



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

## **Accelerating variational quantum eigensolver convergence using parameter transfer**

Downloaded from: <https://research.chalmers.se>, 2023-09-08 04:48 UTC

Citation for the original published paper (version of record):

Skogh, M., Leinonen, O., Lolur, P. et al (2023). Accelerating variational quantum eigensolver convergence using parameter transfer. *Electronic Structure*, 5(3).  
<http://dx.doi.org/10.1088/2516-1075/ace86d>

N.B. When citing this work, cite the original published paper.

PAPER • OPEN ACCESS

## Accelerating variational quantum eigensolver convergence using parameter transfer

To cite this article: Mårten Skogh *et al* 2023 *Electron. Struct.* **5** 035002

View the [article online](#) for updates and enhancements.

### You may also like

- [Filtering variational quantum algorithms for combinatorial optimization](#)  
David Amaro, Carlo Modica, Matthias Rosenkranz et al.
- [Large gradients via correlation in random parameterized quantum circuits](#)  
Tyler Volkoff and Patrick J Coles
- [Understanding the role of oxygen-vacancy defects in Cu<sub>2</sub>O\(111\) from first-principle calculations](#)  
Nanchen Dongfang, Yasmine S Al-Hamdani and Marcella Iannuzzi

# Electronic Structure



## PAPER

# Accelerating variational quantum eigensolver convergence using parameter transfer

### OPEN ACCESS

#### RECEIVED

29 March 2023

#### REVISED

9 July 2023

#### ACCEPTED FOR PUBLICATION

18 July 2023

#### PUBLISHED

4 August 2023

Mårten Skogh<sup>1,2</sup> , Oskar Leinonen<sup>1,3</sup> , Phalgun Lolur<sup>1,4</sup> and Martin Rahm<sup>1,\*</sup> <sup>1</sup> Department of Chemistry and Chemical Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden<sup>2</sup> Data Science & Modelling, Pharmaceutical Science, R&D, AstraZeneca, Gothenburg, Sweden<sup>3</sup> Present address: Department of Physics, University of Oslo, N-0316 Oslo, Norway<sup>4</sup> Present address: National Quantum Computing Centre (NQCC), Rutherford Appleton Laboratory, Harwell Campus, Didcot, Oxfordshire, OX11 0QX, United Kingdom

\* Author to whom any correspondence should be addressed.

E-mail: [martin.rahm@chalmers.se](mailto:martin.rahm@chalmers.se)**Keywords:** quantum computation, potential energy surfaces, quantum chemistrySupplementary material for this article is available [online](#)

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



## Abstract

One impediment to the useful application of variational quantum algorithms in quantum chemistry is slow convergence with large numbers of classical optimization parameters. In this work, we evaluate a quantum computational warm-start approach for potential energy surface calculations. Our approach, which is inspired by conventional computational methods, is evaluated using simulations of the variational quantum eigensolver. Significant speedup is demonstrated relative to calculations that rely on a Hartree–Fock initial state, both for ideal and sampled simulations. The general approach of transferring parameters between similar problems is promising for accelerating current and near-term quantum chemistry calculations on quantum hardware, and is likely applicable beyond the tested algorithm and use case.

## 1. Introduction

Quantum computing has been proposed as a possible next step in the field of computational chemistry, especially when applied to highly correlated systems [1]. Theoretical proof of potential quantum speedup exists for problems involving highly correlated systems [2, 3]. However, these results rely on fault-tolerant quantum computers. While quantum computing has come a long way, fault-tolerant computing is yet to be realized [4–6]. While the technology matures, other avenues to useful quantum computing are being explored. The period leading up to fault-tolerant quantum computing is often referred to as the Noisy Intermediate-Scale Quantum (NISQ) era [7]. Quantum devices in this regime are susceptible to many types of errors and noise [8, 9].

Despite these impediments, algorithms have been designed to leverage existing NISQ hardware for quantum chemistry calculations [1, 3, 10]. Several such approaches rely on so-called Variational Quantum Algorithms (VQAs) [11]. These variational algorithms, a subset of which will be described in more detail below, utilize a parameterized quantum circuit to generate trial states for which a cost (or loss) can be evaluated through measurement on the NISQ device. By coupling the quantum computer to a classical optimization algorithm, the circuit parameters can be optimized such that the cost is minimized. Even though the addition of a classical optimization step allows for some level of noise suppression [11, 12], several works [9, 12–15] have highlighted the challenges of optimizing the high dimensional parameter spaces associated with these types of quantum circuits. While there is no formal proof that VQAs can out-perform classical implementations, they are nonetheless useful in pushing the limits of near-term quantum computation [16, 17].

For chemistry applications, the (currently) most common algorithm is the Variational Quantum Eigensolver (VQE), which has been used to perform several quantum chemistry calculations on NISQ hardware [18–22]. The VQE algorithm is briefly summarized below, and in more detail in, e.g. [3, 19, 23, 24].

The number of parameters and energy evaluations required for VQE calculations can reach into the tens or hundreds of thousands, even for small molecules [12]. The large number of parameters makes calculations slow, and in some instances practically unfeasible. For example, more than 16 000 variational parameters and 69 qubits are required to describe a single hydrogen cyanide molecule using unitary coupled cluster theory including single and double excitations (UCCSD) and a standard  $6-31+G(d)$  basis set. More advanced versions of the VQE algorithm exist that reduce the number of parameters. However, the fundamental challenge of high dimensional optimization remains as the size of the simulated system increases.

In this work, we outline and evaluate a general strategy for accelerating convergence of VQE-type [23–25] algorithms through the use of a parameter transfer (PT) approach. We provide proof-of-concept calculations that demonstrate speedup of VQE-based evaluation of potential energy surfaces (PESs) of molecules, both for ideal simulations and in the presence of sampling noise.

## 2. The VQE algorithm

The objective of any VQA is to find the parameters  $\theta^{\min}$  that minimize the value of some cost (loss) function  $C(\theta)$ . How one defines the cost function varies. Nevertheless, some aspects of the cost function are common to all implementations. The type of cost functions that we will discuss in this work can be seen as functions of (1) a parameterized circuit,  $U(\theta)$ , and (2) a set of measurable operators (observables),  $\{\hat{O}_k\}$ , that sums to the cost.

