

A transferable machine learning approach to predict quantum circuit parameters for electronic structure problems

Davide Bincoletto, Korbinian Stein, Jonas Motyl, Jakob S. Kottmann

Angaben zur Veröffentlichung / Publication details:

Bincoletto, Davide, Korbinian Stein, Jonas Motyl, and Jakob S. Kottmann. 2026. "A transferable machine learning approach to predict quantum circuit parameters for electronic structure problems." *Machine Learning: Science and Technology* 7(3): 035055.
<https://doi.org/10.1088/2632-2153/ae7532>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0





PAPER • OPEN ACCESS

A transferable machine learning approach to predict quantum circuit parameters for electronic structure problems

To cite this article: Davide Bincoletto *et al* 2026 *Mach. Learn.: Sci. Technol.* **7** 035055

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

You may also like

- [Data efficient prediction of excited-state properties using quantum neural networks](#)
Manuel Hagelueken, Marco F Huber and Marco Roth
- [Adaptive pruning-based optimization of parameterized quantum circuits](#)
Sukin Sim, Jonathan Romero, Jérôme F Gonthier et al.
- [Quantum extreme learning of molecular potential energy surfaces and force fields](#)
Gabriele Lo Monaco, Marco Bertini, Salvatore Lorenzo et al.



PAPER

OPEN ACCESS

RECEIVED

8 January 2026

REVISED

23 April 2026

ACCEPTED FOR PUBLICATION

29 May 2026

PUBLISHED

15 June 2026

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.



A transferable machine learning approach to predict quantum circuit parameters for electronic structure problems

Davide Bincoletto¹ , Korbinian Stein¹, Jonas Motyl¹ and Jakob S Kottmann^{1,2,*} ¹ Institute for Computer Science, University of Augsburg, Augsburg, Germany² Center for Advanced Analytics and Predictive Sciences, University of Augsburg, Augsburg, Germany

* Author to whom any correspondence should be addressed.

E-mail: jakob.kottmann@uni-a.de**Keywords:** graph neural networks, transfer learning, quantum circuits, quantum chemistry

Abstract

The individual optimization of quantum circuit parameters is currently one of the main practical bottlenecks in variational quantum eigensolvers for electronic systems. To this end, several machine learning approaches have been proposed to mitigate the problem. However, such method predominantly aims at training and predicting parameters tailored to individual molecules: either a specific structure, or several structures of the same molecule with varying bond lengths. This work explores machine learning based modeling strategies to include transferability between different molecules. We use a well investigated quantum circuit design and apply it to model properties of a variety of molecular systems where we show parameter prediction that is systematically transferable to instances significantly larger than the training instances.

1. Introduction

The electronic structure problem for molecular systems is a central problem in quantum chemistry, providing important insights for real-world applications like materials science and drug discovery. While classical computation methods are sufficiently fast for smaller systems, the field of quantum computing promises to efficiently find the resulting ground state energy of larger quantum chemical systems [1].

The problem can be briefly stated as finding the lowest eigenvalue of a given second-quantized Hamiltonian—a formally $2^{2N} \times 2^{2N}$ matrix given in a sparse format requiring only $O(N^4)$ numbers. Here N denotes the number of Fermionic sites, or orbitals. Since electrons have a spin, each orbital can be occupied by a spin-up or spin-down electron (or both). A natural mapping to qubit systems is then to use $2N$ qubits to represent the occupation of the N orbitals by spin-up and spin-down electrons. Quantum algorithms, like the quantum phase estimation [2], project a given input state onto an eigenstate of the electronic Hamiltonian, where the success probability is proportional to the input state and the corresponding eigenstate. This is however an expensive protocol [3, 4], where recent improvements bring the possibility of successful future applications closer to reality [5–7]. It is therefore advantageous to have methods that can prepare adequate initial states that approximate electronic ground states. Other algorithms, like prolate diagonalization [8], imaginary-time evolution [9] or efforts to simulate quantum dynamics [10] also profit from efficient initial state preparations.

A class of algorithms capable of direct state preparation are variational quantum eigensolvers (VQE) [11, 12] that use the variational principle to optimize parameters of specific circuit designs [13, 14]. While there are good arguments, that VQEs alone will not reach verifiable quantum advantage they are still useful for the task of initial state preparation—once the circuit is trained, the preparation cost is typically cheap. In a nutshell, a VQE protocol takes a parameterized quantum circuit and couples it to a classical optimizer to find the parameters that minimize the expectation value with respect to an electronic Hamiltonian. We provide further technical details in appendix A. Over the years, several circuit designs have emerged [15], most are however not developed enough to be applicable in a black-box fashion. One candidate with such property is the separable pair approximation (SPA) [16], which has proven to be a robust method in several applications, both as a consistent baseline to compare with

other methods [17, 18] and as a tool for targeting diverse chemical properties [19–21]. SPA is a sub-circuit of the more general k -UpCCGSD [22] method as well as recently proposed design schemes such as tiled unitary products [23] or structured layering based on chemical resonance structures [24]. The states prepared by the circuits have also been proven advantageous in quantum subspace diagonalization [25]. This will be employed throughout the work. Further details are presented in appendix B.

The classical optimization step can be a notoriously difficult task [26, 27] for unstructured approaches. This has led to multiple attempts at using alternative techniques, e.g. machine learning, to model VQE parameters with the goal of improving reliability in finding optimal solutions. Modeling quantum circuit parameters has shown success in various related works. Zhang and Yin [28], started with collective optimization of groups of Hamiltonians originating from the same molecule at various conformations. Later Cervera-Lierta *et al* [29], introduced the Meta-VQE, addressing the problem on a more general level: minimizing the parameters of a given parameterized Hamiltonian. This approach is inspired by quantum machine learning algorithms and extends beyond quantum chemistry applications. Alternatively, several works [30–32] employed classical neural network or machine-learning related approaches to directly predict circuit parameters or whole circuits [33] from the Hamiltonian. More recently, Bensoussan *et al* [34] and Miao *et al* [32] introduced complementary machine-learning-assisted frameworks for quantum eigensolvers, leveraging classical ML models to optimize eigensolver hyperparameters and to encode Hamiltonian parameters directly into variational circuits, respectively.

Despite numerous investigations, we are currently not aware of a method that accurately leverages molecular structure and is capable of generalizing to large-size systems. This work explores multiple approaches to model variational quantum parameters of system-adapted quantum circuits based on atomic coordinates. We extend upon the basis of the Meta-VQE concept with classical machine learning techniques to predict parameters that enable initialization of the ansatz circuit in an optimal state. We evaluate our models on hydrogenic systems up to H_{12} (up to H_{16} in the appendix, at this time, limited due to simulation cost), with a major focus on to demonstrate transferability across larger systems, and present results with non-hydrogenic systems (Li_2 and C_2H_4) as a proof of concept of transferability with respect to chemical variety. In fact, while most of the methods require training for different sizes of parameterized Hamiltonians, our method aims at building a scalable model that can extend to multiple molecule sizes and species. We refer to [24, 25] for potential transferability between hydrogenic model chemistries and more general classes of molecules.

This article is structured as follows. In section 2, we introduce the data generation and the modeling approaches. In section 3 we show the results of our experimentation for multiple datasets. Finally, we discuss our findings and list several topics for future work in section 4.

2. Methodology

To tackle the parameter prediction problem, we developed three architectures. For the first model we trained a graph attention network (GAT) [35] with the aim of using the graph structure of the molecule as a basis for a dynamic prediction model, leveraging the message passing between adjacent nodes. Previous applications to chemistry have shown promising results on multiple molecular properties at the graph level, including but not limited to the ground state energy [36–38]. Our algorithms employ a similar problem structure, representing molecular geometries as graphs, where atoms correspond to vertices and the bonds to edges.

