Capacity and quantum geometry of parametrized quantum circuits

Tobias Haug,[1,*] Kishor Bharti,[2] and M. S. Kim[1]

[1] *QOLS, Blackett Laboratory, Imperial College London SW7 2AZ, UK*
[2] *Centre for Quantum Technologies, National University of Singapore 117543, Singapore*

To harness the potential of noisy intermediate-scale quantum devices, it is paramount to find the best type of circuits to run hybrid quantum-classical algorithms. Key candidates are parametrized quantum circuits that can be effectively implemented on current devices. Here, we evaluate the capacity and trainability of these circuits using the geometric structure of the parameter space via the effective quantum dimension, which reveals the expressive power of circuits in general as well as of particular initialization strategies. We assess the representation power of various popular circuit types and find striking differences depending on the type of entangling gates used. Particular circuits are characterized by scaling laws in their expressiveness. We identify a transition in the quantum geometry of the parameter space, which leads to a decay of the quantum natural gradient for deep circuits. For shallow circuits, the quantum natural gradient can be orders of magnitude larger in value compared to the regular gradient; however, both of them can suffer from vanishing gradients. By tuning a fixed set of circuit parameters to randomized ones, we find a region where the circuit is expressive, but does not suffer from barren plateaus, hinting at a good way to initialize circuits. Our results enhance the understanding of parametrized quantum circuits for improving variational quantum algorithms.

Quantum computers promise to tackle challenging problems for classical computers such as drug design, combinatorial optimisation and simulation of many-body physics. While fully-fledged large-scale quantum computers with error correction are not expected to be available for many years, noisy intermediate-scale quantum (NISQ) devices have been investigated as a way to approach computationally hard problems with quantum processors available now and in the near future [1, 2]. Variational quantum algorithms (VQA) [3–6] have been a major hope in achieving a quantum speedup with NISQ devices. The core idea is to update a parametrized quantum circuit (PQC) in a hybrid quantum-classical fashion. Measurements performed on the PQC are fed into a classical computer to propose a new set of variational parameters. A key challenge has been the occurrence of barren plateaus, i.e. the gradients used for optimisation vanish exponentially with increasing number of qubits [7], as well as for various types of cost functions [8], entanglement [9] and noise [10]. Further, the classical optimization part of variational algorithms was shown to be NP-hard [11]. Quantum algorithms that avoid the feed-back loop to circumvent the barren plateau problems have been proposed [12–17]. Besides this approach, initialization strategies [18–20] and layer-wise learning [21] for VQA could help to solve the aforementioned problems. However, tools to evaluate the power of these strategies are lacking. Hardware efficient ansätze have been proposed to tailor a PQC to the restrictions of the hardware [22]. A widely used choice is quantum circuits arranged in layers of single-qubit rotations followed by two-qubit entangling gates. However, a key question is the space of possible states this ansatz type can represent [23, 24].

Here, we introduce the effective quantum dimension G_C and parameter dimension D_C as a quantitative measure of the capacity of a PQC. Parameter dimension D_C measures the total number of independent parameters a quantum state defined by the PQC can express. In contrast, the effective quantum dimension G_C [25, 26] is a local measure to quantify the space of states that can be accessed by locally perturbing the parameters of the PQC. Both measures can be derived from the quantum geometric structure of the PQC via the quantum Fisher information metric (QFI) $\mathcal{F}$ [27, 28]. From the QFI, one can obtain the quantum natural gradient (QNG) for a more efficient optimisation via gradients [27–29]. These methods allow us to evaluate the expressive power, trainability and number of redundant parameters of different PQCs, and find better initialization strategies.

As demonstration of our tools, we provide an in-depth investigation of popular hardware-efficient circuits, composed of layered single-qubit rotations and two-qubit entangling gates in various arrangements. We find striking differences depending on the choice of circuit structure that affect both the expressive power of the PQC in general as well as the quality of specific initialization strategies. We calculate the number of redundant parameters of various PQC types, as well as how fast they converge towards random quantum states as a function of the number of layers. The choice of entangling gate has a pronounced effect on the representation power of particular initialization strategies.