The VQE is a particular implementation of a VQA that aims to identify a set of parameters that minimize the expectation value of a Hamiltonian [26]. The VQE algorithm can consequently be formulated as a minimization problem

$$\min_{\theta} E(\theta) = \min_{\theta} \langle \Phi(\theta) | \hat{H} | \Phi(\theta) \rangle, \quad (1)$$

where the cost is the energy  $E(\theta)$ . Because we rely on the Born–Oppenheimer approximation, energy minimization is performed for fixed nuclear coordinates  $\mathbf{R}$  with the corresponding Hamiltonian  $\hat{H}(\mathbf{R})$ . When formulated in second quantization this electronic Hamiltonian reads as,

$$\hat{H} = E_{NN} + \sum h_{ij} \hat{a}_i^+ \hat{a}_j + \frac{1}{2} \sum h_{ijkl} \hat{a}_i^+ \hat{a}_j^+ \hat{a}_k \hat{a}_l. \quad (2)$$

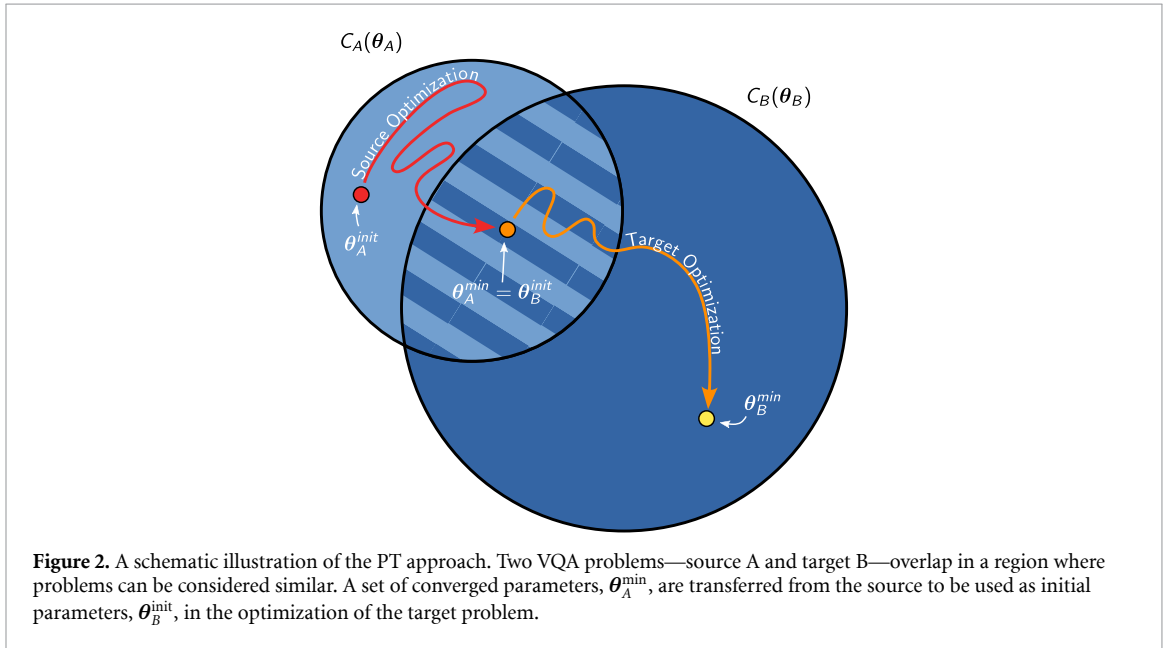
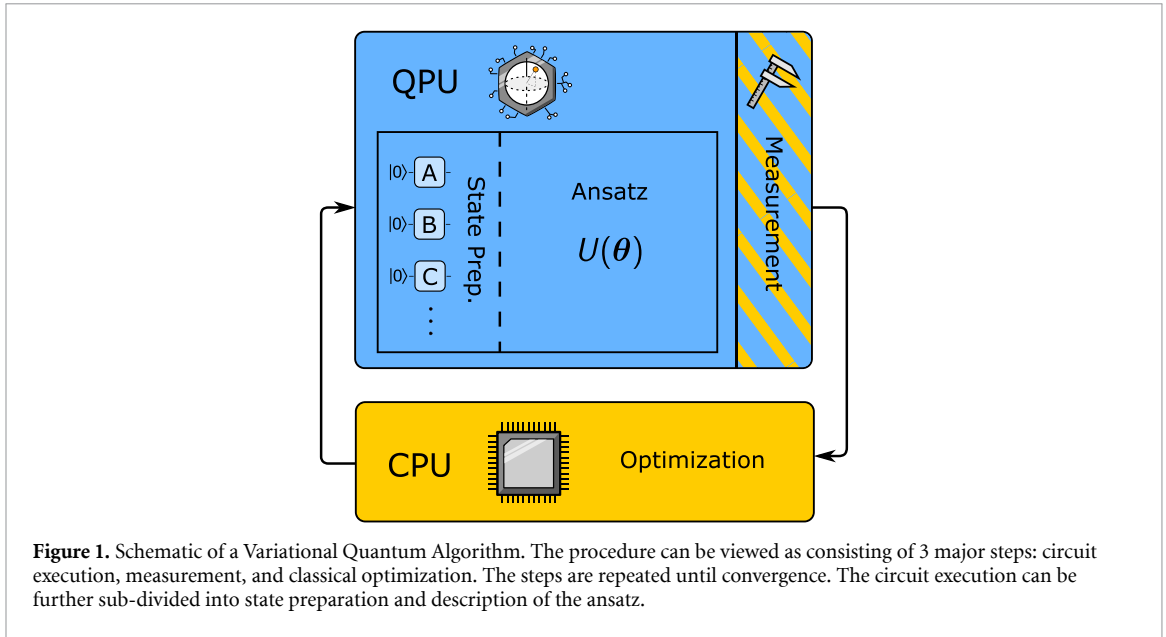
In equation (2),  $h_{ij}$  and  $h_{ijkl}$  are one- and two-electron integrals, and  $E_{NN}$  is the classical nuclear-nuclear repulsion energy,  $\hat{a}_m^+$  and  $\hat{a}_m$  are the fermionic creation and annihilation operators for the spin orbital  $m$  [27, 28]. As we simulate physical systems, the energy minimization of equation (1) is commonly subjected to constraints, such as particle and spin conservation.

