Provably Trainable Rotationally Equivariant Quantum Machine Learning

Maxwell T. West[1,*] Jamie Heredge[1] Martin Sevior[1] and Muhammad Usman[1,2,†]

[1]*School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia*
[2]*Data61, CSIRO, Clayton, VIC 3168, Australia*

Exploiting the power of quantum computation to realize superior machine learning algorithms has been a major research focus of recent years, but the prospects of quantum machine learning (QML) remain dampened by considerable technical challenges. A particularly significant issue is that generic QML models suffer from so-called barren plateaus in their training landscapes—large regions where cost function gradients vanish exponentially in the number of qubits employed, rendering large models effectively untrainable. A leading strategy for combating this effect is to build problem-specific models that take into account the symmetries of their data in order to focus on a smaller, relevant subset of Hilbert space. In this work, we introduce a family of rotationally equivariant QML models built upon the quantum Fourier transform, and leverage recent insights from the Lie-algebraic study of QML models to prove that (a subset of) our models do not exhibit barren plateaus. In addition to our analytical results we numerically test our rotationally equivariant models on a dataset of simulated scanning tunneling microscope images of phosphorus impurities in silicon, where rotational symmetry naturally arises, and find that they dramatically outperform their generic counterparts in practice.

I. INTRODUCTION

Stimulated by steady progress in quantum computing hardware [1], the possibility of using parameterized quantum circuits to carry out machine learning tasks has been thoroughly investigated in recent years as a near-term application of noisy intermediate-scale quantum computers [2–11]. Among the most important discoveries in quantum machine learning (QML) has been the existence of barren plateaus in the training landscapes of variational quantum models [12–15], reminiscent of the vanishing gradients that plagued early neural networks [16]. Barren plateaus have come to be understood to be linked to the expressibility of the ansatz being optimized [14,17–19], with generic, highly expressible models rendered effectively untrainable as the number of qubits increases. This suggests that the architectures of variational QML models should be tweaked on a per-problem basis, with domain-specific knowledge employed to construct ansatzes with inductive biases compatible with the optimal solution. A natural approach along these lines is to build models that explicitly respect the symmetries of their data, so-called *geometric quantum machine learning* (GQML) [20–30]. For example, GQML has been used to classify images with spatial symmetries [28,29], and more general data that satisfies permutation symmetries [23,24,31]. By constraining the search space of the optimization procedure, such symmetry-informed models have in a few instances been proved to be free of barren plateaus [23,32]. Despite a general theory of the trainability of QML models starting to emerge [18,19], however, the number of explicit examples of provably trainable models that have been developed and benchmarked on real data to test their performance in practice remains limited [23,32–34].

In this work, we introduce a family of rotationally equivariant QML models for classifying two-dimensional data with labels that are invariant to rotations by $2\pi/2^k$ for $k \in \mathbb{N}$. In particular, these models are applicable to image data, the consideration of the symmetries of which in the GQML context has previously been limited to reflections [28] and 90° rotations [29]. Our proposed architectures allow for the fraction of resources allocated to the processing of the radial and angular degrees of freedom to be controlled, enabling for an interpolation between a model with no explicit notion of rotational symmetry ($k = 0$) to models that respect continuous rotations ($k \to \infty$). This is facilitated by choosing an encoding strategy that ensures that the representation of the group action on the encoded quantum states is that of the regular representation of $\mathbb{Z}_{2^k}$, which can be diagonalized via a (quantum) Fourier

transform, following which it becomes simple to write down an equivariant model, without needing to resort to explicit symmetrization techniques such as twirling [21]. We study the trainability of our models, finding that the scaling of the variance of the cost function with respect to the trainable parameters of the model differs qualitatively depending on the fraction of the qubits used to encode the radial and angular degrees of freedom. Specifically, we use recently developed techniques [18,19] that involve studying the dynamical Lie algebras of quantum circuits to prove that our models are free from barren plateaus when the number of qubits devoted to encoding the radial information grows at most logarithmically with the total number of qubits.

In addition to our analytical results, we test the performance of our models in practice by benchmarking them against generic, fully expressible models on a ten-class classification task involving the analysis of scanning tunneling microscope (STM) images of phosphorus impurities implanted in silicon. This dataset was chosen as an example of a relatively complicated problem that both possesses rotational symmetry and has important applications to the fabrication of spin-based quantum computers in silicon [35]. We find that our rotationally equivariant models achieve drastically better results and scale much more readily to deeper model sizes than the generic models,

which suffer from significant trainability issues. This is achieved despite the generic models being in principle capable of finding any solution available to the equivariant models, and adds to the growing body of evidence supporting GQML as a promising pathway for quantum machine learning.

II. ROTATIONALLY EQUIVARIANT QUANTUM MACHINE LEARNING

We begin by briefly introducing the basic ideas of GQML needed to motivate our rotationally equivariant architectures. Much more detail can be found in the literature, e.g., Refs. [21,22]. In the standard supervised learning setting we are given data $\mathbf{x}$ drawn from a set $\mathcal{X}$, and attempt to determine associated labels $y(\mathbf{x}) \in \mathcal{Y} = \{1, 2, \ldots, n_{\text{classes}}\}$. The quantum models we employ for this task here are composed of three main components—an initial data encoding layer, a trainable variational section, and a final set of measurements [see Figs. 1(a) and 1(b)]. In the final stage n_{classes} operators $\{M_j\}_{j=1}^{n_{\text{classes}}}$ are measured, with prediction $\hat{y}$ of the model on input $\mathbf{x}$ defined to be the class corresponding to the operator with the highest expectation value, i.e.,

$$\hat{y}_{\boldsymbol{\theta}}(\mathbf{x}) = \arg\max_{j} \langle 0| \mathcal{E}(\mathbf{x})^{\dagger} U_{\boldsymbol{\theta}}^{\dagger} M_j U_{\boldsymbol{\theta}} \mathcal{E}(\mathbf{x}) |0\rangle, \quad (1)$$
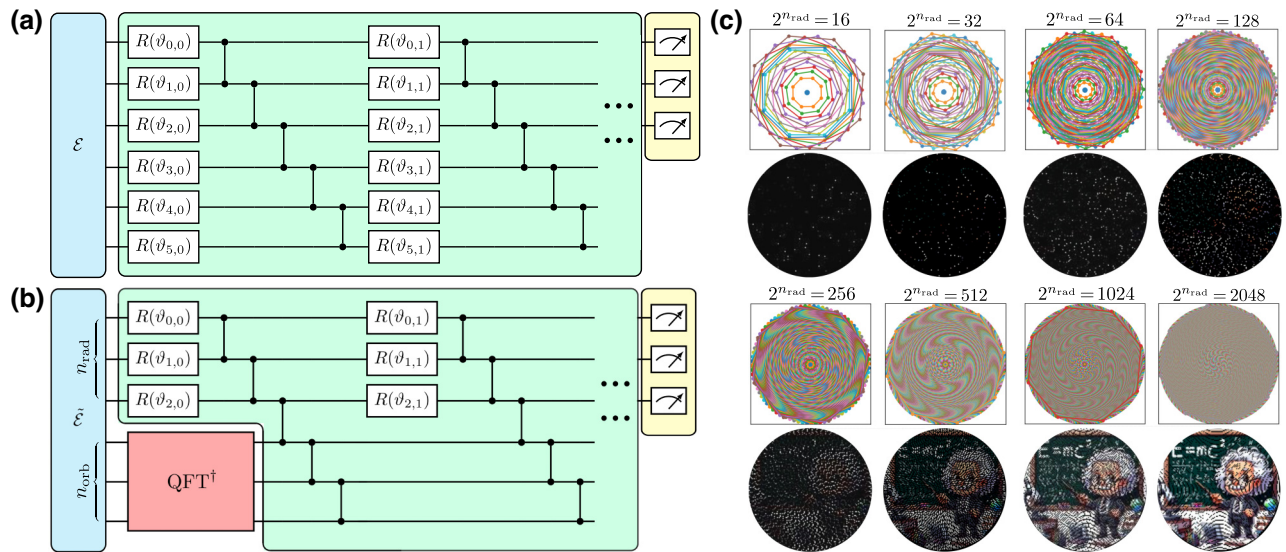