We reveal a transition in the spectrum of the QFI in deep circuits, which leads to a decay of the QNG. For shallow circuits, the QNG can be orders of magnitude larger than the regular gradient. However, both suffer from the barren plateau problem. By tuning the PQCs parameters from zero to a random set of parameters, we find a region where both large gradients and large effective quantum dimension G_C coexist, which could serve

* thaug@ic.ac.uk

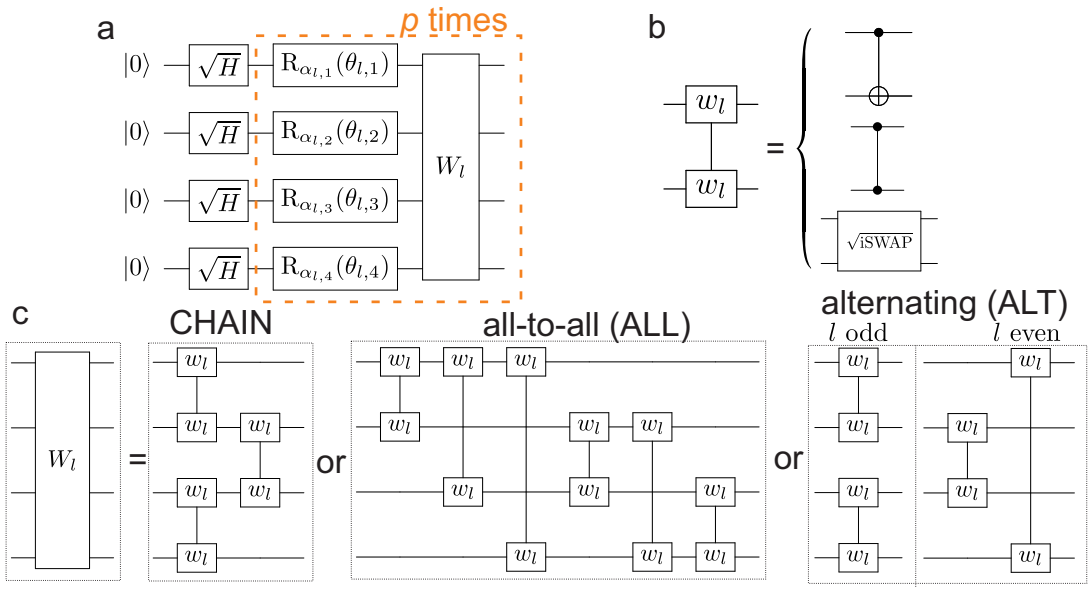

FIG. 1. **a)** Sketch of hardware-efficient parametrized quantum circuit (PQC) $U(\theta)|0\rangle^{\otimes N} = \prod_{l=p}^{1} [W_l V_l(\theta_l)] \sqrt{H_\mathrm{d}}^{\otimes N} |0\rangle^{\otimes N}$ with parameters θ and initial state $|0\rangle$ of all N qubits being in state zero. The PQC consists of an initial layer of $\sqrt{H_\mathrm{d}}$ gates applied to each qubit, where H_d is the Hadamard gate, followed by p repeated layers of parametrized single qubit rotations $V_l(\theta_l)$ and entangling gates W_l. $V_l(\theta_l)$ consists of single-qubit rotations $R_\alpha(\theta_{l,n}) = \exp(-i\sigma_n^\alpha \theta_{l,n}/2)$ at layer l and qubit n around axis $\alpha \in \{x, y, z\}$. **b)** Two-qubit entangling gates w_l considered are CNOT gates (control-σ_x), CPHASE gates (control-σ_z, diag$(1, 1, 1, -1)$) or $\sqrt{\mathrm{iSWAP}}$ gates. **c)** Entangling layer W_l is composed of the two-qubit entangling gates w_l, which are arranged in either a nearest-neighbor one-dimensional chain topology (denoted as CHAIN), all-to-all connection (ALL) or in a alternating fashion (ALT) for even and odd layers l.

as a good set of initial parameters for the training of variational algorithms.

I. PARAMETRIZED QUANTUM CIRCUITS

A PQC generates a quantum state of N qubits

$$|\psi(\theta)\rangle = U(\theta)|0\rangle^{\otimes N}\,, \tag{1}$$

with the unitary $U(\theta)$, the M-dimensional parameter vector θ and product state $|0\rangle^{\otimes N}$ (see Fig.1). The structure of the PQC influences its power to represent quantum states [23, 24, 30]. One way to measure expressiveness is by determining the distance between the distribution of states generated by the circuit and the Haar random distribution of states [23, 24]. This tells us how well the PQC can represent arbitrary states across the Hilbert space. The appearance of barren plateaus or vanishing gradients is connected to the aforementioned measure [7, 31]. The variance of the gradient $\mathrm{var}(\partial_i E) = \langle(\partial_i E)^2\rangle - \langle\partial_i E\rangle^2$ ($\langle.\rangle$ denoting statistical average over many random instances) in respect to the expectation value of a local Hamiltonian H ($E = \langle 0|U^\dagger(\theta) H U(\theta)|0\rangle$) can vanish exponentially with the number of qubits for PQCs with random choice of parameters. The variance decreases also with number of layers p of the PQC until a specific p_r, where it remains constant upon further increase of $p > p_\mathrm{r}$. For local cost functions, it has been shown that in most cases low variance of the gradient of such PQCs correlates with high expressibility [31].

