting this basis set entails using six qubits to encode the wave function into the system register by Jordan–Wigner encoding [80]. For the CH_2O molecule, the active space contained two electrons in the four highest lying MOs. For all calculations we used $k = 1000$ to smooth the boxcar function and a window width $h = \frac{8}{2^t}$. We set the temperature for the tempered-softmax distribution (equation (12)) at $T = 0.0035$. We found that higher values of k (i.e. $k > 5000$) in some cases lead to numerical instabilities. We recall that a thorough analysis of the window size effects is given in section 3 (figure 6). We set the number of readout qubits to $t = 13$ for the H_3^+ molecule and $t = 11$ for the CH_2O molecule. Overall we exploited 19 qubits in both cases.

It is worth noting that QPE methods always involve a rescaling of the Hamiltonian to avoid aliasing of the energy found. In practice, it means working with a rescaled Hamiltonian $H_{SC} = \frac{H_{mol} - E_{min}\mathcal{I}}{\Delta E}$, where ΔE is the relevant spectral range. However, this procedure increases the computational cost, which scales as $\mathcal{O}(\Delta E \epsilon^{-1})$ rather than $\mathcal{O}(\epsilon^{-1})$. In all our numerical simulations we mitigated this issue by working with purely electronic Hamiltonians (i.e. adding *a posteriori* the frozen core contribution and/or the internuclear interaction terms).

5.2. Trihydrogen cation (H_3^+)

As a first application of our differentiable pipeline, we show two examples of geometry optimization of H_3^+. This system, being the simplest polyatomic molecule, has been extensively studied for benchmarking *ab initio* models [81, 82] and it has also attracted attention because it is one of the most common species in molecular hydrogen plasmas found in interstellar clouds and planetary ionospheres [83].

In figure 9 we report a convergence plot of the geometry optimization for the ground state which is stable if the hydrogen atoms are arranged in a triangular shape [84]. We label the energies and bond lengths computed with the differentiable QPE developed in this work as QPE. The initial state used for this calculation is the Hartree–Fock (HF) state, which after Jordan–Wigner mapping reads $|HF\rangle = |110000\rangle$. The initial geometry was set equal to an isosceles triangle with bond lengths slightly distorted from the HF equilibrium geometry. Note that in figure 9(b) the dash-dot lines refer to different bond lengths and that the green and yellow lines are superimposed. As we can see, after 20 iterations the optimization is able to recover the correct equilateral shape (D_{3h} symmetry) in agreement, within chemical accuracy, with the FCI optimized geometry. Interestingly, we notice that the energy obtained using the QPE is slightly below the FCI reference value (orange dashed line in figure 9(a)) due to the fact that the QPE is not a variational method.

As a second application example, we consider the first excited triplet state of the same molecule. Now the initial state is a spin-flipped doubly excited determinant built upon the HF configuration, i.e. $|\psi\rangle = t_{01}^{35}|HF\rangle = |0001010\rangle$. As reported in the literature [85], the triangular geometry is not stable in this excitation and undergoes photodissociation. On the other hand, a linear structure is a stable configuration
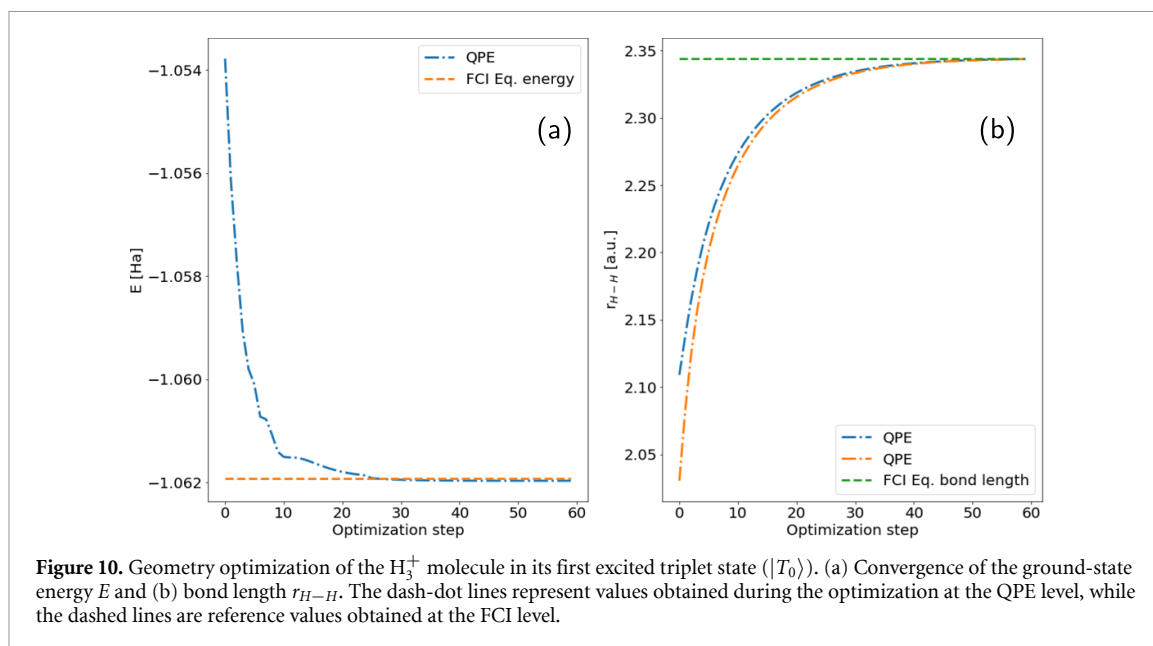
Figure 10. Geometry optimization of the H_3^+ molecule in its first excited triplet state ($|T_0\rangle$). (a) Convergence of the ground-state energy E and (b) bond length r_{H-H}. The dash-dot lines represent values obtained during the optimization at the QPE level, while the dashed lines are reference values obtained at the FCI level.