FIG. 1. Rotationally equivariant quantum machine learning. (a) A generic fully expressible QML model, consisting of arbitrary parameterized single-qubit rotation gates and nearest-neighbor CZ gates. At deep depths, models such as these are known to suffer from *barren plateaus*, regions in their training landscape where the gradients of the cost function vanish exponentially, rendering the models untrainable [18]. (b) The architecture of our rotation-equivariant models, which are less than fully expressible, having been restricted to the subspace of the full Hilbert space that respects rotation symmetry. The combination of our encoding choice $\tilde{\mathcal{E}}$ [see Eq. (4)] and the quantum Fourier transform ensures that the induced representation of the rotation symmetry group will commute with any operation on the first n_{rad} qubits, and operations on the last n_{orb} qubits that are diagonal in the computational basis. Thus, by employing an arbitrary circuit on the first n_{rad} qubits and CZ gates only on the remainder, we construct an equivariant model. (c) Our choice of encoding leads to pixels being sampled from images at the vertices of $2^{n_{\text{rad}}}$ regular $2^{n_{\text{orb}}}$-gons (top row), with examples of the resulting reconstructed images for a 128×128 pixel image with $2^{n_{\text{orb}}} = 8$ (i.e., octagons) and $16 \leq 2^{n_{\text{rad}}} \leq 2048$ (bottom row).

where $\mathcal{E}(\mathbf{x})$ denotes the data encoding unitary for input $\mathbf{x}$, and $U_{\boldsymbol{\theta}}$ the variational component evaluated with parameters $\boldsymbol{\theta} \in \Theta$. Now, if $\mathcal{G} : \mathcal{X} \to \mathcal{X}$ is a symmetry of the data at the level of the classification labels, i.e., $y[g(\mathbf{x})] = y(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, $g \in \mathcal{G}$ [but *not* necessarily $g(\mathbf{x}) = \mathbf{x}$], then we wish to build models with inductive biases that automatically enforce respect for this symmetry, i.e., $\hat{y}_{\boldsymbol{\theta}}(\mathbf{x}) = \hat{y}_{\boldsymbol{\theta}}[g(\mathbf{x})]$ for all $\mathbf{x}, \boldsymbol{\theta}$. The encoding map $\mathcal{E} : \mathcal{X} \to \mathcal{H}$ induces a unitary representation R of the symmetry group $\mathcal{G}$ on $\mathcal{H}$, making the following diagram commute:

$$
\begin{array}{ccc}
\mathcal{X} & \xrightarrow{\ \ g\ \ } & \mathcal{X} \\
\Big\downarrow{\scriptstyle \mathcal{E}} & & \Big\downarrow{\scriptstyle \mathcal{E}} \\
\mathcal{H} & \xrightarrow{\ R(g)\ } & \mathcal{H} .
\end{array}
$$

The prediction of the model on a data point acted upon by $g \in \mathcal{G}$ becomes

$$
\hat{y}_{\boldsymbol{\theta}}(g(\mathbf{x})) = \arg\max_{j} \langle \psi(\mathbf{x})| R_g^{\dagger} U_{\boldsymbol{\theta}}^{\dagger} M_j U_{\boldsymbol{\theta}} R_g |\psi(\mathbf{x})\rangle \qquad (2)
$$

with $|\psi(\mathbf{x})\rangle = \mathcal{E}(\mathbf{x})|0\rangle$ the encoded state of the data. For the predictions of our model to be invariant with respect to the action of the symmetry, then, equating Eqs. (1) and (2) we find a condition for equivariance,

$$
[R(g), U_{\boldsymbol{\theta}}^{\dagger} M_j U_{\boldsymbol{\theta}}] = 0
$$
$$
\text{for all } g \in \mathcal{G}, \ \boldsymbol{\theta} \in \Theta, \ j \in \{1, \ldots, n_{\text{classes}}\}, \qquad (3)
$$

i.e., $U_{\boldsymbol{\theta}}^{\dagger} M_j U_{\boldsymbol{\theta}} \in \mathfrak{comm}(R)$, and the representation of the symmetry group enforces constraints on the architecture of the model. The subspace $\mathfrak{comm}(R)$ will depend on the nature of the encoding map $\mathcal{E}$, to which we now turn for our specific application.

Our encoding strategy $\tilde{\mathcal{E}}$ is a slight modification of the standard amplitude encoding method. For an image, this would typically involve flattening the image into a one-dimensional vector $\mathbf{x}$ and then mapping the elements of that vector to the amplitudes of a set of basis states, i.e., $\mathbf{x} \mapsto 1/\|\mathbf{x}\| \sum_i x_i |i\rangle$. This, however, would lead to complicated and nonlocal elements $R(g)$ of the representation of the symmetry group. In principle, this is not a fundamental issue, and one could construct an equivariant model via (for example) "twirling" (i.e., explicitly projecting onto the invariant subspaces of the representation) [21], but here we instead opt to implement an encoding method more amenable to rotational symmetry, which will allow us to easily write down an equivariant model in terms of familiar quantum gates. Expending some effort modifying the initial data encoding stage in order to simplify the equivariance constraints has previously been successfully employed in the context of reflection symmetry

[28]. Our strategy is parameterized by two hyperparameters, n_{rad} and n_{orb} (respectively the number of qubits assigned to encoding radial and orbital degrees of freedom, with $n = n_{\text{rad}} + n_{\text{orb}}$ the total number of qubits), and involves constructing $2^{n_{\text{rad}}}$ regular $2^{n_{\text{orb}}}$-gons [see Fig. 1(c) for examples with $n_{\text{orb}} = 3$ and $4 \leq n_{\text{rad}} \leq 11$]. The image is sampled from at each vertex of each polygon, and, indexing the polygons by $r \in \{0, 1, \ldots, 2^{n_{\text{rad}}} - 1\}$, and the angular coordinate describing the position on the polygon by $\phi = 2\pi k/2^{n_{\text{orb}}}$, $k \in \{0, 1, \ldots, 2^{n_{\text{orb}}} - 1\}$, the corresponding pixel values $x_{r,\phi}$ are used to construct the state

$$
|\psi(\mathbf{x})\rangle = \frac{1}{\|\mathbf{x}\|} \sum_{r,\phi} x_{r,\phi} |r, \phi\rangle . \qquad (4)
$$

With this choice of data encoding, representation $R(g)$ of the operation of rotating by $2\pi g/2^{n_{\text{orb}}}$ simply acts as

$$
R(g)|\psi(\mathbf{x})\rangle = \sum_{r,\phi} x_{r,\phi} |r, \phi + 2\pi g/2^{n_{\text{orb}}} \ (\text{mod } 2\pi)\rangle . \qquad (5)
$$