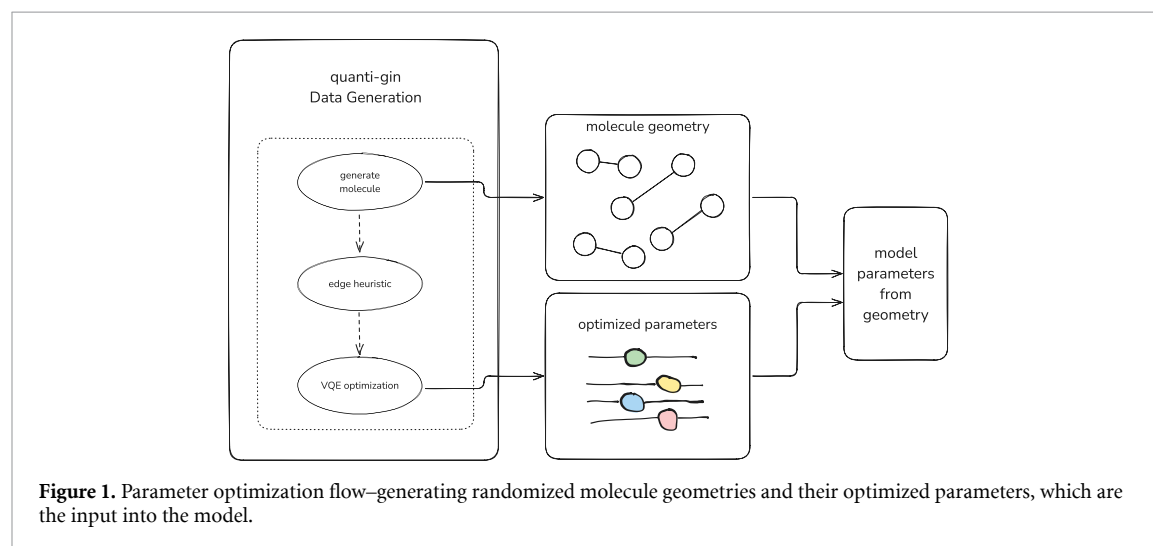
The second and third models are based on the Schrödinger’s network (SchNet) [37]—a neural architecture specifically designed for learning molecular representations. It employs continuous filter convolutions, which enables the accurate modeling of interactions at small distances. The native model predicts ground state energy for a mixed set of molecules. Since the problem we are focusing on is similar, we can use the SchNet as backbone and employ the core structure of the architecture for modeling variational parameters, which then result in the ground state energy. With a new head for the backbone we are able to train two different models based on the SchNet architecture with slightly different architecture approaches, which have different weighting on the employed graphs of our ansatz. The models have been developed sequentially in the order presented with the goal to further refine their capabilities and allow training with reduced training sets. An overview of all the models features is showcased in table 1.

2.1. Data generation

For parameter modeling and analysis, we made use of QUANTI-GIN [39] to generate datasets of molecular geometries containing Hamiltonians, quantum circuits in the form of the SPA [16] and the

Table 1. Overview of the examined neural network architectures. GAT and Linear SchNet models use identical training data and preprocessing, while the Mixed SchNet model employs a smaller, mixed dataset with coordinate reordering based on graph matching.

Model	Training dataset	# Parameters	Input preprocessing
Graph attention network (GAT)	230k linear H_4	302 625	Euclidean distance matrix with angles
Linear SchNet	230k linear H_4	28 273	Euclidean distance matrix with angles
Mixed SchNet	1.000 linear H_4 2.000 random H_6	472 450	reorder coordinates according to perfect matching graph



corresponding optimized parameters following the strategies of reference [24]. The workflow is illustrated in figure 1. The procedure for generating a datapoint with index i (out of n total) is as follows:

1. Generate the coordinates C according to the molecular structure following algorithm 1. See appendix for alternative structures.
2. Estimate the optimal chemical graph (Lewis formula) for the coordinates C . We used scaled Euclidean distances as edge weights and heuristically determined a perfect matching graph G with minimal edge weight.
3. Construct the circuit ansatz U_{SPA} employing the graph G and compute the corresponding orbital-optimized Hamiltonian H_{opt} .
4. Minimize the expectation value $\langle U_{\text{SPA}} | H_{\text{opt}} | U_{\text{SPA}} \rangle$ with a VQE, to obtain the corresponding energy E_{SPA} and parameters θ .
5. Normalize the angles and store the set $(C, H, G, E_{\text{SPA}}, \theta)$ as one instance of the dataset.

Here we applied a constraint on the coordinates C generated by the randomized procedure (algorithm 1). Starting from the origin, each new atom is placed at a random point with a distance between 0.5 and 2.5 Å from the previous existing atom, ensuring a minimum separation of 0.5 Å. This avoids clustering of very small inter-atomic distances as well as generating dissociated structures with no relevant physical interactions. See the code in the appendix for a step-by-step example using the TEQUILA [40] library.

In the experimentation we generated two training datasets: a large one, made of 230.000 linear H_4 instances, and a small one, made of 1.000 linear instances of H_4 and 2.000 random instances of H_6 . The purpose of the large dataset is mainly to get a good estimate of the best possible result when only training on linear H_4 instances. Where the smaller training set was chosen to demonstrate that comparable, or improved, quality can be obtained with significantly smaller training sets that still contain only small instances. For the evaluation we produced several datasets of 500 random instances with varying molecule sizes (ranging from H_2 to H_{12}). Here we chose a minimal basis approach (one orbital, or two qubits, for each hydrogen atom) and refer to other works [10, 20, 41] regarding improved problem

Algorithm 1. Random coordinates generation with minimum separation.

Input: Number of atoms $n \in \mathbb{N}$,
maximum displacement $d_{\max} \in \mathbb{R}^+$,
Output: Coordinate matrix $C \in \mathbb{R}^{n \times 3}$
Set $C \leftarrow \{(0,0,0)\}$
for $i \leftarrow 1$ **to** $n - 1$ **do**
 Select a reference point $\mathbf{r}_{\text{prev}} \in C$ uniformly at random;
 $\delta_{\min} \leftarrow 0$
 While $\delta_{\min} \leq 0.5$ **do**
 Sample $\mathbf{u} \sim \mathcal{U}([0,1]^3)$ at random;
 $\mathbf{r}_{\text{new}} \leftarrow \mathbf{r}_{\text{prev}} + d_{\max} \cdot \mathbf{u}$
 $\delta_{\min} \leftarrow \min_{\mathbf{r} \in C} \|\mathbf{r}_{\text{new}} - \mathbf{r}\|_2$
 Append $\mathbf{r}_{\text{new}}$ to C ;
return: C ;

encodings. In this work, we focus on hydrogenic systems because they are relatively simple to treat, while still containing fundamental geometrical structures that can be leveraged in the study of non-hydrogenic systems, as shown in previous works [24, 42].

2.2. Modeling approaches

Our goal is to build a model which, given the features derived from the molecule (positions and atom types), outputs a value for the parameters (angles) of the VQE circuit. As these angles are continuous values, $\theta \in [-2\pi, 2\pi]$, this is a regression problem. We chose PYTORCH [43] as deep learning framework, due to its wide use in the field. It provides a large amount of built-in primitives, methods and utilities that can be used in the construction of neural networks. This was combined with PYTORCH GEOMETRIC [44] for additional functionalities.