for the $|T_0\rangle$ state; thus, we initialized the geometry as linear with the two terminal hydrogens at different distances from the central one (see figure 10(b)). For this reason, differently from the previous figure, we show two dash-dot lines labeled as QPE since we have only two bond lengths to monitor during the optimization. Again, we see that the optimization allows us to reach energy and bond length with errors below chemical accuracy (i.e. $\Delta E < 1$ mHa and $\Delta r < 0.01$ Å).

In the next section we showcase an application to a larger molecule to understand the impact of our procedure when considering a more elaborate electronic structure.

5.3. Formaldeyhde (CH$_2$O)

Here we show two examples of geometry optimization carried out with our automatically differentiable QPE on a larger molecular system. The formaldehyde molecule is studied in its ground state and first triplet state ($|T_0\rangle$) which corresponds to the spin-flip transition $^3(n \to \pi^*)$. These systems have been extensively studied both theoretically and experimentally [86–88] as representative systems of carbonyl-chemistry and, more broadly, because of their role in photochemistry.

Throughout the optimization we set the $|HF\rangle$ state as the input state of the QPE. For the active space considered, it is $|HF\rangle = |11000000\rangle$. In table 1 we report the results of our optimization and compare them with a geometry optimization performed at the CASCI level (using the same active space) and with experimental data from [89] as reported in [90].

As we can see the predicted length of the CO double bond is within chemical accuracy ($\Delta r < 0.01$ Å) considering both the theoretical calculations at the CASCI level and the experimental values as reference. Also, considering the predicted angles, an error of $\Delta\theta \approx 2°$ represents a good agreement between theory (both methods) and experiment. Instead, by looking at the predicted CH bond length we can see that there is a more significant discrepancy between the theoretical methods ($\Delta r \approx 0.05$ Å). Such a difference may be ascribed to different factors: (i) bias ensuing from the external parameters T and k, (ii) possible numerical instabilities encountered in the AD pipeline and (iii) differences in the optimization procedure and the underlying Hartree–Fock solver. Noteworthy, the results obtained using the differentiable QPE approach agree better with the experimental values than the other theoretical results. The agreement with the experimental results in [91] is even within the limits of chemical accuracy. This fact can be attributed to error cancellation as the active space is truncated.

We now turn to the discussion of the optimization of the molecule in its first triplet state ($|T_o\rangle$). Here, all along the optimization procedure, we set as initial state the triplet state built upon excitation of the HF determinant, $|\psi\rangle = t_0^2|HF\rangle + t_1^3|HF\rangle$ where t_i^j is a single excitation from spin-orbital i to j. We recall that in our convention for labeling the orbitals, α/β spins are interleaved and that spin-orbital labeling starts with the leftmost qubit being indexed with 0.

As we can see from table 2, bond lengths and angles from QPE are within chemical accuracy of the CASCI reference values, thus further supporting that the GCE estimator coupled with the AD pipeline of PennyLane is able to reproduce both total energies and forces. However, with respect to the experiment (and regardless of the theoretical method employed), the chosen active space causes an error of ≈ 0.02 Å, leading

Table 1. Optimized geometry of formaldehyde in the $|GS\rangle$ electronic state. Distances are expressed in Å, angles in degrees. Experimental errors are reported in brackets.