II. PARAMETER DIMENSION

We now introduce the parameter dimension D_C of a PQC as another measure of capacity. As introduction, we take a generic quantum state which is parametrized by in total $M = 2^N$ real and complex coefficients

$$|\psi(a, b)\rangle = \sum_{j=1}^{2^N} (a_j + i b_j)|j\rangle\,, \tag{2}$$

where $|j\rangle$ is the j-th computational basis state and $a_j, b_j \in \mathcal{R}$. One can map the above state to $D_\mathrm{C} = 2^{N+1} - 2$ independent parameters, that lie on the surface of $2^{N+1} - 1$ dimensional sphere (by taking norm and global phase of the quantum state into account). Conversely, for a generic real-valued quantum state with $b_j = 0$, we find $D_\mathrm{C} = 2^N - 1$ independent parameters. Analogous to a generic quantum state, we now define for a PQC denoted by C the parameter dimension D_C as the number of independent parameters that can be represented in the space of quantum states. In general, D_C for N qubits is upper bounded by the generic state Eq. (2).

We define the redundancy

$$R = \frac{M - D_{\mathrm{C}}}{M}\,, \tag{3}$$

which is the fraction of parameters of the PQC that do not contribute to changing the quantum state.

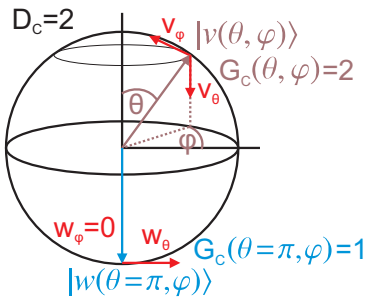

FIG. 2. Example to demonstrate the effective quantum dimension G_{C} and parameter dimension D_{C} for a single qubit parametrized as $|\psi(\theta,\varphi\rangle = \cos(\theta/2)|0\rangle + \exp(i\varphi)\sin(\theta/2)|1\rangle$. $D_{\mathrm{C}} = 2$ is the total number of independent parameters of the quantum state. G_{C} denotes the number of independent directions a quantum state can move by locally perturbing its parameters θ, φ. For a random state $|v\rangle$ ($\theta \notin \{0,\pi\}$) two possible directions exist, along v_θ and v_φ. The particular state $|w(\theta = \pi,\varphi)\rangle$ can only be perturbed in direction w_θ as adjusting φ does not change the state (e.g. $|w(\pi,\varphi+\epsilon)\rangle = |w(\pi,\varphi)\rangle$), thus $G_{\mathrm{C}} = 1$.

III. EFFECTIVE QUANTUM DIMENSION

Now, we explain how the QFI $\mathcal{F}(\theta)$ quantifies the expressive power of a PQC (see Supplemental materials B for an introduction to the QFI and QNG). One can relate $\mathcal{F}(\theta)$ to the distance in the space of pure quantum states, which is given by the Fubini-Study distance

$$\mathrm{Dist}_Q\Big(|\psi(\theta)\rangle,\ |\psi(\theta+\mathrm{d}\theta)\rangle\Big)^2 = \sum_{i,j} F_{ij}(\theta)\mathrm{d}\theta_i\mathrm{d}\theta_j\,, \tag{4}$$

where $\mathrm{Dist}_Q(x,y) = |\langle x|y\rangle|^2$ and the QFI [27, 28]

$$\mathcal{F}_{ij}(\theta) = \mathrm{Re}(\langle\partial_i\psi|\partial_j\psi\rangle - \langle\partial_i\psi|\psi\rangle\langle\psi|\partial_j\psi\rangle)\,. \tag{5}$$

$\mathcal{F}(\theta)$ quantifies the change of the quantum state when adjusting its parameter θ infinitesimally to $\theta + \mathrm{d}\theta$. The singular value decomposition

$$\mathcal{F} = VSV^T\,, \tag{6}$$

gives us V, which is a real-valued unitary with the i-th eigenvector $\alpha^{(i)}$ placed at the i-th column of V, and S, which is a diagonal matrix with the M non-negative eigenvalues $\lambda^{(i)}$ of $\mathcal{F}(\theta)$ along the diagonal. The eigenvalues and eigenvectors obey the equation $\mathcal{F}(\theta)\alpha^{(i)} = \lambda^{(i)}\alpha^{(i)}$. Inserting Eq. (6) into Eq. (4) gives us

$$\mathrm{Dist}_Q\Big(|\psi(\theta)\rangle,\ |\psi(\theta+d\theta)\rangle\Big)^2 = \mathrm{d}\theta^T\mathcal{F}d\theta = d\theta^T VSV^T\mathrm{d}\theta\,. \tag{7}$$