2.2.1. GAT

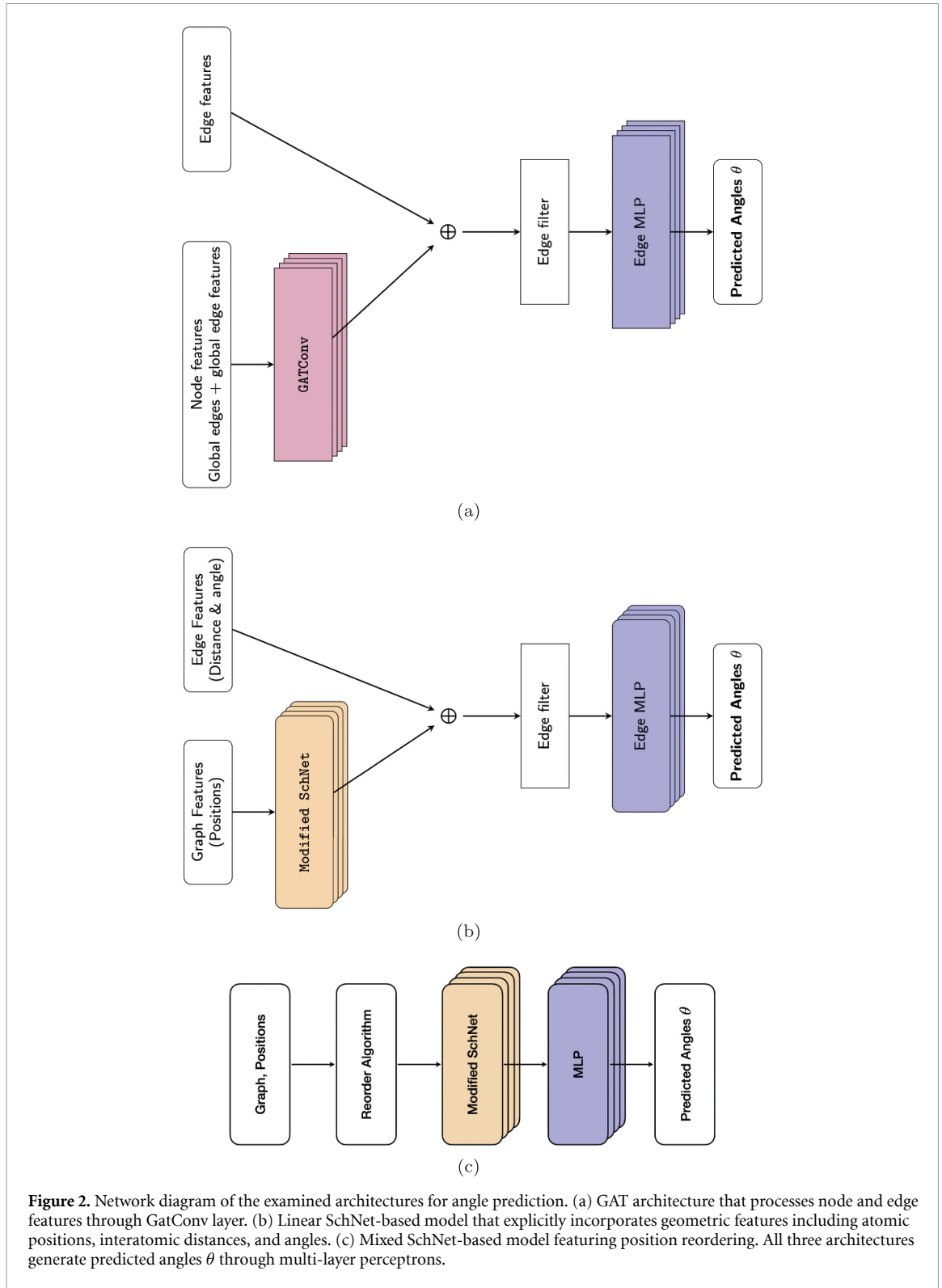
Due to the recent success of attention mechanisms [45], we decided to use this approach as the foundation for our GAT model. Such mechanisms have demonstrated effectiveness in capturing complex dependencies between elements in a sequence or graph, thus, we assume they are suitable for modeling molecular interactions. [46] PYTORCH GEOMETRIC provides a graph attention convolution layer (GATConv) implementation, which applies the attention mechanism on node features using message passing to their respective edges [35]. Here, message passing refers to the process where each node updates its representation by aggregating and transforming feature information (messages) from its neighbors according to the graph structure. This approach allows the model to assign different importance weights to neighboring atoms based on their features and relative positions, which is crucial for accurately capturing the interactions that determine optimal circuit parameters. As we use the layer like a black box for the implementation, we will only describe the architecture of the model built around the graph attention convolution. For more implementation details on the graph attention implementation, refer to [35]. We provide a visual overview of the graph-attention GAT in figure 2(a).

Through the combination of the atom coordinates and edges by our heuristic, we already have a graph that we can apply to the GAT, but there are some important factors to consider. For one, the given graph only contains paired nodes, which, given the edge-based message passing in the GATConv layer, would lead to only the paired atoms exchanging information. Additionally, using 3-dimensional coordinates as node features to learn the relationships introduces a feature of high dimensionality. When limited to them, this is likely to greatly increase the required amount of data, therefore also the time and resources required to train the model. For this reason, we perform the following pre-processing and enhancement steps on the molecule graph data before passing it into the model:

1. Generate global edges with the same edge heuristic from the data generation routine.
2. Generate global and non-global edge features, for edges $\vec{e}$, with euclidean distance and angle:

$$\vec{e}_{uv} = (\|\mathbf{x}_u - \mathbf{x}_v\|_2, \theta_{uv})$$

where $\mathbf{x}_u, \mathbf{x}_v$ are the positions of nodes u and v , and θ_{uv} is the angle of the corresponding edge.



3. For each feature set $\vec{v} = (\mathbf{x}_u, \mathbf{x}_v, \vec{e}_{uv})$ perform normalization of the respective components:

$$\vec{v} = \frac{\vec{v} - \min(\vec{v})}{\max(\vec{v})}$$

Importantly, we keep the node features as the bare coordinates of the atoms, as to leave a possibility for the model to extract useful information from coordinates and their correlations.

We first pass the node features, global edges and their generated features through one or more GATConv layers, whose output x is an internal representation of node features, modeling interactions between the atoms that are in close proximity. Next, we perform the combination of edge features and node representation x , called the edge filter, which uses the non-global edges uv from the edge heuristic described above. In this step, for each edge uv , we concatenate the node features x_u and x_v and the corresponding edge features $\vec{e}_{uv}$:

$$\vec{c}_{uv} = \vec{x}_u \oplus \vec{x}_v \oplus \vec{e}_{uv}$$

Finally, we pass the concatenated features through a multi-layer perceptron (edge MLP) with two linear layers, scaling up from the feature dimension of $\vec{c}_{uv}$ to 256, a ReLU activation function, and a final linear layer that scales down to dimension 1. As the concatenated edge features $\vec{c}_{uv}$ have dimension $[E, N]$, where E is the count of edges and N the feature dimension, we obtain the predicted angles for the circuit $\vec{\theta} \in \mathbb{R}^E$ for every edge between determined pairs.

2.2.2. SchNet backbone

While the architecture described in the previous section has the capability to model the impact of inter-pair influences, it is not tuned specifically for predictions of molecular graph representations. For our second architecture, we turned to SchNet [37]. This approach is based on modeling the aforementioned interactions using continuous filter convolutions, instead of discretized grids that the distances need to be fitted to, which allows for the modeling of small distance impacts. Based on the successful application of SchNet, in order to model quantum chemical properties on molecules with a similar input structure, we propose a new architecture using a modified SchNet as a backbone. This architecture shares the pre-processing and the final layer with the previously described GAT, but replaces the GATConv layers for modeling interatomic interactions. The network diagram is shown in figure 2(b).

In the original SchNet, the forward pass takes the coordinates C and applies a distance expansion using radial-basis functions. These extend the dimensionality of the distance feature from 1 to 50, passing the distance through 50 Gaussian functions with different centers. After that, the embedded information gets passed to a series of three residual blocks, inspired by ResNet [47]. At each layer, our molecule is represented in an atom-wise manner, analogous to pixels in an image. The key mechanism enabling this representation is a series of interaction blocks, which iteratively update atomic feature vectors based on the embedded molecular geometry.

Because we want to process the node representations directly, for concatenating them with the edge information features, we cut off the original SchNet after the interaction blocks and add a custom head to it. We combine the internal node representations with the edge features, where the latter is first passed through the distance expansion layer:

$$\vec{c}_{uv} = \vec{x}_u \oplus \vec{x}_v \oplus f(\vec{e}_{uv})$$

here, $\vec{x}$ are the SchNet node representations and $f(\vec{e})$ the edge features passed through the distance expansion, based on the edge heuristic. After this combination, we pass the features through a simple edge MLP, again reducing down the dimensionality of the output to match the edges and parameters count in the final VQE circuit.