	CASCI(2e, 4o)	QPE(2e, 4o)	Exp. [89]
r_{CO}	1.212	1.211	1.21(1)
r_{CH}	1.055	1.106	1.12(1)
$\angle_{OCH}$	123.18	122.96	121(1)
$\angle_{OCH}$	123.18	122.96	121(1)
$\angle_{HCH}$	113.64	114.08	—

Table 2. Optimized geometry of formaldehyde in its first triplet electronic state ($|T_0\rangle$). Distances are expressed in Å, angles in degrees.

	CASCI(2e, 4o)	QPE(2e, 4o)	Exp. [88]
r_{CO}	1.320	1.325	1.307
r_{CH}	1.063	1.060	1.084
$\angle_{OCH}$	117.36	119.75	—
$\angle_{OCH}$	116.72	119.74	—
$\angle_{HCH}$	125.92	120.51	121.8

to overestimation of the CO bond length and underestimation of the CH bond length. This is in agreement with other results reported in the literature for calculations in restricted active spaces. For example, Angeli *et al* [87] report errors of 0.062 and 0.028 Å for the CO bond length and of 0.005 and 0.028 Å for the CH bond using two different active spaces and including orbital optimization.

The above results demonstrate not only the effective application of the automatically differentiable QPE to a non trivial chemical system but also emphasize the importance of combined numerical and cost analysis when proposing an algorithm. Indeed, our results highlight the need for specific-domain knowledge and stress the importance of gaining practical experience, as similarly demonstrated in the literature for standard computational quantum chemistry, to obtain results in line with other methods and experiments. In particular, our analysis indicates that the extension of experiments to the optimization of high-energy excited states requires on-the-fly assessments to understand how the displacement of atoms induces changes in the electronic structure and accordingly determine how to modify the input molecular state in the quantum circuit. As a possible future development, we will consider the design of a cheap and fast algorithm to make this evaluation feasible.

6. Conclusions

We showed that a smooth estimator allows the seamless integration of the QPE algorithm within the framework of AD. Initially proposed in [41], the circular estimator was confined to the case when exact eigenstates of an Hamiltonian are known and can be efficiently prepared on a quantum register. Here we overcomed this fundamental issue unlocking its practical use. Along with a new, more accurate estimator for the QPE and the possibility of integrating the algorithm in a new programming environment, here we demonstrate a new off-the-shelf strategy to devise quantum algorithms enabling the simulation of molecular properties.

To this end, we showcased the calculation of optimized geometries outperforming the existing literature both from the point of view of system size (number of atoms) and target states, as we present calculations beyond the singlet ground state geometry optimization. The latter feature is of utmost importance when it comes to the simulation of photochemical experiments. The next, natural step of this work is the application of this strategy to the simulation of spectroscopy experiments.

Despite the potential of the approach reported in this study, some aspects currently limit its scalability to larger systems. In particular, we see that the reconstruction of the measurement distribution requires a substantial computational overhead, which can be mitigated by improving the wave function guess. Further, a straightforward improvement may result from the inclusion of a cosine or Kaiser window in the phase estimation circuit. This concentrates the information in the main lobe of the distribution and may lower the number of relevant bitstrings to sample for the circular average. This solution has already proven to give a cubic speedup for the majority rule estimator [66] and may provide similar speedups for the GCE. Alternatives strategies to circumvent this issues could resort to recent variations of the QPE such as in [73, 92, 93].

It should also be emphasized that when developing a quantum algorithm we are not only concerned with finding a solution to a useful problem but also with ensuring that the runtime of the algorithm is feasible and

advantageous compared to other solutions, including classical ones. Despite the favorable scaling of QPE, the answer to this question in the field of quantum chemistry is still open [37, 39, 40, 94]. For what concerns the algorithm proposed in this work, further research is necessary to arrive at cost-effective differentiable solutions. Our study further underlines the importance of expanding the classes of circuits that allow for backpropagation scaling in derivative calculations [75].

Data availability statement

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

Acknowledgments

Appendix A. The physical meaning of differentiating through the measurement distribution

In this section we derive analytic expressions for the energy derivative as defined by equation (15). These expressions allow us to understand the physical origin of the automatically computed gradient.

The main idea is that we can recast the whole differentiable QPE procedure in this expression:

$$E \approx \mathcal{S}\left[\int \langle \Psi | e^{-iHt} | \Psi \rangle e^{i\omega t} dt\right] = \mathcal{S}\left[I(\omega)\right] \tag{A1}$$

where $\mathcal{S}$ is a differentiable functional corresponding to the smoothing procedure outlined in figure 3 of the main text.