Now, we assume that the small variations in θ are in the direction of the i-th eigenvector of $\mathcal{F}$ with $\mathrm{d}\theta = \mathrm{d}\mu\alpha^{(i)}$, where $\mathrm{d}\mu$ is an infinitesimal scalar. We find

$$\mathrm{Dist}_Q\Big(|\psi(\theta)\rangle,\ |\psi(\theta+d\mu\alpha^{(i)})\rangle\Big)^2 = d\mu{\alpha^{(i)}}^T VSV^T\alpha^{(i)}\mathrm{d}\mu = \lambda^{(i)}\mathrm{d}\mu\mathrm{d}\mu\,,$$

where we have used $V^T\alpha^{(i)} = e^{(i)}$, where $e^{(i)}$ is the i-th basis vector. When updating $\theta' = \theta + \mathrm{d}\mu\alpha^{(i)}$, the quantum state changes at a rate that is proportional to $\lambda^{(i)}$. Eigenvalues $\lambda^{(i)} = 0$ are called singularities as there is no change in the quantum state at all, i.e. $|\langle\psi(\theta)|\psi(\theta+\mathrm{d}\mu\alpha^{(i)})\rangle| = 1$. The case $\lambda^{(i)}$ being very small, i.e. $1 \gg \lambda^{(i)} > 0$, is called near singularity and is associated with plateaus in classical machine learning where training slows down [32].

We now define the effective quantum dimension $G_{\mathrm{C}}(\theta)$ for a PQC denoted as C. It is given as the total number of non-zero eigenvalues $\lambda^{(i)}(\theta)$ of the QFI $\mathcal{F}(\theta)$ initialized with parameters θ [25, 26]

$$G_{\mathrm{C}}(\theta) = \sum_{i=1}^{M}\mathcal{I}(\lambda^{(i)}(\theta))\,, \tag{8}$$

where $\mathcal{I}(x) = 0$ for $x = 0$ and $\mathcal{I}(x) = 1$ for $x \neq 0$. $G_{\mathrm{C}}(\theta)$ is a local measure of expressiveness that counts the number of independent directions in the state space that can be accessed by an infinitesimal update of θ.

A straightforward example is a generic single qubit quantum state (see Fig.2)

$$|\psi(\theta,\varphi)\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \exp(i\varphi)\sin\left(\frac{\theta}{2}\right)|1\rangle \tag{9}$$

$$\mathcal{F} = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2(\theta) \end{bmatrix}. \tag{10}$$

The eigenvalues and eigenvectors of the QFI $\mathcal{F}$ are straightforward to calculate with $\lambda_1 = 1$, $\alpha_1 = \{1,0\}$ and $\lambda_2 = \sin^2(\theta)$, $\alpha_2 = \{0,1\}$. The effective quantum dimension $G_{\mathrm{C}}(\theta,\varphi) = D_{\mathrm{C}} = 2$, except for the special case $\theta = n\pi$, n integer, the eigenvalue is $\lambda_2 = 0$ and thus $G_{\mathrm{C}}(n\pi,\varphi) = 1$. Here, any change in the direction of eigenvector α_2 (corresponding to changing φ) will not yield any change in the underlying quantum state. However note that except for these singular parameters we find $G_{\mathrm{C}} = 2$, which is equivalent to the maximal number of independent parameters of a qubit.

Further, consider the single qubit circuit with Pauli z matrix σ_z and Hadamard gate H_{d}

$$U(\theta)|0\rangle = \prod_{i=1}^{M}\left[\exp(-i\frac{\theta_i}{2}\sigma_z)\right]H_{\mathrm{d}}|0\rangle\,. \tag{11}$$

Here, we find $\mathcal{F} = \frac{1}{4}J_{M,M}$, where $J_{M,M}$ is a $M\times M$ matrix filled with ones. A diagonalization gives us $M-1$ eigenvalues with $\lambda = 0$, and one eigenvalue $\lambda_1 = \frac{M}{4}$

with eigenvector $\alpha_1 = \frac{1}{\sqrt{M}} J_{M,1}$. This circuit has a low parameter dimension $D_C = 1$ and a large redundancy of $R = (M-1)/M$, i.e. there are $M-1$ parameter directions which do not yield any change of the quantum state.