The implementation of VQE can be broken down into three main components: quantum circuit execution, measurement, and classical optimization of parameters with respect to the expectation value (figure 1). The quantum circuit itself can be represented by two unitary operators: the state preparation,  $U_{\text{prep}}$ , and the ansatz,  $U(\theta)$ . The state preparation circuit acts on the all-zero state to create some initial state  $|\Phi_{\text{init}}\rangle = U_{\text{prep}}|0 \dots 0\rangle$ . This initial state is then modified through the application of the ansatz to produce a trial state  $|\Phi(\theta)\rangle$ , that is  $U(\theta)|\Phi_{\text{init}}\rangle = |\Phi(\theta)\rangle$ . Measurements of relevant Pauli operators are subsequently used to reconstruct the expectation value that, in turn, is used to guess a new set of parameters to be used in the ansatz. This process is repeated until some convergence criterion is met. While the VQE algorithm may appear straightforward, each step presents challenges that are critical for the quality of the calculation. A more detailed description of how the electronic structure problem is mapped to a quantum circuit ansatz is provided in the supporting information (SI).

A crucial part of the VQE algorithm, which can make or break any attempt at convergence, is providing a suitable initial guess (state),  $|\Phi_{\text{init}}\rangle$ , for the wavefunction (*state preparation* in figure 1). In this work, we explore the general approach of accelerating convergence by transferring parameters from a previously converged calculation or otherwise known solution.

## 3. Parameter Transfer (PT)

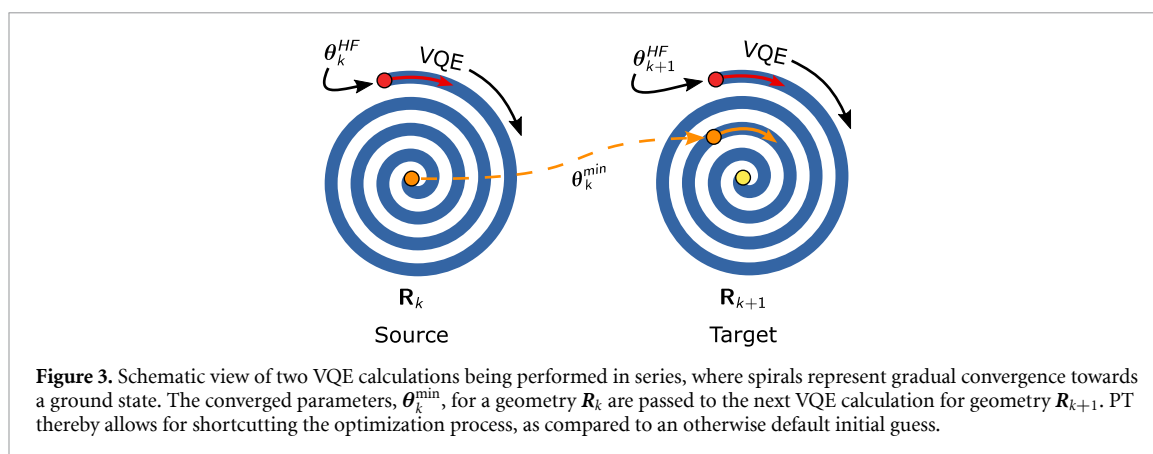
Central to the idea of PT is the existence of an underlying similarity between two problems. In many cases, some form of commonality is apparent. For example, in chemistry, molecular structures might differ by a minor chemical substitution, or be related by symmetry or conformation. An important class of problems also related to similarity is quantum mechanical problems treated at different levels of theory or accuracy. The level of theory might, for instance, be increased, by utilizing larger basis sets, more flexible wavefunctions, or by limiting approximations in the Hamiltonian. Others, including one of us (MS), have shown that PT can be used to accelerate post-Born–Oppenheimer VQE calculations, by relying on related



Born–Oppenheimer approximation-based calculations [29]. In what follows, we will first briefly discuss PT in the context of a general cost function,  $C(\theta)$ . We will then address the case of molecular ground state energies (cf  $E(\theta)$  in equation (1)).

Providing a general definition or measure of similarity is outside the scope of this work. We here view similarity in terms of relations between the shape of the landscapes of two cost functions,  $C_A(\theta_A)$  and  $C_B(\theta_B)$ . Regions around minima of cost functions,  $\theta_A^{\min}$  and  $\theta_B^{\min}$ , are considered similar if the two minima offer approximate solutions to the other, i.e. if  $C_B(\theta_A^{\min}) \approx C_B(\theta_B^{\min})$  and  $C_A(\theta_B^{\min}) \approx C_A(\theta_A^{\min})$ . Note that the two cost functions need not have similar values, nor the same number of parameters. To understand this seeming dichotomy, we can consider a simple example: the separate minimizations in  $x$  of  $y = x^2$  and  $y = x^4 + c$ . Both problems share the same solution in  $x^{\min} = 0$  while differing in cost at an arbitrary value  $x$ , since  $c$  is unbound. For the purpose of optimization, it is important that the two convex regions surrounding each minimum overlap (figure 2). Given that an overlap exist it is possible to start at  $\theta_A^{\min}$  and follow the steepest decent to reach  $\theta_B^{\min}$ .

In the optimization problems considered herein, there is an underlying order to the calculations; one is performed before the other. We use *source* to mean an initial calculation. *Target*, in contrast, refers to the calculation to which the parameters are transferred and that we wish to accelerate.