Accordingly, we can write explicitly the energy derivative differentiating the last expression:

$$\frac{\partial \mu}{\partial \mathbf{R}} = \frac{\delta \mathcal{S}}{\delta I} \frac{\partial I}{\partial \mathbf{R}}. \tag{A2}$$

Here, our focus is on the second factor $\frac{\partial I}{\partial \mathbf{R}}$, as the first does not provide any information about the physics of the system but concerns the dependence of our estimator on the parameters denoted as T and k in the main text.

By expanding $|\Psi\rangle$ on the Hamiltonian eigenbasis we get:

$$I(\omega, \mathbf{R}) = \sum_i |c_i|^2 (\mathbf{R}) \delta (\phi_i (\mathbf{R}) - \omega) \tag{A3}$$

where $\phi_i = f(E_i)$. As expressed in the introduction (section 2.1).

Differentiating and accounting for the finite precision of the quantum register grid we obtain:

$$\frac{\partial I}{\partial \mathbf{R}} = \sum_i \frac{\partial |c_i|^2}{\partial \mathbf{R}} \frac{\partial P}{\partial \phi_i} \frac{\partial \phi_i}{\partial \mathbf{R}} \tag{A4}$$

where P is the parent distribution kernel defined in equation (4) and i is a label running over the electronic eigenstates.

In our approximation, i.e. equation (14), the last equation reduces to:

$$\frac{\partial I}{\partial \mathbf{R}} \approx \frac{\partial |c_i|^2}{\partial \mathbf{R}} \frac{\partial P}{\partial \phi_i} \frac{\partial \phi_i}{\partial \mathbf{R}} \tag{A5}$$

where $|i\rangle$ is the only eigenstate that contributes to the probability P.

We again focus on the terms that bring the molecular physics into the algorithm. Therefore we do not comment the term $\frac{\partial P}{\partial \phi_i}$ which can be obtained analytically from equation (4).

By applying the chain rule to the eigenphase derivative we can easily see the emergence of the true molecular energy derivatives: $\frac{\partial \phi_i}{\partial \mathbf{R}} = \frac{\partial \phi_i}{\partial f} \frac{\partial f}{\partial E_i} \frac{dE_i}{d\mathbf{R}}$. The dependence on f is a purely algorithmic choice and will not be considered here. Instead, following [42] we directly relate the force calculated with respect to the ith eigenstate to molecular integrals derivatives as:

$$\frac{dE_i}{d\mathbf{R}} = \left(\sum_{pq} \gamma_{pq} \frac{dh_{pq}}{d\mathbf{R}} + \sum_{pqrs} \Gamma_{pqrs} \frac{dg_{pqrs}}{d\mathbf{R}} \right) - \left(\sum_{pqm} \gamma_{pq} h_{mq} \frac{dS_{mp}}{d\mathbf{R}} + 2 \sum_{pqrst} \Gamma_{pqrs} g_{tqrs} \frac{dS_{tp}}{d\mathbf{R}} \right) \tag{A6}$$

where γ_{pq} and Γ_{pqrs} are the one- and two-electrons reduced density matrices [95] for the ith eigenstate respectively, h and g are the one- and two-electrons integrals representing all the interactions in the molecular Hamiltonian (i.e. electron-nuclear interactions and electron-electron interactions). All indexes of the summation run over the spin-orbitals. S is the basis overlap matrix.

To obtain a complete expression for the energy derivative, we evaluate $\frac{\partial |c_u|^2}{\partial \mathbf{R}}$. We consider a real-valued wavefunction (without any loss of generality as long as we are not neglecting the presence of external magnetic fields) for which we can write the first-order response of the wavefunction with respect to a perturbation (in our case the perturbation is a nuclear displacement) as [96]:

$$\frac{\partial c_\mu}{\partial \mathbf{R}} = -\sum_\nu \mathcal{G}_{\mu\nu}^{-1} \frac{\partial \mathcal{F}_\nu}{\partial \mathbf{R}} \tag{A7}$$

where $\mathcal{G}$ and $\mathcal{F}$ are the electronic Hessian and the electronic gradient, respectively, defined as:

$$\mathcal{G}_{\mu\nu} = \frac{\partial^2 E}{\partial c_\mu \partial c_\nu} = 2 \left(\langle \mu|H|\nu \rangle - E_{GS} \langle \mu|\nu \rangle \right) \tag{A8}$$

$$\mathcal{F}_\mu = \frac{\partial E}{\partial c_\mu} = 2 \langle \mu|PH|\Psi \rangle. \tag{A9}$$