For the type of PQC as shown in Fig.1, which are arranged in a layer-wise structure with the parametrized gates being Pauli operators, the effective quantum dimension G_C is less than the parameter dimension D_C and the total number of parameters M

$$G_C(\theta) \leq D_C \leq M\,. \tag{12}$$

Given the aforementioned PQC type with a random set of parameters $\theta_{\text{rand}} \in \text{rand}(0, 2\pi)$, we find numeric evidence that $G_C(\theta_{\text{rand}})$ is approximately equivalent to D_C

$$G_C(\theta_{\text{rand}}) \simeq D_C\,. \tag{13}$$

The core intuition is that starting from a sufficiently random initial parameter set, a change of the PQC parameters in the right direction is able to bring one closer to any quantum state that can be expressed by the PQC. For specific choices of parameters such as $\theta = 0$ we find $G_C < D_C$. Moving sufficiently away from these special points, we recover that $G_C \simeq D_C$ (see Fig.6).

We stress that Eq. (13) is not valid for arbitrary quantum circuits, e.g. circuits where the parameters do not enjoy a 2π periodicity. As simple example take the evolution of a single qubit with a single parameter t $U(t)|0\rangle = \exp(-i\sqrt{2}\sigma_z t)\exp(-i\sigma_x t)|0\rangle$. The evolution over all possible t (note the absence of 2π periodicity) will cover all possible quantum states and thus $D_C = 2$, whereas the effective quantum dimension (with only a single parameter t) is $G_C = 1 < D_C$.

We now consider different types of hardware efficient PQC $|\psi(\theta)\rangle = U(\theta)|0\rangle^{\otimes N}$, which are circuits that can be efficiently run on NISQ quantum processors. We choose an initial state $|0\rangle^{\otimes N}$, followed by a single layer of the square root of the Hadamard gate ($\sqrt{H_d}$) on every qubit. Then, we repeat p layers composed of parametrized single qubit rotations and a set of two-qubit entangling gates (see Fig.1a). The single qubit rotations are either chosen randomly to be around the $\{x, y, z\}$ axis, or fixed to a specific axis. The two-qubit entangling gates are either CNOT, CPHASE or $\sqrt{\text{iSWAP}}$ gates (see Fig.1b), that are common native gates in current quantum processors [33]. The entangling gates in each layer are arranged in either a nearest-neighbor chain topology (CHAIN), all-to-all connections (ALL) or in an alternating nearest-neighbor fashion (ALT) (see Fig.1c). The numerical calculations are performed using Yao [34].

IV. RESULTS

As a demonstration of our methods, in Fig.3 we provide an in-depth characterization of a particular PQC as function of number of layers p. Each layer consists of parametric single-qubit rotations around x, y or z axis,