Explicitly, with respect to this basis we have (in the case $n_{\text{orb}} = 2$, for example)

$$
R(0) = \mathbb{I}_{\text{rad}} \otimes \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},
$$

$$
R(1) = \mathbb{I}_{\text{rad}} \otimes \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},
$$

$$
R(2) = \mathbb{I}_{\text{rad}} \otimes \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},
$$

$$
R(3) = \mathbb{I}_{\text{rad}} \otimes \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix},
$$

where $\mathbb{I}_r$ is the identity matrix acting on the radial degrees of freedom. Thus, the rotations act on the angular degrees of freedom as the regular representation of the additive group $\mathbb{Z}_{2^{n_{\text{orb}}}}$, and can be diagonalized by the Fourier transformation over that group [36]. With $\mathbb{Z}_{2^{n_{\text{orb}}}}$ being an Abelian group, this is simply the familiar discrete Fourier transform, and can be implemented efficiently on a quantum computer by means of the quantum Fourier transform (QFT). Operators $O \in \mathfrak{comm}(R)$ are then also block diagonalized by the QFT (in fact, fully diagonalized in this case as the irreducible representations of Abelian groups are one dimensional), and so following the initial operations

$(\mathbb{I}_{\mathrm{rad}} \otimes \mathrm{QFT}^{\dagger}_{\mathrm{orb}}) \circ \tilde{\mathcal{E}}(\mathbf{x})$ we construct our equivariant model from components that are diagonal on the angular degrees of freedom and unconstrained on the radial degrees of freedom. In this work, for simplicity, we take the diagonal component to be chains of nearest-neighbor controlled-Z (CZ) gates [see Fig. 1(b)] and employ a fully expressible ansatz on the radial component, More generally, one could allow gates generated by Pauli strings that contain a Z at exactly k locations (all in the angular register), and I everywhere else, increasing the dimension of the dynamical Lie algebra (DLA) by $\binom{n_{\mathrm{orb}}}{k}$, i.e., a polynomial of degree k in n_{orb}. All such gates are diagonal in the computational basis, thus preserving the symmetry properties of the model. Allowing parameterized Z rotations on the angular register, for example, would only linearly increase $\dim \mathfrak{g}$, while simultaneously allowing strings with all values of $0 \leq k \leq n_{\mathrm{orb}}$ would constitute an increase of $2^{n_{\mathrm{orb}}}$, allowing for full expressibility while remaining symmetry preserving. Alternately, in the other direction, one could decrease $\dim \mathfrak{g}$ by introducing additional constraints on the radial register. A potentially significant drawback of our encoding scheme, requiring as it does (in general) the preparation of highly entangled states, is that the time required to do so scales exponentially with the number of qubits. Moreover, as amplitude encoding creates a state of the same dimension as the input data vector, which we can presumably store and handle classically, simulating the entire quantum circuit is at most within a polynomial factor of what can be done with classical resources. With respect to the time taken necessary for state preparation, however, the situation may in practice be considerably more favourable than the worst case, as the data that one is interested in classifying can be highly structured (indeed, by assumption in the case considered here, containing nontrivial symmetries), potentially significantly lessening the difficulty of constructing the corresponding states. State preparation techniques can, for example, prepare approximations to amplitude encoded images with sufficiently high fidelity to allow for machine learning using drastically shallower circuits than a direct implementation [37]. Nonetheless, it would be interesting to attempt to generalize the techniques presented here to include encoding strategies that can be carried out more efficiently.

We now turn to the trainability of our rotationally equivariant models. Our main analytical result is that, under some conditions, the models are free from barren plateaus. These results utilize the recent insights of Ragone *et al.* [18] and Fontana *et al.* [19], who analyzed the mean and variance of the cost function of QML models by studying the Lie closure of the generators of the circuit. In brief, given a layered quantum circuit of the form

$$\mathcal{U}(\boldsymbol{\theta}) = \prod_{i=1}^{L} U(\boldsymbol{\theta}_i) = \prod_{i=1}^{L}\prod_{j=1}^{J} e^{-i\theta_{ij}H_j}, \qquad (6)$$

one can relate the cost function mean and variance to the dimension of the *dynamical Lie algebra* $\mathfrak{g} = \mathrm{span}\,\langle iH_0, iH_1, \ldots, iH_J\rangle_{\mathrm{Lie}}$, where $\langle \mathcal{G}\rangle_{\mathrm{Lie}}$ is the set formed from repeated nested commutators of the elements of $\mathcal{G}$. In particular, if $\dim \mathfrak{g}$ grows exponentially with the number of qubits then the model necessarily exhibits barren plateaus [18,19]. Intuitively, the relevance of the nested commutators of the generators can be seen by applying the Baker-Campbell-Hausdorff formula to Eq. (6). In this work we show that the dimension of the dynamical Lie algebra of our models is exponential in n_{rad}, but only linear in n_{orb}.

Proposition 1. The dynamical Lie algebra $\mathfrak{g}$ of our rotationally equivariant model has a decomposition

$$\mathfrak{g} \cong \mathfrak{su}(2^{n_{\mathrm{rad}}}) \oplus \mathfrak{su}(2^{n_{\mathrm{rad}}}) \oplus \mathbb{R}^{n_{\mathrm{orb}}+1},$$

and thus dimension

$$\dim \mathfrak{g} = 2 \cdot 4^{n_{\mathrm{rad}}} + n_{\mathrm{orb}} - 1,$$

where n_{rad} (n_{orb}) is the number of qubits used to encode the radial (angular) degrees of freedom.

The exponential dependence on n_{rad} is a result of us employing a fully expressible ansatz on the radial register; one could replace it with a variational structure possessing a DLA growing only polynomially in n_{rad}, but this constraint is not forced by rotational equivariance. Combined with the results of Ref. [18], our main result follows.
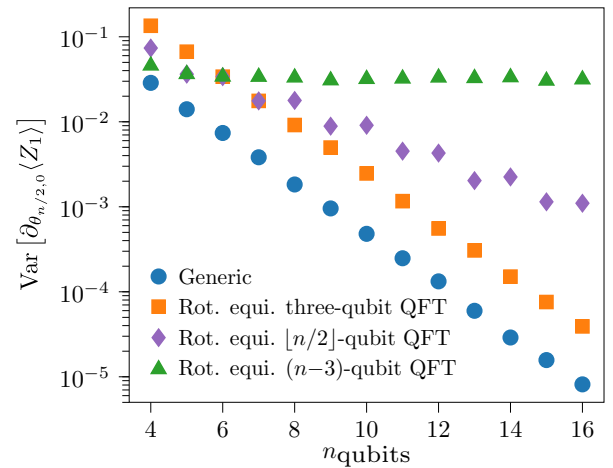

FIG. 2. Gradient scaling. The variance of the derivative of a measured observable with respect to a parameter in the middle of the model under different assignments of the qubits to encoding radial and rotational information. In all cases the variances are estimated by averaging over 1000 randomly initialized 50-layer models. When the number n_{rad} of qubits not involved in the quantum Fourier transform grows faster than $n_{\mathrm{qubits}}/2 + o(n_{\mathrm{qubits}})$, the model suffers from a barren plateau, as exhibited by the exponential vanishing of the cost function gradients.