2.2.3. Mixed Schrödinger network

Based on the experience with the two previous models, we constructed a third model trained on a significantly smaller dataset including mixed instances (linear and randomized instances of H_4 and H_6). For this we used SchNet as a backbone but applied a different head. After the residual interactions, in fact, we take the embedded representation and build new layers on top, see 2(c). As input we have the coordinates $C = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n\} \subset \mathbb{R}^3$ and the corresponding perfect matching graph G . If we want to predict the angles for our SPA circuit we follow the steps:

1. Reorder C to according to the perfect matching graph G , e.g. $C = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\} \rightarrow \tilde{C} = \{\mathbf{r}_1, \mathbf{r}_4, \mathbf{r}_2, \mathbf{r}_3\}$ for the graph of H_4 . A brief discussion over the graph matching preprocessing is in the appendix.
2. Process the coordinates $\tilde{C}$ to obtain pairwise embedding with SchNet backbone. This is done by increasing the dimensionality of the available information for each pair to a parameterized hidden dimension, e.g. $s(\mathbf{r}_1, \mathbf{r}_4) \rightarrow X^{1,4}$.
3. Shape the output of the backbone to the amount of angles we need with a MLP. The circuit ansatz needs one angle for each electron pair, e.g. $\text{MLP}(X^{1,4}) \rightarrow \theta^{1,4}$

Here, the processing steps in the forward pass give the perfect matching graph more emphasis on the output of the model compared to previous architectures. After concatenating the angles of all layers into one tensor, we have a suitable representation for training and evaluating our model. The network diagram is shown in figure 2(c).

This architecture differs slightly from the previously mentioned linear SchNet model and notably was trained on mixed (linear and random) instances. The training dataset for this model contains 1.000 linear H_4 instances and 2.000 random H_6 instances.

3. Evaluation

There are multiple ways to measure the performance of our proposed models. While we can directly compare the output angle values as a metric, we are more interested in the models learning to predict angles that result in close-to minimal energies. For this we do zero-shot, fixed variable evaluation of the circuit. In this evaluation we focus on comparing the result of the baseline energy with the energy that the respective predictor can achieve when directly applying its variables to the circuit. Our goal is to determine and compare the model capability of producing angles, and thus states, close to, or better than, the ground state found by the VQE optimization. The steps for evaluating a single molecule instance are the following:

1. Use the perfect matching graph G and the coordinates C of the evaluated molecule in the forward pass of the model.
2. Map the model predicted angles to the circuit ansatz.
3. Simulate the resulting circuit and return the predicted energy, which is used for determining the model's accuracy.

3.1. Random instance evaluation

For our first evaluation, we utilized the previously defined algorithm 1 for random data generation. We generated 500 samples for each size $s \in \{2, 4, 6, 8, 10, 12\}$ of hydrogenic systems which we then run through the SPA optimization using the edge heuristic. The procedure is equivalent to the example code presented in the appendix, including the orbital optimization step.

The considered metrics are defined as follows:

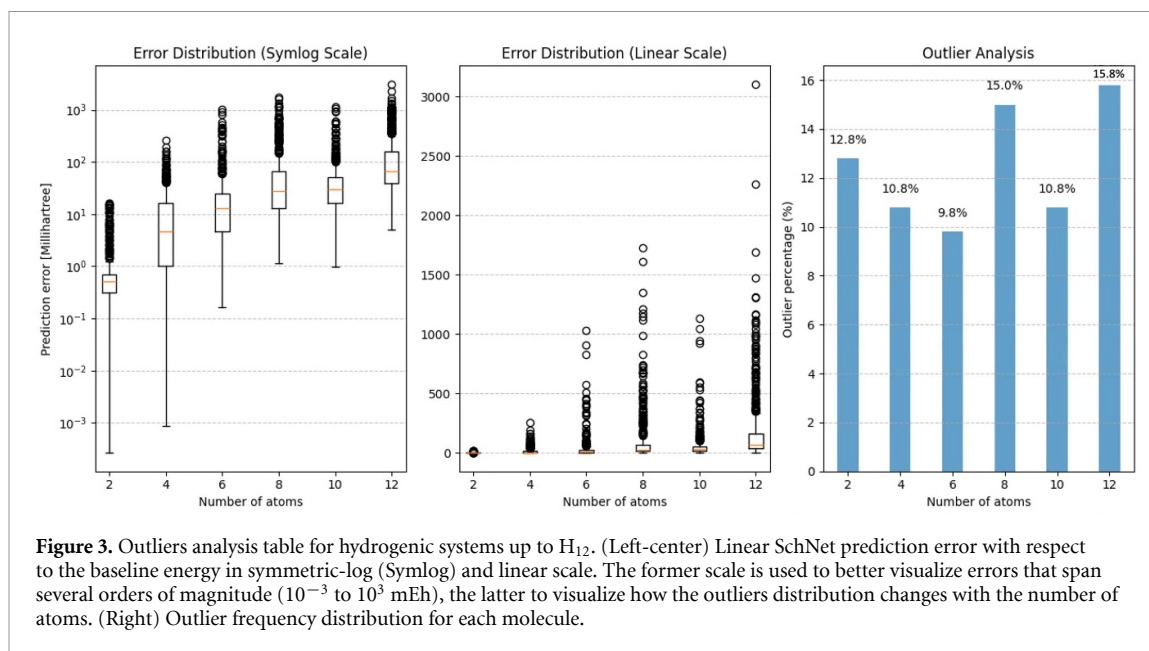
$$\begin{aligned} \text{MSE} &= \frac{\sum_{x \in S} (B_x - M_x)^2}{|S|} \\ \text{ME} &= \frac{\sum_{x \in S} (B_x - M_x)}{|S|} \end{aligned} \quad (1)$$

where B_x is the baseline energy of the x -th sample, and M_x is the corresponding method energy. The mean absolute error is also often referred to as the L_1 loss metric, while the mean squared error (MSE) is commonly known as the L_2 loss metric. We are interested in the former since we do not want to lose information on the instances where $M_x < B_x$. The final results for the fixed variable comparison can be seen in table 2. The best model of each architecture is compared against the baseline, grouped by system size. We observe that for larger system sizes, the SchNet-based models perform better than the GATConv-based model, indicating comparatively best generalization capabilities provided by the architecture. In particular, for the mixed SchNet model, the results indicate that small datasets are sufficient to capture the required information and predict ground state energies with adequate accuracy. The mixture of the dataset also has a positive effect for the generalizability where significant improvements were achieved when testing on randomized structures while the price of computing training instances of H_6 instead of H_4 was relatively small.

To gain more insight about the performance of our models, we carried out detailed comparisons of the energy differences. Figure 3 shows an outlier analysis, i.e. the respective count of samples that lie outside of the standard deviation, comparing the linear SchNet model and the baseline energy. As expected, in our evaluation of the SchNet-based model, the largest amount of energy outliers occur for the H_{12} system, i.e. the instances which differ the most from the linear H_4 training dataset and require the greatest transferability from the models. This is an indicator at potential improvements one can make to the model or dataset.

Table 2. Comparison of fixed variables application methods against baseline, with the models generalizing up to H_{12} from only learning on smaller instances. The exact amount of parameters are shown in table 1. The models were trained on two dataset sizes: large, i.e. 230 k linear instances, and small, i.e. 1.000 linear instances of H_4 and 2.000 random instances of H_6 , and evaluated on 500 instances of H_n .

Atoms	Method	Training size	Model size	ME (mEh)	MSE (mEh)
2	GAT	Large	Medium	14	0.3
	Linear SchNet	Large	Small	1	0.01
	Mixed SchNet	Small	Large	0.2	6×10^{-4}
4	GAT	Large	Medium	43	5
	Linear SchNet	Large	Small	15	1
	Mixed SchNet	Small	Large	6	0.3
6	GAT	Large	Medium	67	15
	Linear SchNet	Large	Small	37	11
	Mixed SchNet	Small	Large	10	2
8	GAT	Large	Medium	107	48
	Linear SchNet	Large	Small	100	54
	Mixed SchNet	Small	Large	34	5
10	GAT	Large	Medium	153	55
	Linear SchNet	Large	Small	58	16
	Mixed SchNet	Small	Large	26	4
12	GAT	Large	Medium	188	113
	Linear SchNet	Large	Small	185	127
	Mixed SchNet	Small	Large	38	6



3.2. Structured instance evaluation

A further evaluation involves assessing model performance on structured instances. For this, we looked at the predicted ground state energies for linear and ring H_n molecules, with the atom bonding lying on the x -axis ranging from 0.5 Å to 4.0 Å. The coordinates were generated using algorithms 2 and 3. Figure 4 shows the errors for the linear SchNet model with a focus on the potential energy curves of the molecules. The linear dataset is redundant since it is also partially in the training set, but it is a good proof of consistency that the SchNet-based models can transfer the learned structure to larger structured instances. Using a log-scale for the plots makes it easier to see that, as expected, the model's accuracy

Algorithm 2. Linear coordinates generation along the z -axis.