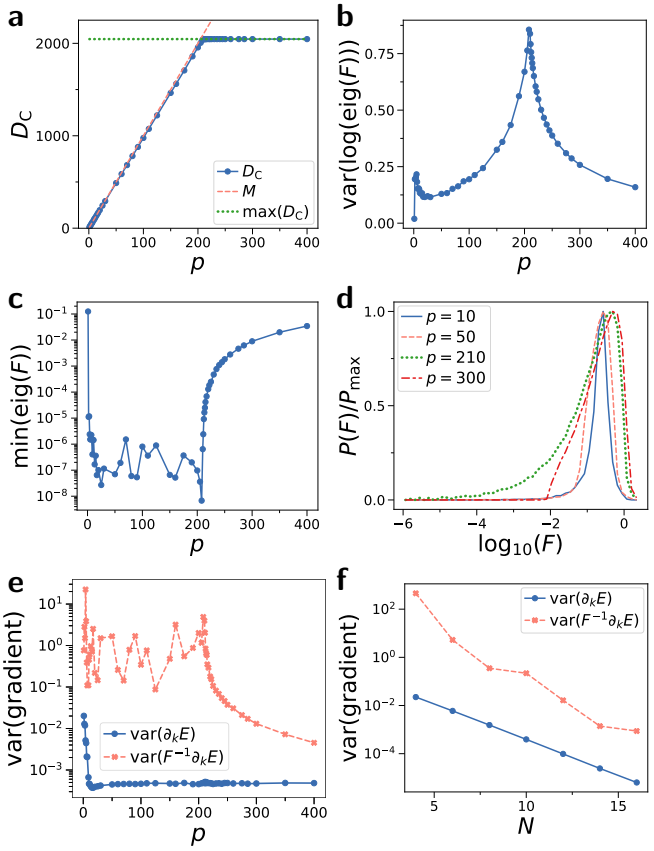

FIG. 3. Properties of PQC consisting of p layers of randomly chosen x, y, z rotations, followed by CNOT gates in a chain topology (see Fig.1) for $N = 10$ qubits. **a)** The parameter dimension D_C of the circuit scales linearly with p, until it levels at a characteristic value $p_c \approx 210$. **b)** Variance of the logarithm of the non-zero eigenvalues of $\mathcal{F}$. The variance peaks around $p \approx p_c$. **c)** Minimal non-zero eigenvalue of $\mathcal{F}$ against p. It increases for $p > p_c$. **d)** Histogram of logarithm of eigenvalues of Fisher information matrix $\mathcal{F}$. The width of the distribution increases with p, with a pronounced tail at small $\mathcal{F}$ developing around $p \approx p_c$, which disappears for $p > p_c$. **e)** Variance of the gradient var($\partial_k E$) and QNG var($\mathcal{F}^{-1}\partial_k E$) in respect to the Hamiltonian $H = \sigma_1^z \sigma_2^z$. The gradient decays until $p \approx 20$, after which it remains constant. The QNG remains larger than the regular gradient, but decreases for $p > p_c$. **f)** Variance of gradients and QNG for varying qubit number N for depth $p = 2N$, showing approximate exponential decrease with N.

which are randomly chosen at every qubit and layer as well as CNOT gates in a chain topology (see CHAIN in Fig.1). The parameter dimension D_C (i.e. number of independent parameters of the quantum state that can be represented by the PQC) increases linearly with p in Fig.3a, until it reaches the maximal possible value for $D_C = 2^{N+1} - 2$ at a characteristic number of layers p_c. This point is reflected in the spectrum of the QFI $\mathcal{F}$, averaged over random instances of the PQC (see Fig.3b-d). Most notably, the variance of the logarithm of the non-zero eigenvalues reaches a maximum for p_c (Fig.3b).

Further, the minimum taken over all eigenvalues becomes minimal (Fig.3c). We can see this more clearly in the distribution of eigenvalues (Fig.3d). With increasing p, the distribution becomes broader, with a pronounced tail of small eigenvalues of $\mathcal{F}$ appearing close to the transition at p_c. Above the transition $p > p_c$, the small eigenvalues suddenly disappear from the distribution. We investigate the variance of the gradient and QNG in Fig.3e. We note that the variance of the regular gradient decays with p, reaching a minimum around $p \approx 20$ [7], upon which it remains constant. The variance of the QNG remains larger than the regular gradient, however the QNG decays for $p > p_c$. In Fig.3f, we numerically find that variance of both regular gradient and QNG vanish exponentially with increasing number of qubits N, demonstrating the barren plateau problem.