Corollary 1. Our rotationally equivariant models are, in the absence of noise, free from barren plateaus if and only if $4^{-n_{rad}}\mathcal{P}_{\mathfrak{g}_s}(\rho) \notin \mathcal{O}(1/b^n)$ for any $b > 2$, where $\rho = |\psi(\mathbf{x})\rangle \langle \psi(\mathbf{x})|$ is the initial state and $\mathcal{P}_{\mathfrak{g}_s}(\rho)$ its purity with respect to the semisimple component of the dynamical Lie algebra $\mathfrak{g}$.

Here $\rho_{\mathfrak{g}}$ denotes the orthogonal projection of ρ onto $\mathfrak{g}$ (with respect to the Hilbert-Schmidt inner product) with purity $\mathcal{P}_{\mathfrak{g}}(\rho) := \text{tr}[(\rho_{\mathfrak{g}})^2]$. Proofs of Proposition 1 and Corollary 1 are given in Appendix A. Corollary 1 gives the constraint within which the rotationally equivariant models must operate in order to not suffer from barren plateaus, and has several immediate consequences. Firstly, regardless of the choice of n_{rad}, the purity $\mathcal{P}_{\mathfrak{g}_s}(\rho)$ cannot vanish faster than 2^{-n}. Secondly, as $\mathcal{P}_{\mathfrak{g}}(\rho) \leq 1$, the number of qubits used in the radial register is bounded by $n_{rad} \leq n/2 + o(n)$. These restrictions respectively correspond to requirements on (a generalized notion of) the initial state entanglement [15] and the circuit expressibility [14], and are linked through the unifying Lie-algebraic theory of Ref. [18]. In Fig. 2 we plot empirical results for the cost function gradient scaling under various choices of n_{rad}/n_{qubits}, the conclusions of which are consistent with our analytical findings. Under the constraints of Corollary 1, then, our rotationally equivariant models join the (short) list of QML architectures provably free from barren plateaus [23,32]. This is an extremely encouraging finding, but nonetheless the true test of machine learning models is to benchmark their performance in practice on a real dataset, which is the task to which we now turn.

III. NUMERICAL RESULTS

We test our models on a ten-class dataset consisting of STM images of phosphorus (P) donor impurities implanted in silicon (Si), with the classification task being to identify, with lattice-site precision, the location of the donor that produced the STM image (see Fig. 3). This is enabled in principle by the remarkable sensitivity of the images to the precise location of the donor in the Si lattice [35,38], and has important applications to the construction of high-fidelity spin-based quantum computers in silicon [35]. This problem has previously been tackled with machine learning approaches [39,40], under the assumption of a fixed orientation of the Si crystal with respect to the direction of motion of the STM. Relaxing that assumption leads naturally to STM images that are randomly rotated, forming the classification task we consider here.

We simulate the STM images resulting from P donors implanted at ten different positions between $3a_0$ and $4.5a_0$ below the surface (with $a_0 \approx 0.54$ nm the Si lattice constant) using a multimillion-atom tight-binding simulation implemented within NEMO-3D [41] (see Appendix B for
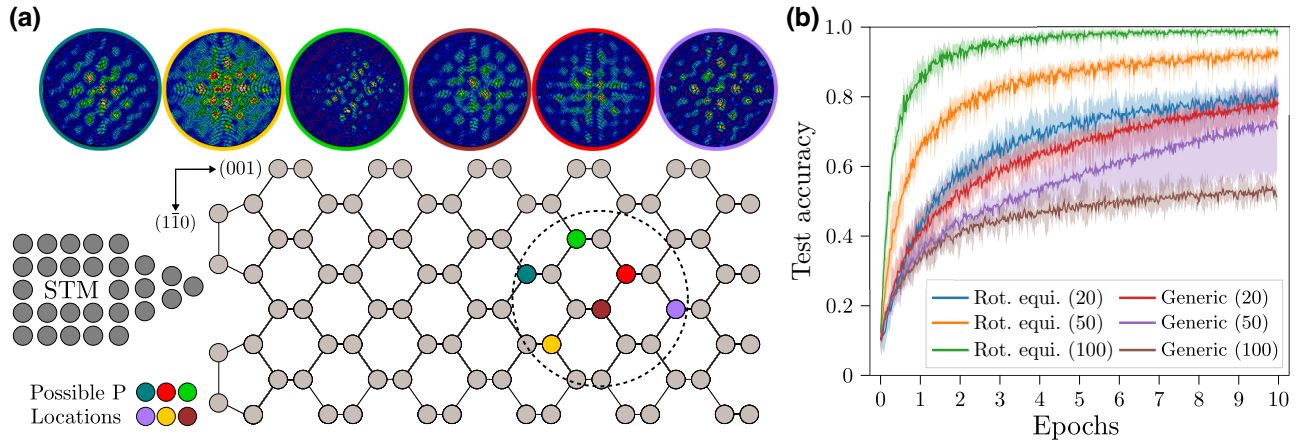

FIG. 3. Pinpointing atomic impurities with symmetry-informed QML. In order to numerically test the performance of our models in practice, we consider a classification task that involves determining the locations of phosphorus donor atoms implanted into silicon by analyzing their STM images. (a) P donors in Si create remarkably rich STM images that strongly depend on their precise locations in the lattice. Here we consider the problem of determining the correct lattice site from ten possibilities corresponding to positions between $3a_0$ and $4.5a_0$ beneath the surface (with $a_0 \approx 0.54$ nm the silicon lattice constant) given a (simulated; see Appendix B) STM image. (b) We randomly initialize and train each model configuration five times, plotting the mean accuracy achieved on a validation set of 500 examples throughout training, and shading the area between the best and worst performing of the training runs. The architecture and number of layers in each model are indicated. Our rotationally equivariant models drastically outperform their generic counterparts on this task, despite possessing fewer trainable parameters and having access to only a fraction of the total Hilbert space. Moreover, we find that increasing the depth of the generic models actually adversely affects their performance in practice, due to the increasing issues with trainability and given only a finite set of training examples, namely 7500 STM images equally distributed amongst the classes. Although one would expect, given enough data, the accuracy of the generic models to eventually match their symmetry restricted counterparts, their decreased efficiency in doing so is a clear disadvantage in practice.

details). This results in 128×128 pixel grayscale images, to which we add Gaussian noise and rotate, before encoding them into quantum states by means of Eq. (4) with $n_{\text{rad}} = 10$ and $n_{\text{orb}} = 3$. This results in us sampling from the images at the vertices of 1024 octagons [see Fig. 1(c) for the general strategy and Fig. 3(a) for reconstructions of the resulting STM images (without noise)]. We also consider a generic model that employs standard amplitude encoding ($\mathbf{x} \mapsto 1/\|\mathbf{x}\| \sum_i x_i |i\rangle$) and a maximally expressible ansatz [see Fig. 1(b)]. The models are implemented via a noiseless statevector simulation in Pennylane [42] and trained with the ADAM optimizer [43] (with a learning rate of 1×10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\epsilon = 1 \times 10^{-8}$) on a dataset of 7500 images over ten epochs, with their accuracy on a validation set of 500 examples recorded throughout the training process. The loss function $\ell_{\boldsymbol{\theta}}$ for an input state $\rho = |\psi(\mathbf{x})\rangle \langle \psi(\mathbf{x})|$ with true label $y \in \{1, 2, \ldots, n_{\text{classes}}\}$ is given by

$$\ell_{\boldsymbol{\theta}}(\rho, Z_y) = -\text{tr}[U_{\boldsymbol{\theta}} \rho U_{\boldsymbol{\theta}}^{\dagger} Z_y], \tag{7}$$

where Z_y denotes the Pauli-Z gate acting on the yth qubit. The results, for models with 20, 50, and 100 layers, are plotted in Fig. 3(b). We train each model five times from different random initializations, and plot the average over the training runs. The shaded areas are bounded by the best and worst of the runs. We find that the rotationally equivariant models drastically outperform their generic counterparts, which suffer from trainability issues that only worsen within increasing model depth. Combined with our analytical results, we conclude that these models offer a significant advantage over a generic ansatz for two-dimensional data with rotational symmetry.