The previous equations hold for the derivatives of the ground state coefficients. Moreover, we point out that $|\mu\rangle$ and $|\nu\rangle$ are N-electrons Slater determinants. The expressions in the Hamiltonian eigenbasis can be recovered considering that the two sets of coefficients are related by the basis change $\mathbf{C}_{\{|u\rangle\}} = \mathbf{V}^\dagger \mathbf{C}_{\{|\mu\rangle\}} \mathbf{V}$ (where $\mathbf{V}$ is the matrix that diagonalizes the Hamiltonian). Finally, in equation (A9) we have introduced the projection operator $P = \mathbf{1} - |\Psi\rangle\langle\Psi|$.

Using equation (A7) and $\frac{\partial |c_u|^2}{\partial \mathbf{R}} = 2c_u \frac{\partial c_u}{\partial \mathbf{R}}$, we can attribute a precise physical meaning to the overlap derivative term: in analogy with *Hooke's law*, the wavefunction relaxation is proportional to the 'force' $-\frac{\partial \mathcal{F}_\nu}{\partial \mathbf{R}}$ and inversely proportional to the 'force constant' $\mathcal{G}$. Indeed, in the absence of external perturbation (i.e. nuclear displacement), the electronic gradient is null and the wavefunction is stable.
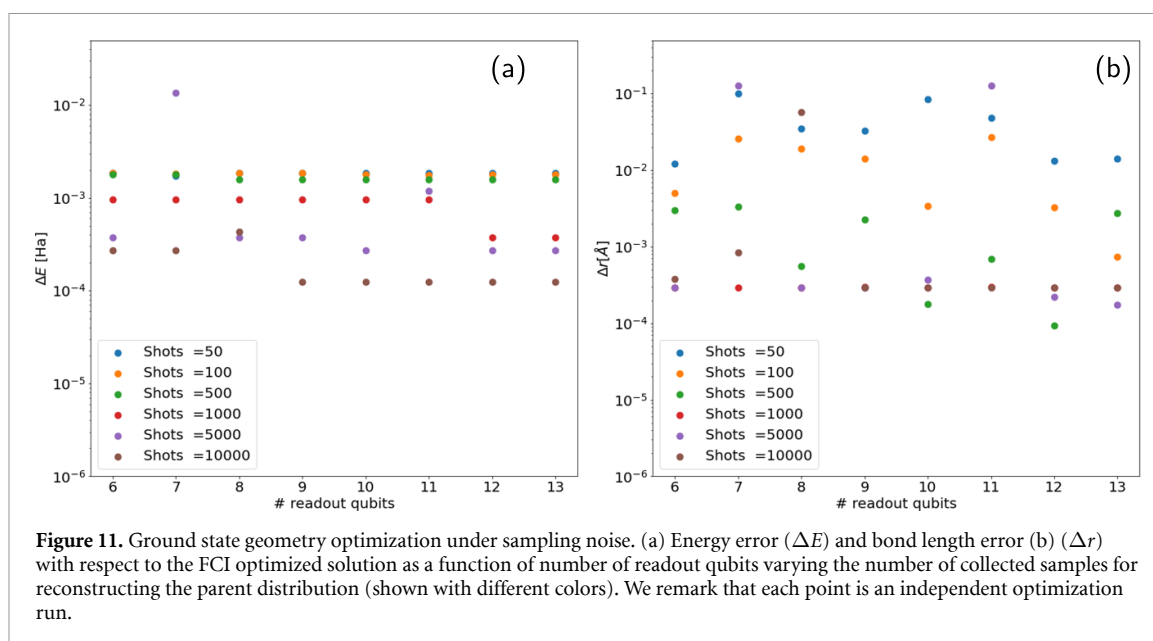
Finally, adding the contribution of equation (A6), we can reconstruct the meaning of equation (A5). In essence, equation (A5) tells us that the measurement distribution induced by the QPE circuit varies as nuclei are displaced due to the variation of the electronic energy (equation (A6)) and the wavefunction coefficients (equation (A7)). These contributions are singled out and modulated by the functional $\mathcal{S}$. Interestingly, if we employed in this context a straight-through estimator (i.e. equating incoming and outcoming gradients of the $\mathcal{S}$ node), we could not distinguish between the contributions of the different eigenstates that enter the variations of the probability distribution in different geometries. Finally, we remark that also other algorithmic choices (e.g. $f \propto \cos^{-1}$ when employing a qubitization [97] based QPE), and the number of readout qubits entering the quantity $\frac{\partial P}{\partial \phi_i}$ play a key role in obtaining the exact gradient.