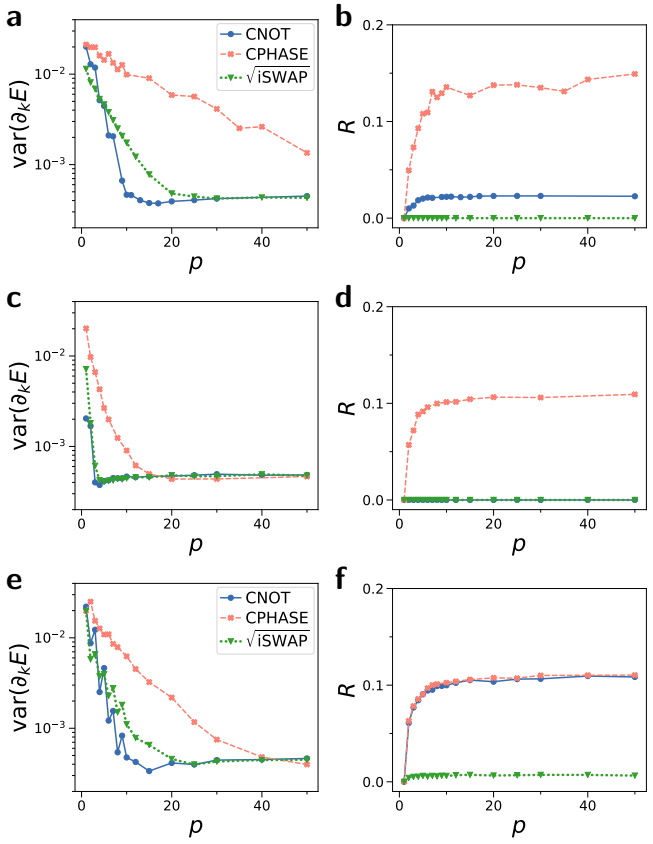

FIG. 4. Variance of gradient and redundancy of different hardware efficient PQCs plotted against number of layers p for $N = 10$ qubits. Each layer consists of single qubit-rotations as randomly chosen rotations around $\{x, y, z\}$ axis. We plot different arrangements of entangling gates (as shown in Fig.1c) with **a,b)** nearest-neighbor one-dimensional chain, **c,d)** all-to-all, **e,f)** alternating nearest-neighbor connections. **a,c,e)** Variance of the gradient $\mathrm{var}(\partial_k E)$ in respect to the Hamiltonian $H = \sigma_1^z \sigma_2^z$. **b,d,f)** Redundancy R (Eq. (3)), which is the fraction of redundant parameters of the PQC.

In Fig.4, we compare different types of PQCs with different entangling gates and arrangements. We keep the single qubit-rotations as randomly chosen rotations around $\{x, y, z\}$ axis. We note that all circuits show the same qualitative behavior regarding the transition in the QFI (see Fig.3 and supplemental materials) as well as suffer from exponential decrease of the variance of the gradient with increasing number of qubits. However, key differences in the different PQCs remain as we demonstrate below. We show the variance of the gradient for random PQC parameters in Fig.4a,c,e for different arrangements of the entangling gates (CHAIN, ALL, ALT) as well as different types of entangling gates (CNOT, CPHASE, $\sqrt{\text{iSWAP}}$). The variance decays with increasing p, until it reaches a constant level, the value of which is the same for all gates and arrangements. However, CPHASE requires the most layers p to converge, followed by $\sqrt{\text{iSWAP}}$ and CNOT. Fig.4b,d,f shows the redundancy R, which is the fraction of redundant parameters of the PQC. It quickly reaches a constant level with increasing p. $\sqrt{\text{iSWAP}}$ has consistently low R, while for CNOT it varies depending on the arrangement of entangling gates. For CPHASE, we have consistently larger R. This can be easily understood when considering that z rotations commute with the entangling CPHASE layer. When two z rotations appear consecutively on the same qubit, they yield a redundant parameter. R for CNOT depends highly on the entangling gates arrangement.

We note that for these PQCs the number of layers p_c at which the transition of the QFI occurs can be estimated from the value of redundancy R. We find $p_c \approx (1 - R_C)D_C/N$, where R_C is the converged value of R. The eigenvalue spectrum of these PQCs and further types of PQCs are discussed in the supplemental materials C.

In Fig.5, we fix the single-qubit rotations around the y-axis and investigate different entangling gates arranged in a nearest-neighbor one-dimensional chain. Depending on the choice of entangling gates, we find that the variance of the gradient decays to a different constant level with increasing p (see Fig.5a). y $\sqrt{\text{iSWAP}}$ matches the variance found in Fig.3e, whereas y CNOT and y CPHASE have higher variance. In Fig.5b we show the maximal D_C for many layers p. D_C scales exponentially for y CNOT ($D_C \propto 2^N$) and y $\sqrt{\text{iSWAP}}$ ($D_C \propto 2^{N+1}$), whereas for y CPHASE we find numerically an approximate quadratic scaling $D_C \propto N^2$.

In Fig.6 we show how G_C and the variance of the gradient changes when tuning the parameters of a PQC defined as $U(a\theta_{\text{rand}})|0\rangle$, $\theta_{\text{rand}} \in [0, 2\pi)$, $a \in [0, 1]$. When adjusting $a = 0$ to $a = 1$, this corresponds to changing the PQC from parameters all zero to a PQC with random parameters. We exemplary show a PQC consisting of layered randomly chosen single qubit rotations around x,y,z axis and entangling gates arranged in a chain. In Fig.6a, we show different types of entangling gates. G_C increases with a, reaching the parameter dimension D_C for $a = 1$. CNOT and $\sqrt{\text{iSWAP}}$ increase faster with a compared to the PQC with CPHASE gates. The variance of the gradient decreases sharply once a particular a is reached. Note that there is a specific range of pa-