IV. SUMMARY AND OUTLOOK

Much as early classical neural networks suffered from significant trainability issues [16] that were solved only after considerable experimentation and a sequence of increasingly optimized architectural decisions, early QML models based on generic circuits look likely to be replaced by carefully designed constructions. In the absence of large-scale fault-tolerant quantum computers on which to easily experiment with different types of QML architectures in the many-qubit regime, symmetry principles represent a promising guide to current research [20–30] that has now led to the discovery of multiple architectures that are provably trainable [23,32], including in this work. Our results have been facilitated by the increased understanding of the nature of QML models afforded by their rapidly developing Lie-algebraic theory [18,19], through which we can study the complexity of our models as they interpolate between the trainable and untrainable regimes, as characterized by the dimensionality of the dynamical Lie algebra.

In future work, it would be extremely interesting to explore generalizations of the QFT-based technique introduced here to other symmetry groups, which in many cases admit Fourier transforms that can be efficiently evaluated on quantum computers [36]. With the encoding set up so that the induced representation on the Hilbert space is given as in this work by the regular representation, it would again be (block) diagonalized by the corresponding Fourier transform. In the case of a non-Abelian group nontrivial blocks will appear, and the set of allowable equivariant operators in the variational ansatz will no longer be arbitrary diagonal operators, but instead (as a result of Schur's lemma) those operators that are multiples of the identity on each of the irreducible components in the decomposition of the regular representation. For any Abelian symmetry group, arbitrary diagonal operators will continue to be equivariant. More generally, we conjecture that the deep connections between Fourier analysis and representation theory may lead to some very interesting variational algorithms—QFTs play a central role in many fault-tolerant quantum algorithms with provable speed ups; by connecting these to the representation theory (and therefore symmetries) of groups, it may be possible to convert these algorithms into geometric QML algorithms that also enjoy provable speed ups.

Finally, we note that imposing restrictions on the expressibility of QML models introduces the potential for efficient classical simulations, which can become possible, for example, in the case of a small dynamical Lie algebra [44,45], or in the presence of sufficiently constraining symmetry [46]. The compromises that can be explored through families of models such as these, which can interpolate between easily trainable but efficiently classically simulable circuits through to classically intractable but barren plateau ridden circuits, then constitute an interesting research direction in the ongoing search for practical quantum advantage in machine learning.

The code that supports the findings of this article is available online [47].

ACKNOWLEDGMENTS

APPENDIX A

In this appendix we prove Proposition 1 and Corollary 1 from the main text.

Proposition 1. The dynamical Lie algebra $\mathfrak{g}$ of our rotationally equivariant model has a decomposition

$$\mathfrak{g} \cong \mathfrak{su}(2^{n_{\text{rad}}}) \oplus \mathfrak{su}(2^{n_{\text{rad}}}) \oplus \mathbb{R}^{n_{\text{orb}}+1},$$

and thus dimension

$$\dim \mathfrak{g} = 2 \cdot 4^{n_{\text{rad}}} + n_{\text{orb}} - 1,$$

where n_{rad} (n_{orb}) is the number of qubits used to encode the radial (angular) degrees of freedom.

Proof. The DLA $\mathfrak{g}$ [17–19] of a layered quantum circuit of the form

$$\mathcal{U}(\boldsymbol{\theta}) = \prod_{i=1}^{L} U(\boldsymbol{\theta}_i) = \prod_{i=1}^{L}\prod_{j=1}^{J} e^{-i\theta_{ij} H_j} \qquad (A1)$$

is given by the Lie closure of the generators $\mathcal{G}$ of its layers,

$$\mathfrak{g} = \text{span}\,\langle i\mathcal{G}\rangle_{\text{Lie}} = \text{span}\,\langle iH_0, iH_1, \ldots, iH_J\rangle_{\text{Lie}}, \quad (A2)$$

where $\langle \mathcal{G}\rangle_{\text{Lie}}$ is the set formed from repeated nested commutators of the elements of $\mathcal{G}$. In our rotationally equivariant case generators H_j of the circuit are the operators that, when exponentiated, produce arbitrary rotations on the first n_{rad} qubits, and $\text{CZ}_{i,i+1}$ gates between nearest-neighbor qubits. We thus have $X_i, Y_i, Z_i \in \mathcal{G}$ for all $i \leq n_{\text{rad}}$, as well as $\text{cz}_{i,i+1}$, where $\exp(-i(\text{cz}_{i,i+1})) = \text{CZ}_{i,i+1}$ for all $i \leq n_{\text{qubits}} - 1$. Explicitly, in the computational basis in the two-qubit case we have

$$\text{CZ}_{1,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \exp \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i\pi \end{pmatrix}$$

$$= \exp -i(\text{cz}_{1,2}),$$

from which we can readily obtain expressions for the CZ generators in terms of Pauli operators,

$$\text{cz}_{i,i+1} = -\frac{\pi}{4}(\mathbb{I} - Z_i - Z_{i+1} + Z_i Z_{i+1}).$$

We therefore wish to determine the dimension of the vector space spanned by the Lie closure of the set

$$\{X_i, Y_i, Z_i\}_{i=1}^{i=n_{\text{rad}}} \cup \{\mathbb{I} - Z_i - Z_{i+1} + Z_i Z_{i+1}\}_{i=1}^{i=n_{\text{qubits}}-1}.$$

First we restrict our attention to $i \leq n_{\text{rad}}$. We have

$$[X_i, \mathbb{I} - Z_i - Z_{i+1} + Z_i Z_{i+1}] = 2i(Y_i - Y_i Z_{i+1}),$$
$$[Y_i, \mathbb{I} - Z_i - Z_{i+1} + Z_i Z_{i+1}] = -2i(X_i - X_i Z_{i+1}).$$

So $\{X_i Z_{i+1}, Y_i Z_{i+1}\}_{i \leq n_{\text{rad}}} \subset \mathfrak{g}$. By commuting these terms with appropriate single-qubit operators we can generate all two-qubit nearest-neighbor operators acting within the first n_{rad} qubits. It then follows by the argument presented in Appendix I of Ref. [17] that all $4^{n_{\text{rad}}}$ of the n-body Pauli strings with support on the first n_{rad} qubits lie within the DLA [giving complete expressibility on those qubits as $\dim \mathfrak{u}(2^{n_{\text{rad}}}) = 4^{n_{\text{rad}}}$]. Furthermore, for each of these operators, we can construct (with the aid of $\text{cz}_{n_{\text{rad}},n_{\text{rad}}+1}$) a corresponding operator that also applies $Z_{n_{\text{rad}}+1}$, doubling the number of linearly independent operators we have found within $\mathfrak{g}$. Finally, the remaining $n_{\text{orb}} - 1$ generators of the CZ gates commute both amongst themselves and with all the other linearly independent operators in the DLA, bringing the total dimension to $2(4^{n_{\text{rad}}}) + n_{\text{orb}} - 1$.