The underlying principle of PT is common practice in conventional quantum chemistry. For example, PT can be useful for speeding up convergence of self-consistent field optimization in PES sampling, either by transferring parameters between wavefunctions of neighboring geometries or different levels of theory. PES, in turn, form the basis for evaluating equilibria and transition state geometries, reaction rates, spectroscopic constants, and other properties of molecules. PT is also related to the field of transfer learning in machine learning [30, 31]. The implementation of PT in quantum computation has thus far been limited. Promising speedup for the Quantum Approximate Optimization Algorithm [32–37] and for VQA models [38] have been demonstrated with similar approaches. IBM has implemented extrapolation (or bootstrapping) methods in their Qiskit framework for the purpose of accelerating PES sampling of molecules [39]. However, while some methods are available, testing and characterization of these methods is lacking. Overall, little is known about the efficiency of PT when applied to the VQE, or in the presence of noise.

Here, we combine PT with VQE and apply it to the problem of PES sampling of molecules. Figure 3 illustrates how PT connects, and ideally accelerates, consecutive variational VQE optimizations on a PES, by providing improved initial guesses. We emphasize that this combination of PT with VQE is but a small first step towards extending PT to quantum computation of chemistry more generally. Some of us (MS) have already built upon this first step and explore PT between levels of theory in another work [29].

The initial geometry,  $\mathbf{R}_0$ , in a PT-VQE PES evaluation is, naturally, performed without any transfer (as there is no previous solution to transfer from). The subsequent calculation of a neighboring geometry,  $\mathbf{R}_1$ , is then initiated using parameters  $\theta_0^{\min}$ , optimized for  $\mathbf{R}_0$ . Extending to the general case of neighboring geometries  $\mathbf{R}_k$ , and  $\mathbf{R}_{k+1}$ , we utilize a converged source ground state,  $|\Phi(\theta_{\mathbf{R}_k}^{\min})\rangle$ , as the initial guess for the target ground state  $|\Phi(\theta_{\mathbf{R}_{k+1}}^{\min})\rangle$ . The difference in geometry (the *transfer step*) is then  $\mathbf{d} = \mathbf{R}_{k+1} - \mathbf{R}_k$  and the *transfer distance* is  $d = \|\mathbf{R}_{k+1} - \mathbf{R}_k\|$ . Here we intentionally leave the norm unspecified for the general case since what constitutes *distance* may vary between systems. We will apply PT to accelerate exploration of relatively simple one-dimensional PES sampling using VQE. For this reason, we refer to Euclidean distances (2-norm).

#### 4. Quantifying speedup

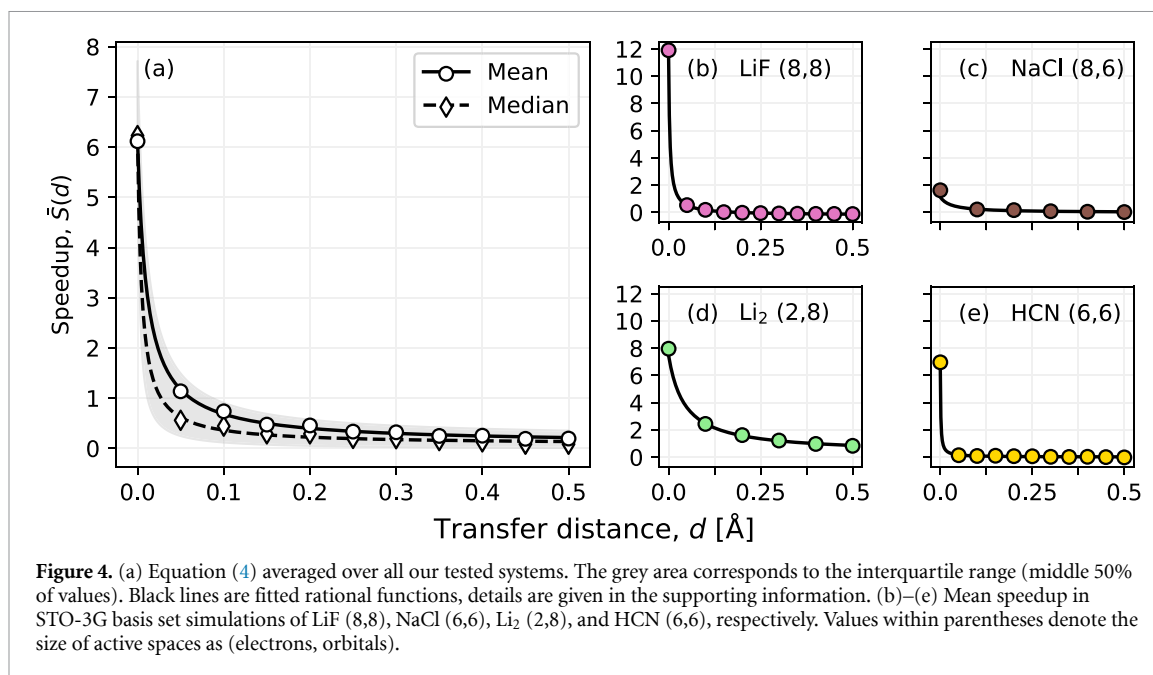
To quantify the efficacy of PT-accelerated VQE, we compare it to the de facto standard method for chemically inspired ansätze: Hartree–Fock (HF), a mean-field description of electron interactions [28]. We define *speedup* for a given target geometry  $\mathbf{R}$  and transfer step  $\mathbf{d}$  as

$$S = S(\mathbf{R}, \mathbf{d}) = \frac{N_{\text{HF}}(\mathbf{R}, \mathbf{d})}{N_{\text{PT}}(\mathbf{R}, \mathbf{d})} - 1, \quad (3)$$

where,  $N_{\text{HF}}$  and  $N_{\text{PT}}$  are the number of times energy expectation values,  $\langle E \rangle$ , needs to be evaluated for convergence to be reached with a HF and a PT initial guess, respectively. In our following evaluation of PT, we will additionally present speedup averaged over all  $n$  considered combinations  $(\mathbf{R}, \mathbf{d})$  where  $\|\mathbf{d}\| = d$ ,

$$\bar{S} = \bar{S}(d) = \frac{1}{n} \sum_{(\mathbf{R}, \mathbf{d})} S(\mathbf{R}, \mathbf{d}) . \quad (4)$$

Some optimization algorithms require additional evaluations as part of a setup routine, as discussed in the supporting information.