Input: Number of atoms $n \in \mathbb{N}$,
iteration index $k \in \{0, \dots, T-1\}$ of $T \in \mathbb{N}$,
boundaries $d_{\min}, d_{\max} \in \mathbb{R}^+$
Output: Coordinate matrix $C \in \mathbb{R}^{n \times 3}$
 $d_{\text{step}} \leftarrow d_{\min} + (d_{\max} - d_{\min}) \cdot \frac{k}{T-1}$
 $z_i \leftarrow i \cdot d_{\text{step}}, \quad i = 0, 1, \dots, n-1$
Initialize $C \leftarrow \mathbf{0}_{n \times 3}$;
Set third column of C to $(z_0, z_1, \dots, z_{n-1})$;
return C ;

Algorithm 3. Ring coordinates generation in the xy -plane.

Input: Number of atoms $n \in \mathbb{N}$,
iteration index $k \in \{0, \dots, T-1\}$ of $T \in \mathbb{N}$,
boundaries $d_{\min}, d_{\max} \in \mathbb{R}^+$
Output: Coordinate matrix $C \in \mathbb{R}^{n \times 3}$
 $d_{\text{step}} \leftarrow d_{\min} + (d_{\max} - d_{\min}) \cdot \frac{k}{T-1}$
 $r \leftarrow \frac{d_{\text{step}}}{2 \sin(\frac{\pi}{n})}$
Initialize $C \leftarrow \mathbf{0}_{n \times 3}$;
for $i \leftarrow 0$ **to** $n-1$ **do**
 $\theta \leftarrow \frac{2\pi i}{n}$;
 $x \leftarrow r \cdot \cos(\theta)$;
 $y \leftarrow r \cdot \sin(\theta)$;
 Set $C[i, 0] \leftarrow x, C[i, 1] \leftarrow y, C[i, 2] \leftarrow 0$;
return C ;

drops as molecule size increases. Note however, that the prediction still remains physically sound while the chains and rings dissociate.

The mixed SchNet was evaluated on the ring instances in the same way with results shown in figure 5. Notably, the mixed model was able to generalize from linear and random instances to structured ring instances. Comparing this results with the linear SchNet ones shows that the accuracy improved by one order of magnitude, which is remarkable given the size of the training set.

3.3. Extension to non-hydrogenic systems

As a proof of concept for the transferability of the described architecture to heterogeneous chemical species, we present an example application on the molecules Li_2 and C_2H_4 . To prove it, we used the Mixed SchNet to predict the SPA parameter of a H_2 molecule and then we used that parameter in a zero-shot simulation of the two molecules. For both of them we want to predict the π -bond formed by the 2p orbitals. In the case of C_2H_4 we used standard SPA to simulate the other bonds. In order to fit the H_2 parameter we rescaled the interatomic distance among the two hydrogen atoms. An empirical search led the scaling factors 0.43 for Li_2 and 1.25 C_2H_4 . Results are presented in figure 6, where we compare, first, the predicted energy with a standard SPA, and then the trend of the predicted parameter with the parameter from the standard SPA.

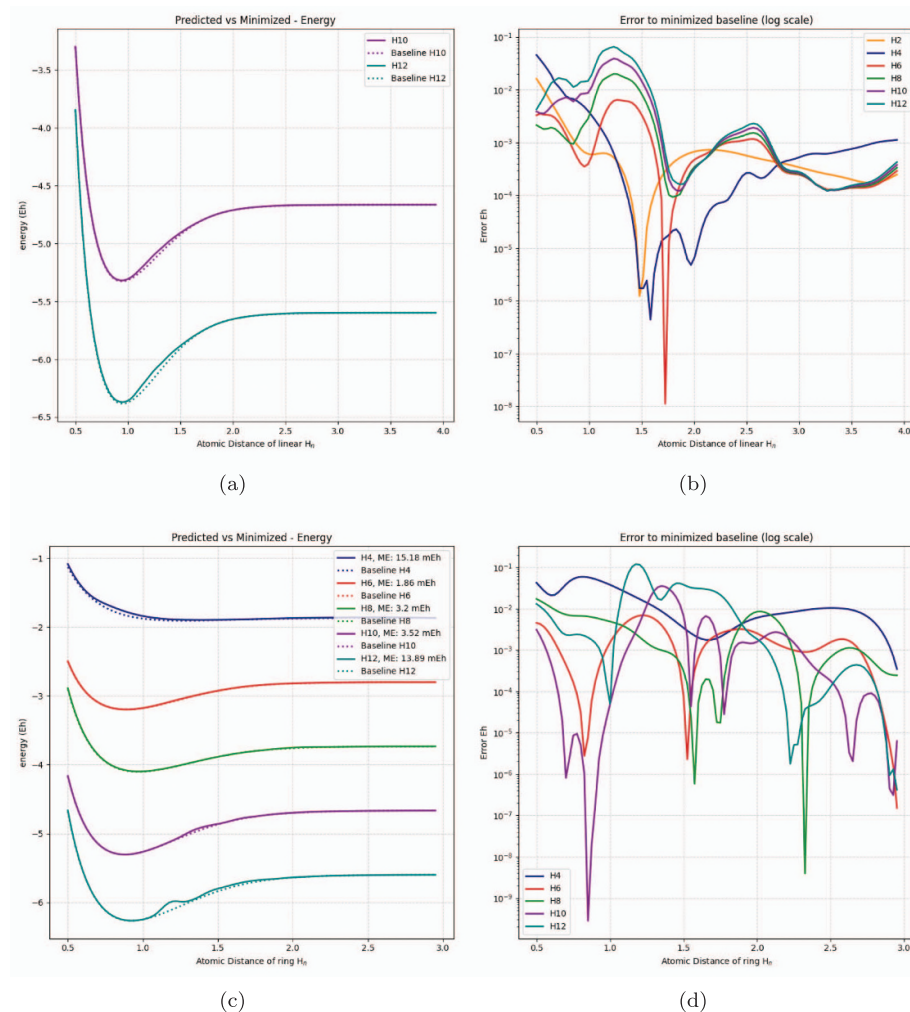


Figure 4. Evaluating linear SchNet model on linear and ring instances up to H_{12} , with the model being trained on 230k linear H_4 instances. Values are compared against the baseline SPA energy. We present additional illustrative demonstrations on larger systems with mixed SchNet in the form of linear H_{14} and H_{16} molecules in the appendix. (a) Model predictions for linear H_{10} and H_{12} . (b) Model prediction errors on log-scale. (c) Model predictions for ring H_n . (d) Model prediction errors on log-scale.