We now turn to the task of finding a direct sum decomposition of $\mathfrak{g}$. Being a subalgebra of $\mathfrak{u}(2^{n_{\text{qubits}}})$, $\mathfrak{g}$ is a reductive Lie algebra, meaning that we have a (unique up to rearrangements of the terms) decomposition [48]

$$\mathfrak{g} \cong \mathfrak{g}_1 \oplus \mathfrak{g}_2 \oplus \cdots \oplus \mathfrak{g}_k, \qquad (A3)$$

where the $\mathfrak{g}_i$ are simple Lie algebras for $i < k$, and $\mathfrak{g}_k$ is the Abelian centre of $\mathfrak{g}$. We have seen that, in our case, $\mathfrak{g}$ is spanned by the $2 \cdot 4^{n_{\text{rad}}}$ Pauli strings that apply arbitrary Pauli operators on the first n_{rad} qubits and either I or Z on the $(n_{\text{rad}} + 1)$th qubit, along with the generators of the final $n_{\text{orb}} - 1$ CZ gates. Pulling out the operators $I^{\otimes(n_{\text{rad}}+1)}$ and $I^{\otimes n_{\text{rad}}} \otimes Z$ (which commute with all elements of $\mathfrak{g}$), we find that $\mathfrak{g}$ consists of an $(n_{\text{orb}} + 1)$-dimensional centre, along with the operators in the set $\text{span}\{A \otimes I \otimes I^{\otimes(n_{\text{orb}}-1)} \mid A \in \mathfrak{su}(2^{n_{\text{rad}}})\} \oplus \text{span}\{B \otimes Z \otimes I^{\otimes(n_{\text{orb}}-1)} \mid B \in \mathfrak{su}(2^{n_{\text{rad}}})\}$. This does not quite give us a decomposition of the form of Eq. (A3), however, as the second of those summands does not form a Lie subalgebra of $\mathfrak{g}$. Equivalently, however, we can consider the set

$$\mathfrak{g}_+ \oplus \mathfrak{g}_-$$
$$:= \text{span}\left\{A \otimes \frac{(I+Z)}{\sqrt{2}} \otimes I^{\otimes(n_{\text{orb}}-1)} \,\middle|\, A \in \mathfrak{su}(2^{n_{\text{rad}}})\right\}$$
$$\oplus \text{span}\left\{B \otimes \frac{(I-Z)}{\sqrt{2}} \otimes I^{\otimes(n_{\text{orb}}-1)} \,\middle|\, B \in \mathfrak{su}(2^{n_{\text{rad}}})\right\},$$

which contains the same operators and is written in the form of a direct sum of two ideals of $\mathfrak{g}$ [both of which are

isomorphic to $\mathfrak{su}(2^{n_{\mathrm{rad}}})$]. We therefore have the decomposition

$$
\begin{aligned}
\mathfrak{g} &\cong \mathfrak{g}_+ \oplus \mathfrak{g}_- \oplus \mathfrak{g}_k \\
&\cong \mathfrak{su}(2^{n_{\mathrm{rad}}}) \oplus \mathfrak{su}(2^{n_{\mathrm{rad}}}) \oplus \mathbb{R}^{n_{\mathrm{orb}}+1},
\end{aligned}
\tag{A4}
$$

which is what we wanted to show. ∎

Our next result makes use of the notion of the $\mathfrak{g}$-purity $\mathcal{P}_{\mathfrak{g}}(O) = \mathrm{tr}O_{\mathfrak{g}}^2$ of an operator O with respect to an operator subalgebra $\mathfrak{g}$ [18,49], with $O_{\mathfrak{g}}$ the orthogonal (with respect to the Hilbert-Schmidt norm) projection of O onto $\mathfrak{g}$.

Corollary 2. Our rotationally equivariant models are, in the absence of noise, free from barren plateaus if and only if $4^{-n_{\mathrm{rad}}}\mathcal{P}_{\mathfrak{g}_s}(\rho) \notin \mathcal{O}(1/b^n)$ for any $b > 2$, where $\rho = |\psi(\mathbf{x})\rangle \langle \psi(\mathbf{x})|$ is the initial state and $\mathcal{P}_{\mathfrak{g}_s}(\rho)$ its purity with respect to the semisimple component of the dynamical Lie algebra $\mathfrak{g}$.

Proof. We rely heavily on Theorem 1 of Ref. [18], which states that, given a DLA $\mathfrak{g}$ in the form of Eq. (A3), an input state ρ and a loss function of the form $\ell_{\boldsymbol{\theta}}(\rho, O) = \mathrm{tr}[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^{\dagger}O]$, then, if $O \in i\mathfrak{g}$ or $\rho \in i\mathfrak{g}$, we have (in the absence of noise)

$$
\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \mathrm{tr}[\rho_{\mathfrak{g}_k} O_{\mathfrak{g}_k}]
\tag{A5}
$$

and

$$
\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = \sum_{j=1}^{k-1} \frac{\mathcal{P}_{\mathfrak{g}_j}(\rho)\mathcal{P}_{\mathfrak{g}_j}(O)}{\dim \mathfrak{g}_j}.
\tag{A6}
$$

Recalling the loss function of our rotation-equivariant models [Eq. (7)] we see that it is in the required form to apply this theorem, with $O = Z_y \in i\mathfrak{g}$. Together with the decomposition of Eq. (A4), we are ready to apply Eqs. (A5) and (A6) for a given input state $\rho = |\psi(\mathbf{x})\rangle \langle \psi(\mathbf{x})|$ with true label $y \in \{1, 2, \ldots, n_{\mathrm{classes}}\}$. Firstly, $Z_y \in i\mathfrak{su}(2^{n_{\mathrm{rad}}})$ has zero projection onto the centre of $\mathfrak{g}$, so $\mathbb{E}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, O)] = 0$. To calculate the variance of the loss function, we introduce orthonormal (with respect to the Hilbert-Schmidt inner product) bases for $\mathfrak{g}_\pm$,

$$
\left\{ 2^{-n_{\mathrm{qubits}}/2} P \otimes \frac{(I \pm Z)}{\sqrt{2}} \otimes I^{\otimes(n_{\mathrm{orb}}-1)} \right\}_{P \in P_{n_{\mathrm{rad}}} \backslash \{I^{\otimes n_{\mathrm{rad}}}\}},
\tag{A7}
$$

where $P_{n_{\mathrm{rad}}}$ is the set of n_{rad}-qubit Pauli strings. Denoting these basis operators by $\{B_j^{(\pm)}\}_{j=1}^{4^{n_{\mathrm{rad}}}-1}$, it follows that the purities of Z_y with respect to $\mathfrak{g}_\pm$ are given by

$$
\begin{aligned}
\mathcal{P}_{\mathfrak{g}_\pm}(Z_y) &= \mathrm{tr}\left[\sum_{j=1}^{4^{n_{\mathrm{rad}}}-1} \mathrm{tr}(B_j^{(\pm)\dagger}Z_y)B_j^{(\pm)} \right]^2 \\
&= \sum_{j=1}^{4^{n_{\mathrm{rad}}}-1} |\mathrm{tr}(B_j^{(\pm)\dagger}Z_y)|^2 = 2^{n_{\mathrm{qubits}}-1},
\end{aligned}
\tag{A8}
$$