**Figure 4.** (a) Equation (4) averaged over all our tested systems. The grey area corresponds to the interquartile range (middle 50% of values). Black lines are fitted rational functions, details are given in the supporting information. (b)–(e) Mean speedup in STO-3G basis set simulations of LiF (8,8), NaCl (6,6), Li<sub>2</sub> (2,8), and HCN (6,6), respectively. Values within parentheses denote the size of active spaces as (electrons, orbitals).

## 5. Methods

All simulations were performed using the VQE algorithm implemented in Qiskit 0.39.3 [39] combined with Gaussian 16, version B.01-AVX2 [40]. All calculations relied on Unitary Coupled Cluster Singles and Doubles (UCCSD) theory, while basis sets and active spaces were varied as indicated in figure 4. We denote the number of active orbitals and electrons using the notation (electrons, orbitals). Active spaces were determined using Qiskit’s *ActiveSpaceTransformer*, limiting the active space to those orbitals closest to the Fermi level. Molecular Hamiltonians were mapped to simulated qubits using parity encoding. All simulations implemented two-qubit reduction, made possible by considering alpha and beta spin parity conservation [41]. The resulting quantum circuits thus range in size from 2 to 14 qubits in the noise-free simulations (details are provided in the SI).

We relied on Qiskit’s Sequential Least Squares Programming (SLSQP) [42] optimization algorithm for exact simulations. Constrained Optimization By Linear Approximation (COBYLA) [43] and the Simultaneous Perturbation Stochastic Approximation (SPSA) [44] optimization algorithms were both evaluated for sampling-based simulations. Convergence of the SPSA algorithm was determined over the last 100 energy evaluations, by comparing  $\langle E \rangle$  averaged over the first and last sets of 50 values. The calculation was considered converged if the difference between the two sets were below a threshold of 0.0001 Ha.

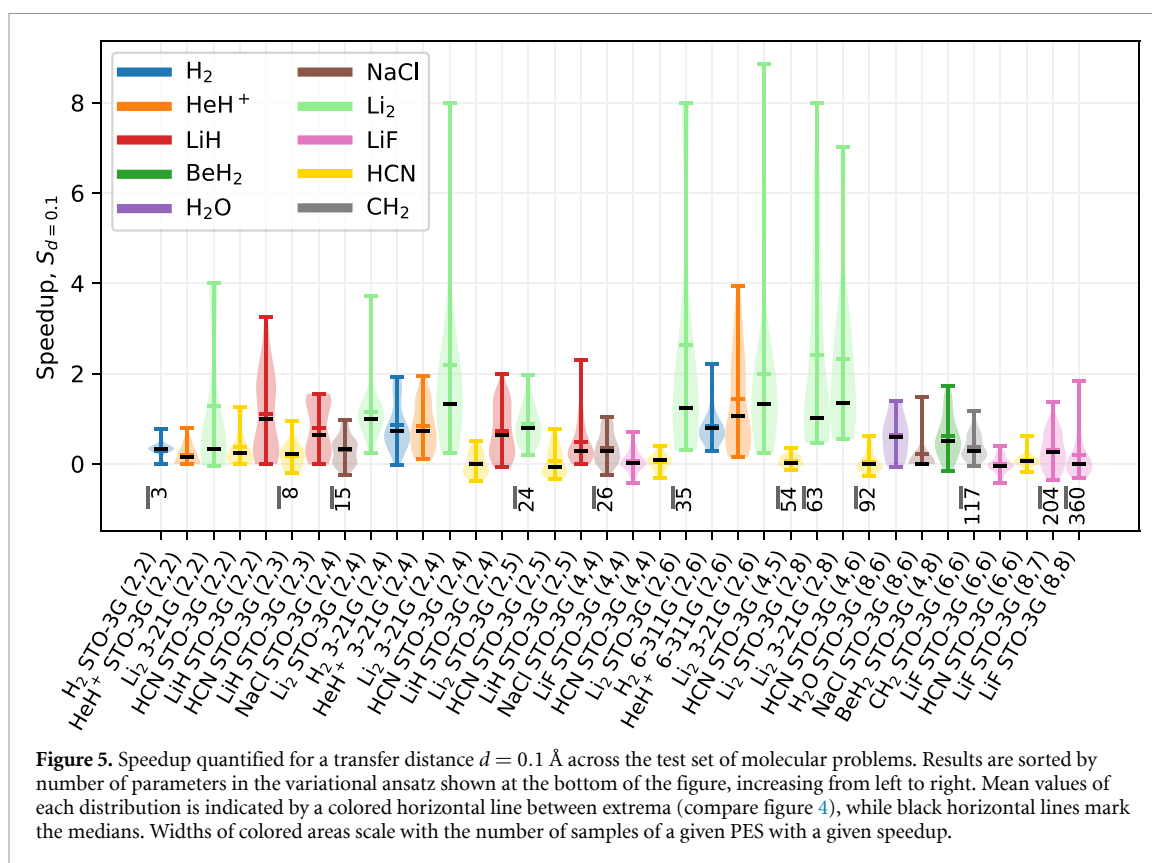
*Sampling and noise:* Each combination of target and transfer geometry were run with the QASM simulator in Qiskit, which performs sampling from an ideal statevector simulation (as opposed to exact statevector). As a practical compromise between speed of simulation and sampling precision when using the SPSA optimizer we took 32 786 ( $2^{15}$ ) samples for each calculation. When using the COBYLA optimizer 1048 576 ( $2^{20}$ ) samples were instead used, owing to the algorithm’s lower noise-resilience. These choices retain a substantial contribution to stochastic noise in the simulations, and is similar in magnitude to what can be expected in current and near-term hardware experiments [20]. To evaluate average performance, multiple calculations were performed for both the HF starting guess and the PT implementation, the results of which were averaged. Speedup for each point  $(R, d)$  was evaluated as an average over all possible combinations of 10–15 separate calculations initialized with HF and PT. To limit computational cost associated with these simulations, the considered circuit sizes were reduced to include only 2 (H<sub>2</sub> and HeH<sup>+</sup>) and 4 qubits (LiH).