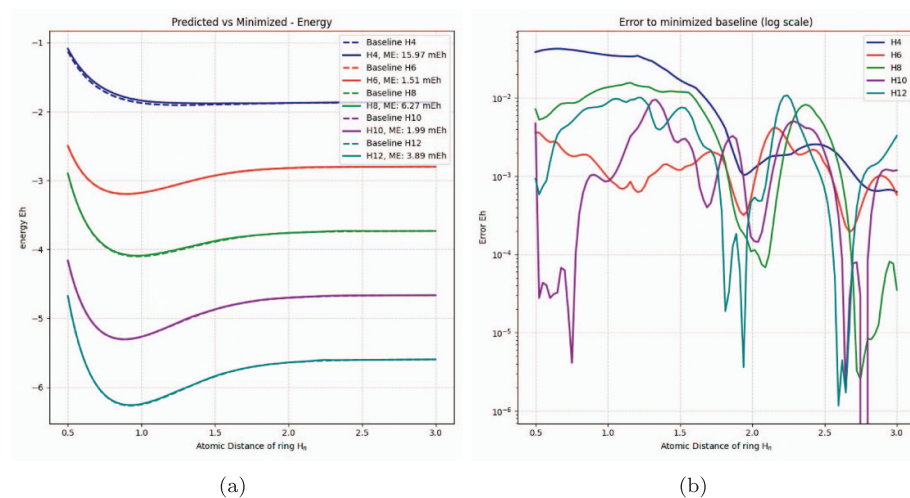
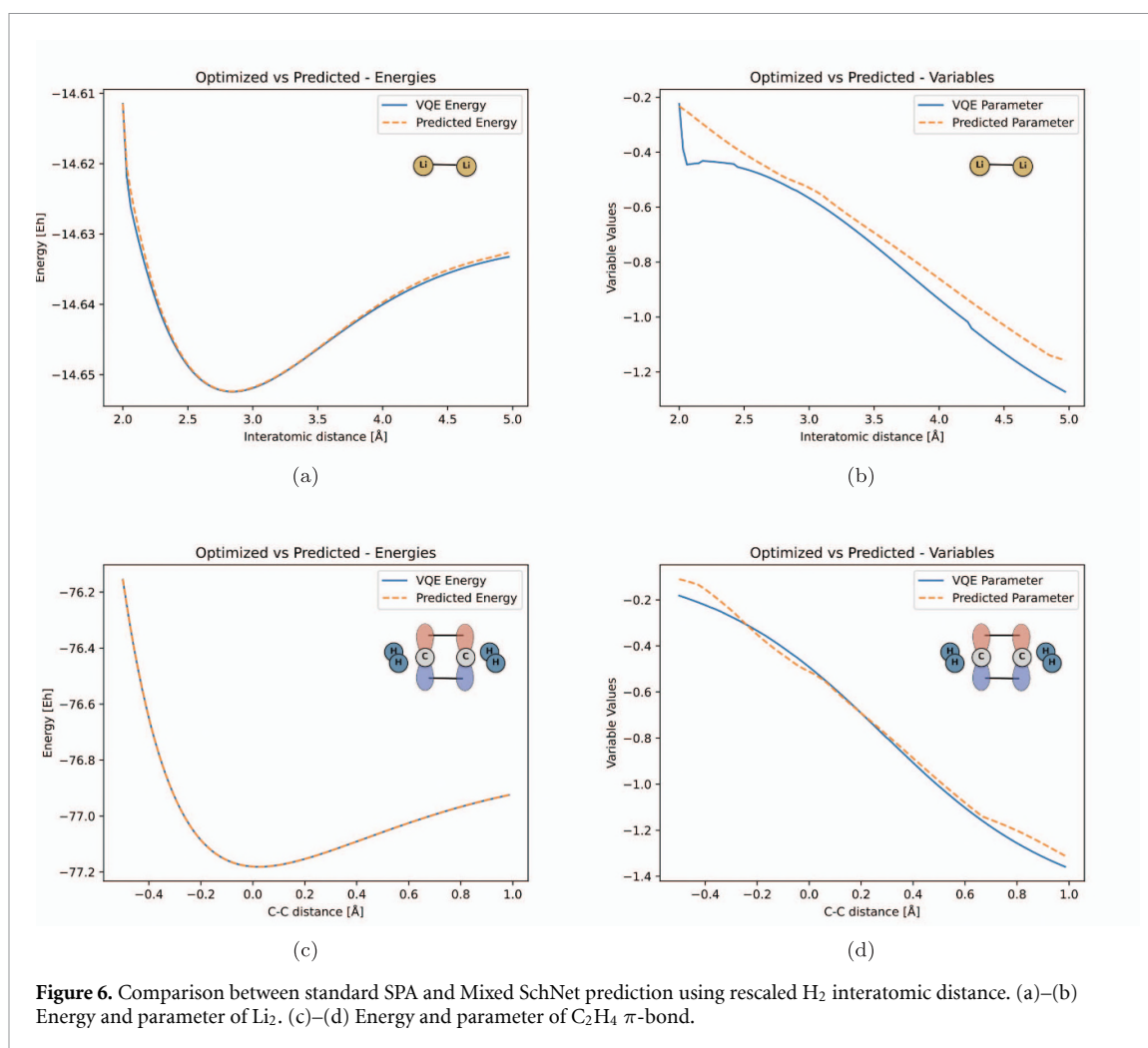


Figure 5. Evaluating mixed SchNet model on ring instances, generalizing up to H_{12} instances, with the training dataset containing 1,000 linear instances of H_4 and 2,000 random instances of H_6 . We present additional illustrative demonstrations on larger systems in the form of linear H_{14} and H_{16} molecules in the appendix. (a) Model predictions for ring H_n , compared against baseline SPA energy. (b) Model prediction errors on log-scale, compared against baseline SPA energy for H_n .



4. Conclusion and outlook

In this work, we have explored the concept of modeling variational quantum parameters on SPA circuits, experimenting with three advanced approaches. These have shown success in modeling the circuit parameters of hydrogenic systems, generalizing for unseen system sizes up to H₁₂. Notably, the forward pass of all three models requires only the coordinates and the edges of the best perfect matching graph for their predictions, showing that such fundamental data representation is sufficient to model molecular systems. In particular, the use of edges, which require a preprocessing step, show that perfect matching graphs accurately describe electronic correlation in a local manner.

The GAT and Linear SchNet models were trained on linear instances only, showing a good starting point for accurate results and generalizability to random and structured geometries. However, the proposed SchNet-based architectures proved best in predicting angles, yielding good variational parameters for the SPA ansatz. By further experimenting with such architecture, we have shown that by mixing the structure of the geometries in the training dataset we achieve a significant increase in accuracy. For this, we used linear and random molecule structures, showing that much of the necessary information can be extracted from such arrangements. This is true also for transferability, in fact, the model accurately predicts unseen instances, i.e. larger structured and random instances up to H₁₂. Moreover, this last mixed model was trained with relatively small amount of data, demonstrating a better capability in capturing correlation information among atoms in the molecular structure.

Regarding the geometries considered for the evaluations, we have seen that the errors grow with system size, as more potentially relevant interactions arise, and also depend on the symmetry of the molecule. Highly asymmetric random geometries with small interatomic distances exhibit the largest errors. In contrast, longer bond lengths tend to be less error-prone, as the variational parameters naturally converge to $\pi/2$ for large interatomic distances. A possible source of error in our results is also the graph heuristic, as it becomes increasingly difficult to determine the appropriate molecular graph for

large, asymmetric systems, particularly at shorter bond lengths. In this regard, one can use optimized heuristics in order to not scan all the possible graphs and prevent the scaling.

Following this work, we propose multiple paths for further development. Given larger and extended datasets and increased computational resources, we assume it is possible to enhance the models, improving on accuracy and transferability across various geometries. This could facilitate initialization of larger system sizes, as well as generalization to non-hydrogenic systems. In addition, computational complexity can be increased not only through larger training datasets but also through greater quality in the VQE-based solutions. This is due to the challenge in finding the global minimum of the optimization landscape. Finally, there are many possibilities for improving the models through alternative machine learning techniques, both in terms of architecture and learning strategies, such as semi-supervised learning or reinforcement learning.

A natural next step is to extend the most promising (mixed SchNet) approach of this work from minimal basis hydrogenic systems to (organic) molecules in more chemically accurate bases. Here we see the challenges on a rather technical side resulting from the many technicalities introduced by the quantum chemical methodology necessary for the task. Regarding the used chemical basis sets, one way to improve would be to switch to numerically optimized orbitals [10, 41]. A grander challenge remains in adopting the methodology to more expressive circuit designs—ideal candidates would be extended graph-based designs [24], localized hierarchical approaches [23, 48] or established coupled-cluster method [22]. Note that all three contain the here used separable pair design as substructure.

We hope to provide a solid foundation for such future endeavors with this work.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/tequilahub/variational-parameter-modeling> [49].

Funding

This work has been funded by the Hightech Agenda Bayern and the Munich Quantum Valley through the Lighthouse project KID-QC².

Conflict of interest

All authors declare no conflict of interest.

Appendix A. Variational quantum eigensolvers

Given that the expectation value of an Hamiltonian can be expressed as a sum of expectation values of its components

$$\langle H \rangle = \sum_{i\alpha} h_{\alpha}^i \langle \sigma_{\alpha}^i \rangle + \sum_{ij\alpha\beta} h_{\alpha\beta}^{ij} \langle \sigma_{\alpha}^i \sigma_{\beta}^j \rangle + \dots \quad (2)$$

with $h \in \mathbb{R}$ being real numbers and σ Pauli operators, the variational quantum eigensolvers (VQE) routine can be summarized with the following steps:

1. Prepare a quantum state $|\Psi(\vec{\theta})\rangle$ with a parameterized ansatz circuit on the quantum computer
2. Measure the expectation values $\langle H_i \rangle$ and combine them classically to approximate $\langle H \rangle$.
3. Update the parameters $\vec{\theta}$ in the state preparation, by minimizing $\langle H \rangle$ with an optimization algorithm, e.g. gradient descent.