where we used the fact that $\mathrm{tr}(B_j^{(\pm)\dagger}Z_y) = 0$ for $B_j^{(\pm)} \neq I^{\otimes(y-1)} \otimes Z \otimes I^{\otimes(n_{\mathrm{rad}}-y)} \otimes (I \pm Z)/2 \otimes I^{\otimes(n_{\mathrm{orb}}-1)}$. From Eq. (A6), then, we have

$$
\begin{aligned}
\mathrm{Var}_{\boldsymbol{\theta}}[\ell_{\boldsymbol{\theta}}(\rho, Z_y)] &= \sum_{j \in \{+,-\}} \frac{\mathcal{P}_{\mathfrak{g}_j}(\rho)\mathcal{P}_{\mathfrak{g}_j}(Z_y)}{\dim \mathfrak{g}_j} \\
&= \frac{2^{n_{\mathrm{qubits}}-1}}{4^{n_{\mathrm{rad}}}-1} \sum_{j \in \{+,-\}} \mathcal{P}_{\mathfrak{g}_j}(\rho) \\
&= \frac{2^{n_{\mathrm{qubits}}-1}}{4^{n_{\mathrm{rad}}}-1} \mathcal{P}_{\mathfrak{g}_s}(\rho),
\end{aligned}
\tag{A9}
$$

where $\mathfrak{g}_s$ is the semisimple component of the dynamical Lie algebra $\mathfrak{g}$. This expression vanishes exponentially in n_{qubits} if and only if $4^{-n_{\mathrm{rad}}}\mathcal{P}_{\mathfrak{g}_s}(\rho) \in \mathcal{O}(1/b^{n_{\mathrm{qubits}}})$ for some $b > 2$. ∎

APPENDIX B

In this appendix we discuss the relevant details of the simulations that produced the STM images considered in this work. The physical principle upon which STMs rely is quantum tunneling [50]. A metallic tip is swept across the surface to be imaged (at a height of the order of about 1 nm), and for a given bias voltage V between the tip and surface, the current I caused by electrons tunneling from the surface to the tip is measured, producing a two-dimensional image of the surface (constant-height mode). Alternatively, one can adjust the height of the tip during the sweep so as to maintain a constant tunneling current and an image can be inferred from the recorded heights (constant-current mode). The tunneling current is given by Bardeen's tunneling formula [51],

$$
\begin{aligned}
I(V) = \frac{2\pi e}{\hbar} \sum_{\mu\nu} &f(E_\mu)[1 - f(E_\nu + eV)]|M_{\mu\nu}|^2 \\
&\times \delta(E_\mu - E_\nu - eV),
\end{aligned}
\tag{B1}
$$

where the E_μ (E_ν) are the eigenenergies of the sample (tip), f the Fermi-Dirac distribution function, and $M_{\mu\nu}$ the

tunneling matrix elements,

$$M_{\mu\nu} = -\frac{\hbar^2}{2m} \int_{\Sigma} [\chi_\nu^*(\mathbf{r})\nabla\psi_\mu(\mathbf{r}) - \psi_\mu(\mathbf{r})\nabla\chi_\nu^*(\mathbf{r})] \cdot d\mathbf{S},$$

(B2)

where Σ is a surface separating the tip from the surface being imaged. Expanding the tip wave function in terms of the spherical harmonics Y_{lm} and the spherical modified Bessel functions of the second kind k_l gives, in the notation of Ref. [52],

$$\chi_\nu(\mathbf{r}) = \sum_\beta C_{\nu\beta} k_\beta[\kappa_\nu(\mathbf{r} - \mathbf{r}_0)] Y_\beta(\theta, \phi)$$

(B3)

with κ_ν the vacuum decay constant and $\beta \in \{s, p_x, p_y, p_z, d_{xy}, d_{yz}, d_{z^2-x^2-y^2}, d_{xz}, d_{x^2-y^2}\}$. This decomposition yields an expression for $M_{\mu\nu}$ in terms of differential operators acting on the surface wave function due to Chen [53],

$$|M_{\mu\nu}|^2 \propto \left| \sum_\beta C_{\nu\beta} \hat{D}_\beta \psi_\mu(\mathbf{r}_0) \right|^2,$$

(B4)

where the $\hat{D}_\beta$ are differential operators. It is known that, for the Si:P system considered here, the dominant contribution comes from the $d_{z^2-x^2-y^2}$ orbital [35,38], with corresponding differential operator [53]

$$\hat{D}_{z^2-x^2-y^2} = \frac{2}{3}\frac{\partial^2}{\partial z^2} - \frac{1}{3}\frac{\partial^2}{\partial x^2} - \frac{1}{3}\frac{\partial^2}{\partial y^2}.$$

The remaining challenge is to calculate the ground-state wave function of the P impurity, which we do via a multimillion-atom $sp^3d^5s^*$ atomistic tight-binding calculation implemented with the NEMO–3D system [41,54]. The tight-binding parameters have previously been fitted to accurately reproduce the donor energy spectrum [55] and the Si band structure [54], and is capable of accurately predicting important physical quantities, including the Stark and strain-induced hyperfine shift [56] and the anisotropic electron g factor in strained Si [57]. The calculations were performed over a $(40 \text{ nm})^3$ domain consisting of around 3.1 million atoms, and include the effects of central cell corrections [55], the formation of Si dimer rows on the surface due to the 2×1 surface reconstruction of Si [58], and hydrogen passivation [59]. The agreement that has been demonstrated between experimental STM images and images calculated using this model is remarkable [35].

---

[1] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, The future of quantum computing with superconducting qubits, J. Appl. Phys. **132**, 160902 (2022).

[2] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[3] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, Nat. Commun. **11**, 1 (2020).

[4] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature **567**, 209 (2019).

[5] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[6] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).

[7] M. T. West, S. L. Tsang, J. S. Low, C. D. Hill, C. Leckie, L. C. L. Hollenberg, S. M. Erfani, and M. Usman, Towards quantum enhanced adversarial robustness in machine learning, Nat. Mach. Intell. **5**, 581 (2023).

[8] S. L. Tsang, M. T. West, S. M. Erfani, and M. Usman, Hybrid quantum-classical generative adversarial network for high resolution image generation, IEEE Trans. Quantum Eng. **4**, 3102419 (2023).

[9] M. Schuld, Supervised quantum machine learning models are kernel methods, ArXiv:2101.11020.

[10] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, Nat. Phys. **17**, 1013 (2021).

[11] H. Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, Science **376**, 1182 (2022).

[12] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 1 (2018).

[13] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 1 (2021).

[14] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[15] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1 (2021).

[16] Y. Bengio, P. Simard, and P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Netw. **5**, 157 (1994).

[17] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Diagnosing barren plateaus with tools from quantum optimal control, Quantum **6**, 824 (2022).

[18] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. O. Marrero, M. Larocca, and M. Cerezo, A unified theory of barren plateaus for deep parametrized quantum circuits, ArXiv:2309.09342.

[19] E. Fontana, D. Herman, S. Chakrabarti, N. Kumar, R. Yalovetzky, J. Heredge, S. H. Sureshbabu, and M. Pistoia, The adjoint is all you need: Characterizing barren plateaus in quantum ansätze, ArXiv:2309.07902.

[20] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, Exploiting symmetry in variational quantum machine learning, PRX Quantum **4**, 010328 (2023).

[21] M. Ragone, P. Braccia, Q. T. Nguyen, L. Schatzki, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo, Representation theory for geometric quantum machine learning, ArXiv:2210.07980.