## 6. Results and discussion

### 6.1. PT in noise-free VQE

To evaluate PT, we first look at noise-free (statevector) simulations. While not being faithful models of NISQ hardware, noise-free simulations can provide valuable insight into the limits of quantum algorithms. By removing the effects of noise, algorithmic errors can be separated from those caused by imperfect hardware





**Figure 5.** Speedup quantified for a transfer distance  $d = 0.1 \text{ \AA}$  across the test set of molecular problems. Results are sorted by number of parameters in the variational ansatz shown at the bottom of the figure, increasing from left to right. Mean values of each distribution is indicated by a colored horizontal line between extrema (compare figure 4), while black horizontal lines mark the medians. Widths of colored areas scale with the number of samples of a given PES with a given speedup.

and choices of peripherals such as optimizers. We will later return to assess the effects of sampling noise in a subset of examples.

Our test set consist of 36 different quantum chemistry problems: PES of 10 molecules treated with different basis sets and active spaces, which are detailed below. For each case, we consider the dissociation of a single atom from a molecule, making a one-dimensional PES. In triatomics, such as  $\text{BeH}_2$ ,  $\text{H}_2\text{O}$  and  $\text{HCN}$ , a single H atom is dissociated, while all bond angles remain fixed.

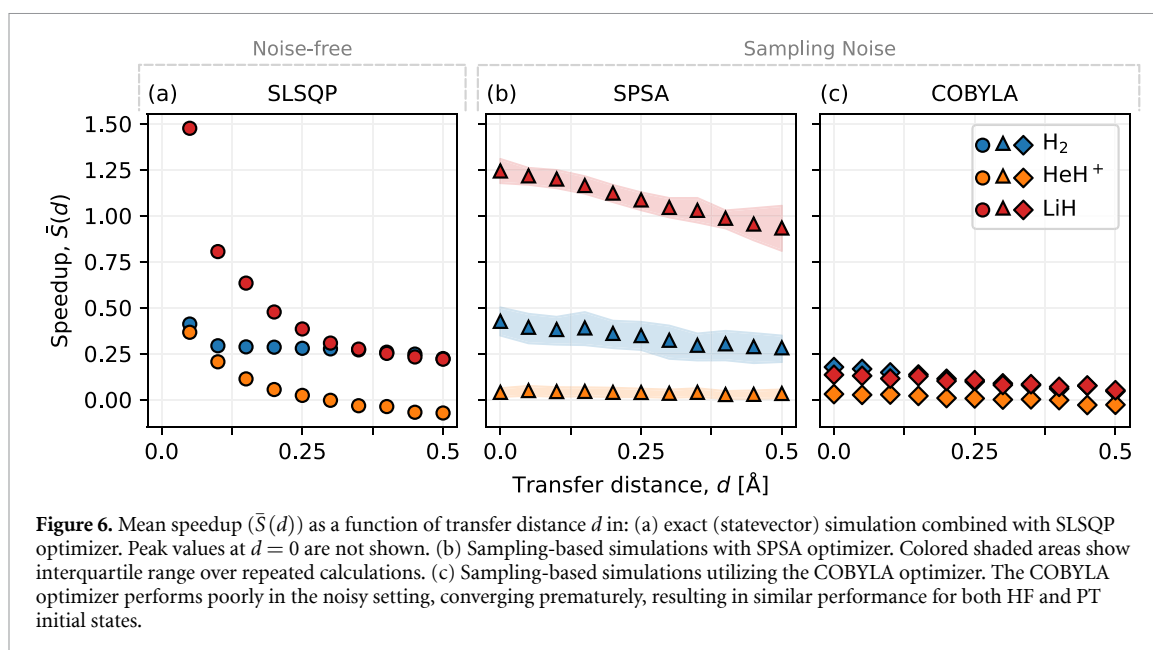
The molecules we study cover a range of electronic structures in terms of electron correlation, including both archetypical ionic dimers ( $\text{HeH}^+$  and  $\text{NaCl}$ ) and covalently bound molecules ( $\text{H}_2$  and  $\text{H}_2\text{O}$ ). Our test set was selected to represent a variety of parameter set sizes, quantum circuits and Hamiltonians of relevance to current and near-term hardware experiments, while still allowing for classical simulation.

The speedup quantified for a given calculation will inevitably be linked to the choice of optimization algorithm. Some such algorithms will function well for noiseless simulations, whereas others are more suitable in the presence of noise. For noise-free optimization, we here rely on the SLSQP optimizer. SLSQP is a representative optimizer for convex optimization in the absence of noise, commonly used in statevector simulations of VQE [45–48]. As we later implement sampling noise, we will subsequently use the SPSA algorithm, which is better adapted for optimization in the presence of noise. We stress that it is outside the scope of this work to provide benchmarking of optimization algorithms (such benchmarking has been done by others [49–51]). Nevertheless, we expect the general conclusion of our work—useful speedup with PT—to hold for *reasonable* choices of optimizers.

One important limit to PT is the transfer distance  $d$ . As  $d$  approaches zero in the absence of noise the transfer perfectly corresponds to the already converged solution. Figure 4(a) shows an expected decrease in average and median speedup over our test set as a function of increasing transfer distance. Figures 4(b)–(e) serve to illustrate the variety in speedup between calculations in our test set. We attribute these differences to the suitability of the HF initial guess: there is less benefit to PT in systems that are well-described by a HF state. In some cases, such as the H dissociation from  $\text{HCN}$ , the speedup is marginal or non-existent, except for very short transfer distances.

For a transfer distance of  $0.1 \text{ \AA}$ , which is a practical choice for PES sampling, we note an average speedup of 70% (figure 4(a)). Note that the choice of a fixed transfer distance for all systems is far from optimal, and  $0.1 \text{ \AA}$  is here merely chosen to facilitate comparison. In figure 5 we look further at the speedup corresponding to this specific transfer distance, but additionally averaged over entire bond dissociation curves, from





compressed bonds to equilibrated geometries and dissociation. Results in figure 5 correspond to calculations with increasing number of circuit parameters from left to right. An overall trend is not apparent, except for some molecules (like  $\text{Li}_2$ ) where an improvement with the number of parameters may be present. The PT approach appears to perform well under noiseless conditions even for calculations exceeding 100 parameters.