VQEs require lower quantum resources in terms of circuit depth and qubit coherence time than traditional quantum algorithms such as quantum phase estimation [2, 6]. This is due to a hybrid quantum–classical structure. The relevant aspect of VQE is the decomposition and computation of the expectation value of an operator as the sum of individual or grouped expectation values. The trade-off

of such algorithm is a significant overhead due to measurement repetitions and classical processing. Moreover, the measurement procedure can be affected by noise, which can impact its accuracy and precision.

Appendix B. The separable pair model

In the following we will give a high-level description of the separable pair approximation (SPA)—a variational circuit design that has been proven effective with regards to convergence rate and consistency. We will limit ourselves to hydrogenic systems in minimal atomic bases and refer to the original works for general molecules [16, 24, 25].

Hydrogenic systems are a collection of hydrogen atoms in 3D (spatial) space. For N hydrogen atoms the input data consists solely of N 3-tuples representing the spatial coordinates. For each of the N atoms a 3D function, called basis function, is assigned. These basis functions resemble the ground state of the isolated hydrogen atom and are used to define the electronic Hamiltonian. For simplicity we can imagine each of these orbitals as quantum state for an electron localized at one of the atoms.

In this work, we use the approximation which assumes that the total wavefunction is a product of individual electron pairs and comes in a highly optimized form with regards to gate-count and circuit-depth. In contrast to other VQE variants, SPAs can be automated and have shown to give consistent results. They are however not a isolated method but can be identified as sub-structures in various VQE approaches, such as improved graph-based designs [24], tiled-unitary products [23], quantum number preserving fabrics [48] or decomposed approaches constructed from linear combinations over different SPA circuits [25].

The energy expectation values of SPA circuits are not invariant with respect to orbital rotations—meaning unitary rotations in the space of the basis functions that define the Hamiltonian can lead to better energies. This is not a surprise, as changing the basis functions also changes the interpretation which orbitals are assigned to the individual electron pairs. Note that these basis rotations do not change the spectrum of the Hamiltonian itself, they merely improve the quality of the SPA. As they can be optimized classically, we do not explicitly include them into the circuit within this work.

Appendix C. Additional results on larger hydrogenic systems

The choice of H_{12} as the largest system considered in the main set of evaluations is primarily motivated by computational convenience, as the current VQE implementation becomes increasingly demanding at larger system sizes. This limitation, however, should not be interpreted as a fundamental barrier of the approach. To illustrate its broader applicability, we additionally present representative results for larger linear hydrogen chains, specifically H_{14} and H_{16} , in figure 7. For these extended systems, the predicted energies remain in good agreement with the baseline SPA results, providing further evidence that the model retains its accuracy beyond the training regime. These observations support the expectation that the proposed methodology generalizes effectively to larger systems, even as the underlying quantum simulation becomes the dominant computational bottleneck.

Appendix D. Discussion over graph matching preprocessing

The mixed SchNet architecture incorporates a graph-matching preprocessing routine prior to evaluation. In this work, we employed a brute-force strategy to select, at each step, the graph that minimizes a suitably defined scaled distance sum. While this approach is conceptually simple and guarantees optimal matching within the chosen metric, it is not strictly required. We adopt it here because, for the training set sizes considered, the associated computational overhead remains modest. More generally, the number of possible graph pairings grows combinatorially as $\mathcal{O}((n-1)!!)$, where n denotes the number of orbitals. Despite this rapid scaling, empirical observations from our prior experiments indicate that the graph-matching stage does not become a practical bottleneck up to systems of roughly 16 orbitals (corresponding to 32 qubits). In this regime, the computational cost is instead dominated by the quantum simulation itself, which quickly eclipses the preprocessing effort. We note that more efficient alternatives do exist. In particular, we have developed heuristic strategies inspired by classical valence bond theory

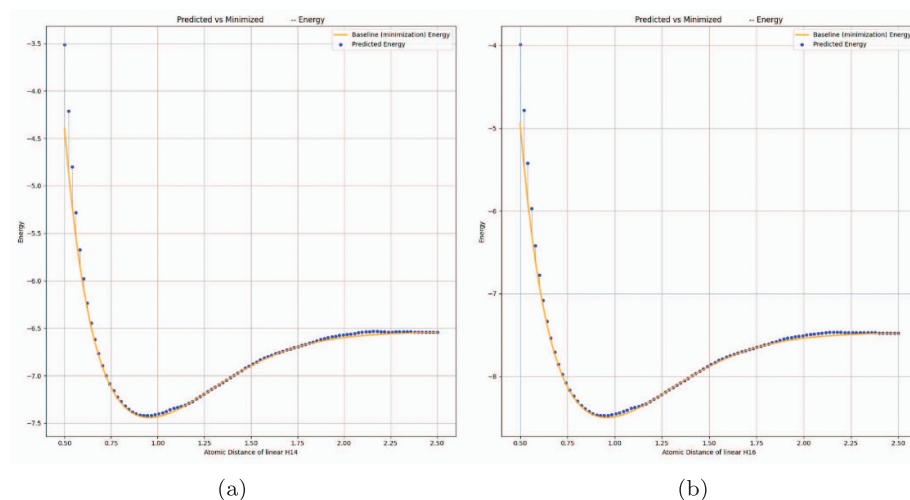


Figure 7. Mixed SchNet predictions for linear H_{14} seen in (a) and linear H_{16} seen in (b), compared against the minimized energy. Resulting in a mean error of 38.36 mEh (a) and 43.14 mEh (b).

that reduce the need to exhaustively enumerate all possible graphs. However, a detailed discussion of these methods would substantially broaden the scope of the present work and detract from its primary focus. For this reason, we defer their exposition to future studies.

Listing 1. Tequila H4 example

```

1 def calculate_optimization():
2     geometry = """
3 H 1.0 1.0 1.0
4 H 2.0 1.0 1.0
5 H 4.0 2.0 2.0
6 H 5.0 2.0 2.0"""
7
8     edges = [[0, 1], [2, 3]]
9     mol = tq.Molecule(geometry=geometry, basis_set="sto-3g")
10
11     U = mol.make_ansatz("HCB-SPA", edges=edges)
12     export_to(U, filename="h4_circuit.pdf")
13
14     # construct initial guess
15     initial_guess = np.eye(4)
16     for edge in edges:
17         initial_guess[edge[0]][edge[0]] = 1.0
18         initial_guess[edge[0]][edge[1]] = 1.0
19         initial_guess[edge[1]][edge[0]] = -1.0
20         initial_guess[edge[1]][edge[1]] = 1.0
21
22     opt_orbitals = tq.chemistry.optimize_orbitals(
23         mol.use_native_orbitals(),
24         U,
25         initial_guess=initial_guess,
26         use_hcb=True,
27         silent=True,
28     )
29
30     print("\n\n\n====optimized orbitals====\n\n\n", opt_orbitals.mo_coeff)
31
32     mol = opt_orbitals.molecule
33     H = mol.make_hardcore_boson_hamiltonian()
34
35     print("\n\n\n====running optimization====\n\n\n")
36     E = tq.ExpectationValue(H=H, U=U)
37     result = tq.minimize(E, silent=True)