[22] Q. T. Nguyen, L. Schatzki, P. Braccia, M. Ragone, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo, Theory for equivariant quantum neural networks, PRX Quantum **5**, 020328 (2024).

[23] L. Schatzki, M. Larocca, Q. T. Nguyen, F. Sauvage, and M. Cerezo, Theoretical guarantees for permutation-equivariant quantum neural networks, npj Quantum Inf. **10**, 12 (2024).

[24] F. Sauvage, M. Larocca, P. J. Coles, and M. Cerezo, Building spatial symmetries into parameterized quantum circuits for faster training, Quantum Sci. Technol. **9**, 015029 (2024).

[25] A. Skolik, M. Cattelan, S. Yarkoni, T. Bäck, and V. Dunjko, Equivariant quantum circuits for learning on weighted graphs, npj Quantum Inf. **9**, 47 (2023).

[26] M. Larocca, F. Sauvage, F. M. Sbahi, G. Verdon, P. J. Coles, and M. Cerezo, Group-invariant quantum machine learning, PRX Quantum **3**, 030341 (2022).

[27] H. Zheng, G. Kang, G. S. Ravi, H. Wang, K. Setia, F. T. Chong, and J. Liu, in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2023), Vol. 1, p. 236.

[28] M. T. West, M. Sevior, and M. Usman, Reflection equivariant quantum neural networks for enhanced image classification, Mach. Learn.: Sci. Technol. **4**, 035027 (2023).

[29] S. Y. Chang, M. Grossi, B. L. Saux, and S. Vallecorsa, in *2023 International Conference on Quantum Computing and Engineering* (IEEE, 2023).

[30] R. D. East, G. Alonso-Linaje, and C. Y. Park, All you need is spin: SU(2) equivariant variational quantum circuits based on spin networks, ArXiv:2309.07250.

[31] J. Heredge, C. Hill, L. Hollenberg, and M. Sevior, Permutation invariant encodings for quantum machine learning with point cloud data, Quantum Machine Intelligence **6**, 1 (2024).

[32] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of barren plateaus in quantum convolutional neural networks, Phys. Rev. X **11**, 041011 (2021).

[33] N. Diaz, D. García-Martín, S. Kazi, M. Larocca, and M. Cerezo, Showcasing a barren plateau theory beyond the dynamical Lie algebra, ArXiv:2310.11505.

[34] R. Wiersema, E. Kökcü, A. F. Kemper, and B. N. Bakalov, Classification of dynamical Lie algebras for translation-invariant 2-local spin systems in one dimension, ArXiv:2309.05690.

[35] M. Usman, J. Bocquel, J. Salfi, B. Voisin, A. Tankasala, R. Rahman, M. Simmons, S. Rogge, and L. Hollenberg, Spatial metrology of dopants in silicon with exact lattice site precision, Nat. Nanotechnol. **11**, 763 (2016).

[36] A. M. Childs and W. Van Dam, Quantum algorithms for algebraic problems, Rev. Mod. Phys. **82**, 1 (2010).

[37] M. T. West, A. C. Nakhl, J. Heredge, F. M. Creevey, L. C. Hollenberg, M. Sevior, and M. Usman, Drastic circuit depth reductions with preserved adversarial robustness by approximate encoding for quantum machine learning, ArXiv:2309.09424.

[38] M. T. West and M. Usman, Influence of sample momentum space features on scanning tunnelling microscope measurements, Nanoscale **13**, 16070 (2021).

[39] M. Usman, Y. Z. Wong, C. D. Hill, and L. C. Hollenberg, Framework for atomic-level characterisation of quantum computer arrays by machine learning, npj Comput. Mater. **6**, 19 (2020).

[40] M. T. West and M. Usman, Framework for donor-qubit spatial metrology in silicon with depths approaching the bulk limit, Phys. Rev. Appl. **17**, 024070 (2022).

[41] G. Klimeck, S. S. Ahmed, H. Bae, N. Kharche, S. Clark, B. Haley, S. Lee, M. Naumov, H. Ryu, F. Saied, M. Prada, M. Korkusinski, T. B. Boykin, and R. Rahman, Atomistic simulation of realistically sized nanodevices using NEMO 3-D—Part I: Models and benchmarks, IEEE Trans. Electron Dev. **54**, 2079 (2007).

[42] V. Bergholm, *et al.*, Pennylane: Automatic differentiation of hybrid quantum-classical computations, ArXiv:1811.04968.

[43] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, ArXiv:1412.6980.

[44] M. L. Goh, M. Larocca, L. Cincio, M. Cerezo, and F. Sauvage, Lie-algebraic classical simulations for variational quantum computing, ArXiv:2308.01432.

[45] M. Cerezo, M. Larocca, D. García-Martín, N. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing, ArXiv:2312.09121.

[46] E. R. Anschuetz, A. Bauer, B. T. Kiani, and S. Lloyd, Efficient classical algorithms for simulating symmetric quantum systems, Quantum **7**, 1189 (2023).

[47] https://github.com/maxwest97/rotation-equivariant-qnn.

[48] A. W. Knapp and A. W. Knapp, *Lie Groups Beyond an Introduction* (Springer, 1996), Vol. 140.

[49] R. Somma, G. Ortiz, H. Barnum, E. Knill, and L. Viola, Nature and measure of entanglement in quantum phase transitions, Phys. Rev. A **70**, 042311 (2004).

[50] C. J. Chen, *Introduction to Scanning Tunneling Microscopy* (Oxford University Press, USA, 2021), Vol. 69.

[51] J. Bardeen, Tunnelling from a many-particle point of view, Phys. Rev. Lett. **6**, 57 (1961).

[52] G. Mándi and K. Palotás, Chen's derivative rule revisited: Role of tip-orbital interference in STM, Phys. Rev. B **91**, 165406 (2015).

[53] C. J. Chen, Tunneling matrix elements in three-dimensional space: The derivative rule and the sum rule, Phys. Rev. B **42**, 8841 (1990).

[54] T. B. Boykin, G. Klimeck, and F. Oyafuso, Valence band effective-mass expressions in the $sp^3d^5s^*$ empirical tight-binding model applied to a Si and Ge parametrization, Phys. Rev. B **69**, 115201 (2004).

[55] M. Usman, R. Rahman, J. Salfi, J. Bocquel, B. Voisin, S. Rogge, G. Klimeck, and L. L. C. Hollenberg, Donor hyperfine Stark shift and the role of central-cell corrections in tight-binding theory, J. Phys.: Condens. Matter **27,** 154207 (2015).

[56] M. Usman, C. D. Hill, R. Rahman, G. Klimeck, M. Y. Simmons, S. Rogge, and L. C. Hollenberg, Strain and electric field control of hyperfine interactions for donor spin qubits in silicon, Phys. Rev. B **91,** 245209 (2015).

[57] M. Usman, H. Huebl, A. Stegner, C. Hill, M. Brandt, and L. Hollenberg, Measurements and atomistic theory of electron g-factor anisotropy for phosphorus donors in strained silicon, Phys. Rev. B **98,** 035432 (2018).

[58] B. Craig and P. Smith, The structure of the Si(100)2 × 1: H surface, Surf. Sci. **226,** L55 (1990).

[59] S. Lee, F. Oyafuso, P. Von Allmen, and G. Klimeck, Boundary conditions for the electronic structure of finite-extent embedded semiconductor nanostructures, Phys. Rev. B **69,** 045316 (2004).