Appendix B. Effect of sampling noise

In figure 11, we illustrate the effect of sampling on the geometry optimization of H_3^+ in its singlet ground state using the strategy defined in the main text, see sections 2 and 3.

Both in terms of energy error (ΔE, figure 11(a)) and bond length error (Δr, figure 11(b)), the estimate of the number of samples to be measured outlined in figure 7 is consistent with the findings reported here. Indeed, the initial state used to perform the geometry optimization is always the Hartree–Fock determinant, which, starting from the initial geometry (the same one used for the calculations presented in the main text), exhibits an overlap with the FCI solution greater than 90%.

Interestingly, by looking at figure 11, we may be tempted to conclude that the GCE variance does not actually depend on the number of readout qubits, as very similar trends are observed with respect to the number of shots regardless of the number of readout qubits; also, the final errors with respect to the exact geometry look quite independent on t. As we have discussed in section 3.1, this is not the case. The results

Figure 11. Ground state geometry optimization under sampling noise. (a) Energy error (ΔE) and bond length error (b) (Δr) with respect to the FCI optimized solution as a function of number of readout qubits varying the number of collected samples for reconstructing the parent distribution (shown with different colors). We remark that each point is an independent optimization run.

shown in figure 11 arise from the interplay of two effects: (i) the window sizes that we have used in these calculations shrink exponentially with the number of qubits, as reported in figure 4. This implies that the number of bitstrings included in the circular average (for the calculations reported in figure 11) is always the same regardless of the number of readout qubits. Clearly, this approach works also with a lower number of qubits because the input state used allows sampling from a large number of bitstrings without adding information from spurious excited states. (ii) These results do not refer to the a single-point energy evaluation but are obtained from different independent optimization processes. Therefore the errors reported here combine the errors on the function and gradient evaluations with those in the optimization process.

ORCID iDs

Davide Castaldo ⓘ https://orcid.org/0000-0001-8622-175X
Soran Jahangiri ⓘ https://orcid.org/0000-0002-9988-8841
Agostino Migliore ⓘ https://orcid.org/0000-0001-7780-2296
Juan Miguel Arrazola ⓘ https://orcid.org/0000-0002-0619-9650
Stefano Corni ⓘ https://orcid.org/0000-0001-6707-108X

References

[1] Talia D, Trunfio P, Carretero J and Garcia-Blas J 2022 *Front. Big Data* **5** 4
[2] Mennucci B and Corni S 2019 *Nat. Rev. Chem.* **3** 315
[3] Ciccotti G, Dellago C, Ferrario M, Hernández E and Tuckerman M 2022 *Eur. Phys. J. B* **95** 3
[4] Chang C, Deringer V L, Katti K S, Van Speybroeck V and Wolverton C M 2023 *Nat. Rev. Mater.* **8** 309–13
[5] Di Felice R *et al* 2023 *J. Chem. Theory Comput.* **19** 7056
[6] Ufimtsev I S and Martinez T J 2008 *J. Chem. Theory Comput.* **4** 222
[7] Liu F, Luehr N, Kulik H J and Martínez T J 2015 *J. Chem. Theory Comput.* **11** 3131
[8] Seritan S *et al* 2021 *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **11** e1494
[9] Pederson R *et al* 2023 *J. Chem. Theory Comput.* **19** 25–32
[10] Almlöf J and Taylor P R 1985 *Int. J. Quantum Chem.* **27** 743
[11] Martis E, Radhakrishnan R and Badve R 2011 *J. Appl. Pharm. Sci.* **1** 2–10
[12] Luo S, Li T, Wang X, Faizan M and Zhang L 2021 *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **11** e1489
[13] Koch H, Jensen H J A, Jo/rgensen P, Helgaker T, Scuseria G E and Schaefer III H F 1990 *J. Chem. Phys.* **92** 4924
[14] Assaraf R and Caffarel M 2000 *J. Chem. Phys.* **113** 4028
[15] Wang S, Li C and Evangelista F A 2021 *J. Chem. Theory Comput.* **17** 7666
[16] Hagan M and Wiebe N 2023 *Quantum* **7** 1181
[17] Low G H, Su Y, Tong Y and Tran M C 2023 *PRX Quantum* **4** 020323
[18] Babbush R, Huggins W J, Berry D W, Ung S F, Zhao A, Reichman D R, Neven H, Baczewski A D J L and Lee J 2023 *Nat. Commun.* **14** 4058
[19] Kang C, Bauman N P, Krishnamoorthy S and Kowalski K 2022 *J. Chem. Theory Comput.* **18** 6567
[20] Steudtner M *et al* 2023 arXiv:2303.14118
[21] Baydin A G, Pearlmutter B A, Radul A A and Siskind J M 2018 *J. Mach. Learn. Res.* **18** 1
[22] Battaglia S 2023 *Quantum Chemistry in the Age of Machine Learning* (Elsevier) pp 577–616
[23] Kasim M F, Lehtola S and Vinko S M 2022 *J. Chem. Phys.* **156** 084801