## 6.2. Effects of measurement sampling

Even with fault-tolerant quantum computing, which is yet to be realized, one needs repeated sampling of expectation values of the quantum system. Such sampling from a noise-less state introduces a stochastic variability, or sampling noise. In this section, we reevaluate three of our test set calculations,  $\text{H}_2$  (2,2),  $\text{HeH}^+$  (2,2), and  $\text{LiH}$  (2,3), in the presence of sampling noise. Due to the presence of sampling noise, SLSQP is no longer suitable (nor practically feasible) and we use the SPSA and COBYLA optimization algorithms (figure 6).

The most notable difference in the performance of PT when sampling noise is introduced is a markedly different dependence of speedup on transfer distance. Instead of a rapid decline, as in figure 4, speedup appear to decrease linearly with increasing transfer distance. We attribute this difference to the inherently larger challenge of optimization in the presence of noise. Figure 6(c) serves to highlight the importance of choosing optimizers well suited for the task at hand. Even though convergence is reached with the COBYLA optimizer, the number of evaluations at convergence *decreases* with increasing sampling noise, implying that figure 6(c) reflects a poor choice in optimizer, not a failure of PT.

The large effect of even relatively low noise levels in combination with improper choices of optimizers, highlights a challenge that NISQ algorithms face when being compared to conventional methods. We further discuss optimizer suitability in the supporting information. Overall, our evaluation suggests the potential for substantial PT-driven speedup both without and in the presence of sampling noise.

## 7. Conclusions

In this work, we have outlined and provided a proof-of-concept for PT, an approach for accelerating consecutive VQE calculations. In our test set, PT provides an average speedup of 70%, as long as a suitable optimization algorithm is used, compared to when a simulation is initialized by a standard HF initial state. Speedup is apparent both in ideal and sampled statevector simulations, provided that source and target geometries are reasonably ( $\sim 0.1$  Å) close. Results can nevertheless vary substantially between molecular systems, as expected. The PT approach does not appear sensitive to problem complexity, quantified in terms of number of quantum circuit parameters. Simulations subjected to sampling noise also appear less sensitive to the choice of transfer distance. Combined, these results lead us to conclude that PT is an attractive approach for accelerating VQE convergence even for larger more correlated systems. While this study is limited to PT within the context of PES sampling, we also stress that PT can be considered a general principle extendable to other types of calculations. In particular, calculations involving transfer of parameters between different levels of theory is a promising avenue, as shown in [29].

## Data availability statement

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

## Acknowledgments

This research relied on computational resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE, NSC and PDC partially funded by the Swedish research council through Grant Agreement No. 2018-05973 and by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at C3SE and NSC partially funded by the Swedish research council through Grant Agreement No. 2022-06725.

## Funding

This research has been supported by funding from the Wallenberg Center for Quantum Technology (WACQT) and from the EU Flagship on Quantum Technology H2020-FETFLAG-2018-03 Project 820363 OpenSuperQ.

## ORCID iDs

Mårten Skogh  <https://orcid.org/0000-0002-6604-3359>

Oskar Leinonen  <https://orcid.org/0009-0004-0149-1531>

Phalgun Lolur  <https://orcid.org/0000-0002-3409-3415>

Martin Rahm  <https://orcid.org/0000-0001-7645-5923>

## References

- [1] Cao Y *et al* 2019 Quantum chemistry in the age of quantum computing *Chem. Rev.* **119** 10856–915
- [2] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge University Press)
- [3] McArdle S, Endo S, Aspuru-Guzik A, Benjamin S and Yuan X 2020 Quantum computational chemistry *Rev. Mod. Phys.* **92** 015003
- [4] Aharonov D and Ben-Or M 2008 Fault-tolerant quantum computation with constant error rate *SIAM J. Comput.* **38** 1207–82
- [5] Knill E, Laflamme R and Zurek W H 1998 Resilient quantum computation *Science* **279** 342–5
- [6] Kitaev A Y 2003 Fault-tolerant quantum computation by anyons *Ann. Phys., NY* **303** 2–30
- [7] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
- [8] Georgopoulos K, Emary C and Zuliani P 2021 Modeling and simulating the noisy behavior of near-term quantum computers *Phys. Rev. A* **104** 062432
- [9] Wang S, Fontana E, Cerezo M, Sharma K, Sone A, Cincio L and Coles P J 2021 Noise-induced barren plateaus in variational quantum algorithms *Nat. Commun.* **12** 6961
- [10] Elfving V E, Broer B W, Webber M, Gavartin J, Halls M D, Lorton K P and Bochevarov A 2020 How will quantum computers provide an industrially relevant computational advantage in quantum chemistry? (arXiv:2009.12472)
- [11] Cerezo M *et al* 2021 Variational quantum algorithms *Nat. Rev. Phys.* **3** 625–44
- [12] Lolur P, Rahm M, Skogh M, García-Álvarez L and Wendin G 2021 Benchmarking the variational quantum eigensolver through simulation of the ground state energy of prebiotic molecules on high-performance computers *AIP Conf. Proc.* **2362** 030005
- [13] Bittel L and Kliesch M 2021 Training variational quantum algorithms is NP-hard *Phys. Rev. Lett.* **127** 120502
- [14] Cerezo M and Coles P J 2021 Higher order derivatives of quantum neural networks with barren plateaus *Quantum Sci. Technol.* **6** 035006
- [15] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 Barren plateaus in quantum neural network training landscapes *Nat. Commun.* **9** 4812
- [16] McCaskey A J, Parks Z P, Jakowski J, Moore S V, Morris T D, Humble T S and Pooser R C 2019 Quantum chemistry as a benchmark for near-term quantum computers *npj Quantum Inf.* **5** 99
- [17] Lubinski T, Johri S, Varosy P, Coleman J, Zhao L, Necaie J, Baldwin C H, Mayer K and Proctor T 2021 Application-oriented performance benchmarks for quantum computing (arXiv:2110.03137)
- [18] Arute F *et al* Google AI Quantum and Collaborators 2020 Hartree-Fock on a superconducting qubit quantum computer *Science* **369** 1084–9
- [19] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O’Brien J L 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213
- [20] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets *Nature* **549** 242–6
- [21] Colless J I, Ramasesh V V, Dahlen D, Blok M S, Kimchi-Schwartz M E, McClean J R, Carter J, de Jong W A and Siddiqi I 2018 Computation of molecular spectra on a quantum processor with an error-resilient algorithm *Phys. Rev. X* **8** 011021
- [22] Hempel C *et al* 2018 Quantum chemistry calculations on a trapped-ion quantum simulator *Phys. Rev. X* **8** 031022
- [23] Grimsley H R, Economou S E, Barnes E and Mayhall N J 2019 An adaptive variational algorithm for exact molecular simulations on a quantum computer *Nat. Commun.* **10** 3007
- [24] Tang H L, Shkolnikov V O, Barron G S, Grimsley H R, Mayhall N J, Barnes E and Economou S E 2021 Qubit-ADAPT-VQE: an adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor *PRX Quantum* **2** 020310
- [25] Cervera-Lierta A, Kottmann J S and Aspuru-Guzik A 2021 The meta-variational quantum eigensolver (Meta-VQE): learning energy profiles of parameterized Hamiltonians for quantum simulation *PRX Quantum* **2** 020329