```

ORCID iDs

Davide Bincoletto  0009-0007-9212-6493

Jonas Motyl  0009-0006-2037-2663

Jakob S Kottmann  0000-0002-4156-2048

References

- [1] McArdle S, Endo S, Aspuru-Guzik A, Benjamin S C and Yuan X 2020 Quantum computational chemistry *Rev. Mod. Phys.* **92** 015003
- [2] Aspuru-Guzik A, Dutoi A D, Love P J and Head-Gordon M 2005 Simulated quantum computation of molecular energies *Science* **309** 1704–7
- [3] Reiher M, Wiebe N, Svore K M, Wecker D and Troyer M 2017 Elucidating reaction mechanisms on quantum computers *Proc. Natl Acad. Sci.* **114** 7555–60
- [4] von Burg V, Low G H, Häner T, Steiger D S, Reiher M, Roetteler M and Troyer M 2021 Quantum computing enhanced computational catalysis *Phys. Rev. Res.* **3** 033055
- [5] Rocca D, Cortes C L, Gonthier J F, Ollitrault P J, Parrish R M, Anselmetti G L and Streif M 2024 Reducing the runtime of fault-tolerant quantum simulations in chemistry through symmetry-compressed double factorization *J. Chem. Theory Comput.* **20** 4639–53
- [6] Cortes C L, Rocca D, Gonthier J F, Ollitrault P J, Parrish R M, Anselmetti G L R and Streif M 2024 Assessing the query complexity limits of quantum phase estimation using symmetry-aware spectral bounds *Phys. Rev. A* **110** 022420
- [7] Ollitrault P J, Gonthier Jérôme F, Rocca D, Anselmetti G-L, Degroote M, Moll N, Santagati R and Streif M 2025 Improving the runtime of quantum phase estimation for chemistry through basis set optimization *J. Chem. Theory Comput.* **21** 12034–47
- [8] Stroschein T, Castaldo D and Reiher M 2025 Ground and excited-state energies with analytic errors and short time evolution on a quantum computer
- [9] Motta M, Sun C, Tan A T K, O'Rourke M J, Ye E, Minnich A J, Brandao F G S L and Chan G K-L 2020 Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution *Nat. Phys.* **16** 205–10
- [10] Langkabel F and Bande A 2022 Quantum-compute algorithm for exact laser-driven electron dynamics in molecules *J. Chem. Theory Comput.* **18** 7082–92
- [11] 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
- [12] Bharti K et al 2022 Noisy intermediate-scale quantum algorithms *Rev. Mod. Phys.* **94** 015004
- [13] Tilly J et al 2022 The variational quantum eigensolver: a review of methods and best practices *Phys. Rep.* **986** 1–128
- [14] Cerezo M et al 2021 Variational quantum algorithms *Nat. Rev. Phys.* **3** 625–44
- [15] Anand A, Schleich P, Alperin-Lea S, Jensen P W K, Sim S, Díaz-Tinoco M, Kottmann J S, Degroote M, Izmaylov A F and Aspuru-Guzik A 2022 A quantum computing view on unitary coupled cluster theory *Chem. Soc. Rev.* **51** 1659–84
- [16] Kottmann J S and Aspuru-Guzik A 2022 Optimized low-depth quantum circuits for molecular electronic structure using a separable-pair approximation *Phys. Rev. A* **105** 032449
- [17] Weber M, Anand A, Cervera-Lierta A, Kottmann J S, Kyaw T H, Li B, Aspuru-Guzik A, Zhang C and Zhao Z 2022 Toward reliability in the nisq era: robust interval guarantee for quantum measurements on approximate states *Phys. Rev. Res.* **4** 033217
- [18] Schleich P, Boen J, Cincio L, Anand A, Kottmann J S, Tretiak S, Dub P A and Aspuru-Guzik A 2023 Partitioning quantum chemistry simulations with clifford circuits *J. Chem. Theory Comput.* **19** 4952–64
- [19] Gil E S, Oppel M, Kottmann J S and González L 2025 Sharc meets tequila: mixed quantum-classical dynamics on a quantum computer using a hybrid quantum-classical algorithm *Chem. Sci.* **16** 596–609
- [20] Schleich P, Kottmann J S and Aspuru-Guzik A 2022 Improving the accuracy of the variational quantum eigensolver for molecular systems by the explicitly-correlated perturbative [2]r12-correction *Phys. Chem. Chem. Phys.* **24** 13550–64
- [21] del Arco Santos F J and Kottmann J S 2025 A hybrid qubit encoding: splitting fock space into fermionic and bosonic subspaces *Quantum Sci. Technol.* **10** 035018
- [22] Lee J, Huggins W J, Head-Gordon M and Whaley K B 2018 Generalized unitary coupled cluster wave functions for quantum computation *J. Chem. Theory Comput.* **15** 311–24
- [23] Burton H G A 2024 Accurate and gate-efficient quantum ansätze for electronic states without adaptive optimization *Phys. Rev. Res.* **6** 023300
- [24] Kottmann J S 2023 Molecular quantum circuit design: a graph-based approach *Quantum* **7** 1073
- [25] Kottmann J S and Scala F 2024 Quantum algorithmic approach to multiconfigurational valence bond theory: insights from interpretable circuit design *J. Chem. Theory Comput.* **20** 3514–23
- [26] 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** 1–6
- [27] Bittel L and Kliesch M 2021 Training variational quantum algorithms is np-hard *Phys. Rev. Lett.* **127** 120502
- [28] Zhang D-B and Yin T 2020 Collective optimization for variational quantum eigensolvers *Phys. Rev. A* **101** 032311
- [29] Cervera-Lierta A, Kottmann J S and Aspuru-Guzik A 2021 Meta-variational quantum eigensolver: learning energy profiles of parameterized hamiltonians for quantum simulation *PRX Quantum* **2** 020329
- [30] Ceroni J, Stetina T F, Kieferova M, Marrero C O, Arrazola J M and Wiebe N 2023 Generating approximate ground states of molecules using quantum machine learning
- [31] Truger F, Barzen J, Leymann F and Obst J 2025 Warm-starting the vqe with approximate complex amplitude encoding (<https://doi.org/10.5220/0013513400004525>)
- [32] Miao J, Hsieh C-Y and Zhang S-X 2024 Neural-network-encoded variational quantum algorithms *Phys. Rev. Appl.* **21** 014053
- [33] Nakaji K, Björn Kristensen L, Campos-Gonzalez-Angulo J A, Vakili M G, Huang H, Bagherimehrab M, Gorgulla C, Wong F, McCaskey A, Kim J-S, Nguyen T, Rao P and Aspuru-Guzik A 2024 The generative quantum eigensolver (gqe) and its application for ground state search
- [34] Bensoussan A, Chachkarova E, Even-Mendoza K, Fortz S and Lenihan C 2025 Accelerq: accelerating quantum eigensolvers with machine learning on quantum simulators *Proc. ACM Program. Lang.* **9** 2279–309

- [35] Veličković P, Cucurull G, Casanova A, Romero A, Lió P and Bengio Y 2018 Graph attention networks
- [36] Gilmer J, Schoenholz S S, Riley P F, Vinyals O and Dahl G E 2017 Neural message passing for quantum chemistry *Int. Conf. Machine Learning* (PMLR) pp 1263–72
- [37] Schütt K, Kindermans P-J, Felix H E S, Chmiela S, Tkatchenko A and Müller K-R 2017 Schnet: a continuous-filter convolutional neural network for modeling quantum interactions *Advances in neural information processing systems* p 30
- [38] von Glehn I, Spencer J S, and Pfau D 2023 A self-attention ansatz for ab-initio quantum chemistry
- [39] Stein K 2024 quanti-gin: A customizable data generator for quantum simulation (available at: <https://github.com/nylser/quanti-gin>)
- [40] Kottmann J S *et al* 2021 TEQUILA: a platform for rapid development of quantum algorithms *Quantum Sci. Technol.* **6** 024009
- [41] Kottmann J S, Schleich P, Tamayo-Mendoza T and Aspuru-Guzik A 2021 Reducing qubit requirements while maintaining numerical precision for the variational quantum eigensolver: A basis-set-free approach *J. Phys. Chem. Lett.* **12** 663
- [42] Bincoletto D and Kottmann J S 2026 A physics-informed measurement protocol for expectation values of fermionic observables *Digit. Discov.* **5** 1257–68
- [43] Ansel J *et al* 2024 Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation *Proc. of the 29th ACM Int. Conf. on Architectural Support for Programming Languages and Operating Systems, Volume 2, ASPLOS '24* (Association for Computing Machinery) pp 929–47
- [44] Fey M and Lenssen J E 2019 Fast graph representation learning with pytorch geometric
- [45] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention is all you need *CoRR* **30** 1–15
- [46] Lee J B, Rossi R A, Kim S, Ahmed N K and Koh E 2019 Attention models in graphs: a survey *ACM Trans. Knowl. Discov. Data* **13** 1–25
- [47] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
- [48] Anselmetti G-L R, Wierichs D, Gogolin C and Parrish R M 2021 Local, expressive, quantum-number-preserving vqe ansätze for fermionic systems *New J. Phys.* **23** 113010
- [49] Bincoletto D, Motyl J and Kottmann J S 2026 variational-parameter-modeling: Code for prediction of quantum circuit parameters *GitHub* (available at: <https://github.com/tequilahub/variational-parameter-modeling>)