[24] Tamayo-Mendoza T, Kreisbeck C, Lindh R and Aspuru-Guzik A 2018 *ACS Cent. Sci.* **4** 559
[25] Bartlett R A, Gay D M and Phipps E T 2006 *Computational Science–ICCS 2006: 6th Int. Conf., (Reading, UK, 28–31 May 2006, Proc., Part IV)* (Springer) pp 525–32
[26] Frank S A 2022 *Front. Ecol. Evol.* **10** 1010278
[27] Xu W, Chen X and Coleman T F 2016 *J. Comput. Finance* **19** 71–96
[28] Di Matteo O and Woloshyn R 2022 *Phys. Rev.* A **106** 052429
[29] Cerezo M *et al* 2021 *Nat. Rev. Phys.* **3** 625
[30] Zhu S, Hung S-H, Chakrabarti S and Wu X 2020 *Proc. 41st ACM SIGPLAN Conf. on Programming Language Design and Implementation* pp 272–85
[31] Fang W, Ying M and Wu X 2022 arXiv:2211.04507
[32] Matteo O D, Izaac J, Bromley T R, Hayes A, Lee C, Schuld M, Száva A, Roberts C and Killoran N 2023 *ACM Trans. Quantum Comput.* **4** 1
[33] Bergholm V *et al* 2018 arXiv:1811.04968
[34] Broughton M *et al* 2020 arXiv:2003.02989
[35] Luo X-Z, Liu J-G, Zhang P and Wang L 2020 *Quantum* **4** 341
[36] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge University Press)
[37] Lee S *et al* 2023 *Nat. Commun.* **14** 1952
[38] Fomichev S *et al* 2023 arXiv:2310.18410
[39] Rubin N C *et al* 2023 *PRX Quantum* **4** 040303
[40] Zini M S, Delgado A, dos Reis R, Casares P A M, Mueller J E, Voigt A-C and Arrazola J M 2023 *Quantum* **7** 1049
[41] Cruz P M, Catarina G, Gautier R and Fernández-Rossier J 2020 *Quantum Sci. Technol.* **5** 044005
[42] O'Brien T E *et al* 2022 *Phys. Rev. Res.* **4** 043210
[43] Cleve R, Ekert A, Macchiavello C and Mosca M 1998 *Proc. R. Soc.* A **454** 339
[44] Ekert A and Jozsa R 1996 *Rev. Mod. Phys.* **68** 733
[45] Rumelhart D E, Hinton G E and Williams R J 1986 *Nature* **323** 533
[46] Arrazola J M *et al* 2021 arXiv:2111.09967
[47] Li J, Yang X, Peng X and Sun C-P 2017 *Phys. Rev. Lett.* **118** 150503
[48] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev.* A **98** 032309
[49] Schuld M, Bergholm V, Gogolin C, Izaac J and Killoran N 2019 *Phys. Rev.* A **99** 032331
[50] Izmaylov A F, Lang R A and Yen T-C 2021 *Phys. Rev.* A **104** 062443
[51] Banchi L and Crooks G E 2021 *Quantum* **5** 386
[52] Kyriienko O and Elfving V E 2021 *Phys. Rev.* A **104** 052417
[53] Wierichs D, Izaac J, Wang C and Lin C Y-Y 2022 *Quantum* **6** 677
[54] Wiersema R, Lewis D, Wierichs D, Carrasquilla J and Killoran N 2023 arXiv:2303.11355
[55] Bengio Y, Léonard N and Courville A 2013 arXiv:1308.3432
[56] Jang E, Gu S and Poole B 2017 Categorical reparameterization with gumbel-softmax (arXiv:1611.01144 [stat.ML])
[57] Mardia K V and Jupp P E 2009 *Directional Statistics* (Wiley)
[58] Angeli C, Cimiraglia R, Dallo F, Guareschi R and Tenti L 2013 *J. Chem. Educ.* **90** 1639
[59] Zheng A and Casari A 2018 *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists* (O'Reilly Media, Inc.)
[60] Moore A J, Wang Y, Hu Z, Kais S and Weiner A M 2021 *New J. Phys.* **23** 113027
[61] Eriksson J, Ollila E and Koivunen V 2009 *2009 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* pp 3565–8