- [26] Yuan X, Endo S, Zhao Q, Li Y and Benjamin S 2019 Theory of variational quantum simulation *Quantum* **3** 191
- [27] Szabo A and Ostlund N S 1996 *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications)
- [28] Helgaker T, Jorgensen P and Olsen J 2014 *Molecular Electronic-Structure Theory* (Wiley)
- [29] Kovyrshin A, Skogh M, Broo A, Mensa S, Sahin E, Crain J and Tavernelli I 2023 A quantum computing implementation of nuclearelectronic orbital (NEO) theory: toward an exact pre-born–oppenheimer formulation of molecular quantum systems *J. Chem. Phys.* **158** 214119
- [30] Weiss K, Khoshgoftaar T M and Wang D 2016 A survey of transfer learning *J. Big Data* **3** 9
- [31] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, Xiong H and He Q 2021 a comprehensive survey on transfer learning *Proc. IEEE* **109** 43–76
- [32] Shaydulin R, Lotshaw P C, Larson J, Ostrowski J and Humble T S 2022 Parameter transfer for quantum approximate optimization of weighted maxcut (arXiv:2201.11785)
- [33] Jing H, Wang Y and Li Y 2022 Data-driven quantum approximate optimization algorithm for cyber-physical power systems (arXiv:2204.00738)
- [34] Galda A, Liu X, Lykov D, Alexeev Y and Safro I 2021 Transferability of optimal QAOA parameters between random graphs (arXiv:2106.07531)
- [35] Egger D J, Mareček J and Woerner S 2021 Warm-starting quantum optimization *Quantum* **5** 479
- [36] Tate R, Farhadi M, Herold C, Mohler G and Gupta S 2022 Bridging classical and quantum with SDP initialized warm-starts for QAOA (arXiv:2010.14021)
- [37] Okada K N, Nishi H, Kosugi T and Matsushita Y 2022 Systematic study on the dependence of the warm-start quantum approximate optimization algorithm on approximate solutions (arXiv:2209.02942)
- [38] Liu H-Y, Sun T-P, Y-c W, Han Y-J and Guo G-P 2023 Mitigating barren plateaus with transfer-learning-inspired parameter initializations *New J. Phys.* **25** 013039
- [39] Aleksandrowicz G et al 2019 Qiskit: an open-source framework for quantum computing (<https://doi.org/10.5281/zenodo.2562111>)
- [40] Hartree D R and Hartree W 1935 Self-consistent field, with exchange, for beryllium *Proc. R. Soc. A* **150** 9–33
- [41] Frisch M J et al 2016 Gaussian 16 Gaussian Inc.
- [42] Kraft D 1988 A software package for sequential quadratic programming 88-28 DLR German Aerospace Center – Institute for Flight Mechanics, Koln, Germany
- [43] Powell M J D 1998 Direct search algorithms for optimization calculations *Acta Numer.* **7** 287–336
- [44] Spall J C 1998 Implementation of the simultaneous perturbation algorithm for stochastic optimization *IEEE Trans. Aerosp. Electron. Syst.* **34** 817–23
- [45] Gao Q, Jones G O, Motta M, Sugawara M, Watanabe H C, Kobayashi T, Watanabe E, Ohnishi Y, Nakamura H and Yamamoto N 2021 Applications of quantum computing for investigations of electronic transitions in phenylsulfonyl-carbazole TADF emitters *npj Comput. Mater.* **7** 70
- [46] Rattew A G, Hu S, Pistoia M, Chen R and Wood S 2020 A domain-agnostic, noise-resistant, hardware-efficient evolutionary variational quantum eigensolver (arXiv:1910.09694)
- [47] Choudhary K 2021 Quantum computation for predicting electron and phonon properties of solids *J. Phys.: Condens. Matter* **33** 385501
- [48] Yalouz S, Koridon E, Senjean B, Lasorne B, Buda F and Visscher L 2022 Analytical nonadiabatic couplings and gradients within the state-averaged orbital-optimized variational quantum eigensolver *J. Chem. Theory Comput.* **18** 776–94
- [49] Lockwood O 2022 An empirical review of optimization techniques for quantum variational circuits (arXiv:2202.01389)
- [50] Romero J, Babbush R, McClean J R, Hempel C, Love P J and Aspuru-Guzik A 2018 Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz *Quantum Sci. Technol.* **4** 014008
- [51] Lavrijsen W, Tudor A, Müller J, Iancu C and de Jong W 2020 Classical optimizers for noisy intermediate-scale quantum devices 2020 *IEEE Int. Conf. on Quantum Computing and Engineering (QCE)* pp 267–77