[62] Christiansen O 2004 *J. Chem. Phys.* **120** 2140
[63] Ollitrault P J, Baiardi A, Reiher M and Tavernelli I 2020 *Chem. Sci.* **11** 6842
[64] Baiardi A, Stein C J, Barone V and Reiher M 2017 *J. Chem. Theory Comput.* **13** 3764
[65] Yurchenko S N, Thiel W and Jensen P 2007 *J. Mol. Spectrosc.* **245** 126
[66] Rendon G, Izubuchi T and Kikuchi Y 2022 *Phys. Rev.* D **106** 034503
[67] Feynman R P 1939 *Phys. Rev.* **56** 340
[68] Knill E, Ortiz G and Somma R D 2007 *Phys. Rev.* A **75** 012328
[69] Gilyén A, Arunachalam S and Wiebe N 2019 *Proc. 30th Annual ACM-SIAM Symp. on Discrete Algorithms* (SIAM) pp 1425–44
[70] Huggins W J, Wan K, McClean J, O'Brien T E, Wiebe N and Babbush R 2022 *Phys. Rev. Lett.* **129** 240501
[71] Sugisaki K, Wakimoto H, Toyota K, Sato K, Shiomi D and Takui T 2022 *J. Phys. Chem. Lett.* **13** 11105
[72] O'Brien T E, Senjean B, Sagastizabal R, Bonet-Monroig X, Dutkiewicz A, Buda F, DiCarlo L and Visscher L 2019 *npj Quantum Inf.* **5** 113
[73] Yi C, Zhou C and Takahashi J 2023 arXiv:2306.07008
[74] Brassard G, Hoyer P, Mosca M and Tapp A 2002 *Contemp. Math.* **305** 53
[75] Bowles J, Wierichs D and Park C-Y 2023 arXiv:2306.14962
[76] Davide C 2024 (Available at: https://github.com/davidecast/DiffQPE) (Accessed 01 July 2023)
[77] Nocedal J and Wright S J 1999 *Numerical Optimization* (Springer)
[78] Turney J M *et al* 2012 *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2** 556
[79] Sun Q *et al* 2018 *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **8** e1340
[80] Jordan P, von Neumann J and Wigner E P 1993 *The Collected Works of Eugene Paul Wigner: Part A: The Scientific Papers* (Springer) pp 298–333
[81] Tennyson J 1995 *Rep. Prog. Phys.* **58** 421
[82] Polyansky O L, Alijah A, Zobov N F, Mizus I I, Ovsyannikov R I, Tennyson J, Lodi L, Szidarovszky T and Csaszar A G 2012 *Phil. Trans. R. Soc.* A **370** 5014
[83] McCall B J and Oka T 2000 *Science* **287** 1941
[84] Aguado A, Roncero O, tablero C, Sanz C and Paniagua M 2000 *J. Chem. Phys.* **112** 1240
[85] Sanz C, Roncero O, tablero C, Aguado A and Paniagua M 2001 *J. Chem. Phys.* **114** 2182
[86] Krylov A I, Sherrill C D and Head-Gordon M 2000 *J. Chem. Phys.* **113** 6509
[87] Angeli C, Borini S, Ferrighi L and Cimiraglia R 2005 *J. Mol. Struct.* **718** 55
[88] Clouthier D and Ramsay D 1983 *Annu. Rev. Phys. Chem.* **34** 31
[89] Chase Jr M W, Davies JR C A, Downey Jr J R, Frurip D J, McDonald R A and Syverud A N 1985 *J. Phys. Chem. Ref. Data* **14** 94

[90] Schuetz C, Frenklach M, Kollias A and Lester W Jr 2003 *J. Chem. Phys.* **119** 9386
[91] Takagi K and Oka T 1963 *J. Phys. Soc. Japan* **18** 1174
[92] Ni H, Li H and Ying L 2023 arXiv:2302.02454
[93] Ding Z and Lin L 2023 *PRX Quantum* **4** 020331
[94] Delgado A *et al* 2022 *Phys. Rev.* A **106** 032428
[95] Helgaker T, Jorgensen P and Olsen J 2013 *Molecular Electronic-Structure Theory* (Wiley)
[96] Helgaker T and Jørgensen P 1988 *Adv. Quantum Chem.* **19** 183
[97] Low G H and Chuang I L 2019 *Quantum* **3** 163