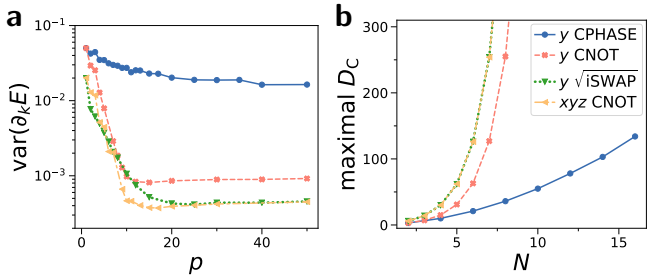

FIG. 5. Capacity of PQCs with y rotations and different entangling gates. The entangling layer is arranged as a nearest-neighbor one-dimensional chain. Three of the PQCs have y rotations, and as reference we show a PQC with randomized x, y or z rotations and CNOT gates. **a)** Variance of the gradient $\mathrm{var}(\partial_k E)$ in respect to the Hamiltonian $H = \sigma_1^z \sigma_2^z$. **b)** Maximal parameter dimension D_C of the PQCs as function of number of qubits N. For y CPHASE we find an approximate powerlaw $D_\mathrm{C} \propto N^2$.

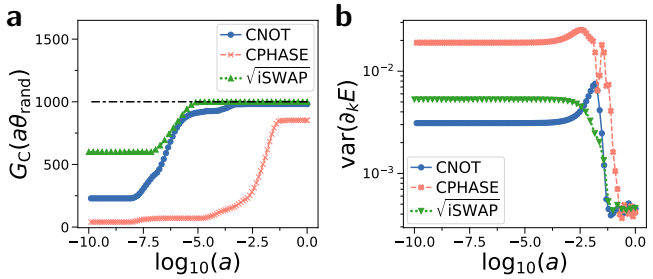

FIG. 6. Tuning the PQC parameters $\theta = a\theta_\mathrm{rand}$, where $a = (0, 1]$ and $\theta_\mathrm{rand} \in [0, 2\pi)$ for circuits composed of random x, y and z rotations and entangling gates arranged in a chain configurations. **a)** The effective quantum dimension G_C as function of $\log_{10}(a)$. Black dashed-dotted line is number of parameters M. **b)** Variance of the gradient $\mathrm{var}(\partial_k E)$ in respect to Hamiltonian $H = \sigma_1^z \sigma_2^z$. All plots show number of layers $p = 100$ and $N = 10$ qubits.

rameters $\log_{10}(a) \approx -2.5$ where the PQCs have nearly maximal G_C and the variance of gradients remains large.

Finally, in Fig.7 we show the scaling of $G_\mathrm{C}(\theta = 0, N)$ with number of qubits N for a PQC with entangling gates in a chain arrangement initialized with $\theta = 0$, corresponding to the point $a = 0$ in Fig.6. Numerically, we find linear scaling of $G_\mathrm{C}(\theta = 0, N)$ for CPHASE entangling gates, quadratic scaling for CNOT gates and higher order polynomial or even exponential scaling for $\sqrt{\mathrm{iSWAP}}$ gates.

V. DISCUSSION

We investigated the capacity and trainability of hardware efficient PQCs using the quantum geometric structure of the parameter space. We introduced the notion of parameter dimension D_C and effective quantum dimension G_C which are global and local measures respectively

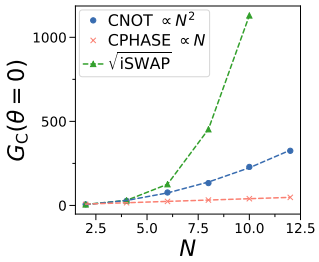

FIG. 7. Effective quantum dimension $G_\mathrm{C}(\theta = 0)$ plotted against number of qubits N for a circuit consisting of randomly chosen parametrized rotations around x, y or z axis with parameters $\theta = 0$, and two-qubit entangling gates arranged in a nearest-neighbor chain. We compare CNOT, CPHASE and $\sqrt{\mathrm{iSWAP}}$ entangling gates. From numerical results, we find G_C scales quadratically for CNOT gates, linearly for CPHASE gates and higher order polynomial or even exponential scaling for $\sqrt{\mathrm{iSWAP}}$. Number of layers p is chosen such that $G_\mathrm{C}(\theta = 0)$ is maximized.

of the space of quantum states that can be accessed by the PQC. Both can be derived from the QFI. We applied these concepts on exemplary PQCs composed of layers of single-qubit rotations and different types of entangling gates arranged in various geometries (see Fig.1). For comparable circuit depth p, we find strong numerical evidence that PQCs constructed from CNOT or $\sqrt{\mathrm{iSWAP}}$ gates have lower variance of the gradient, and thus higher expressibility compared to PQCs with CPHASE gates. For a specific type of PQC composed of y rotations and CPHASE gates, D_C scales only quadratically with number of qubits, which may imply that this PQC can be efficiently simulated on classical computers. We find that the redundancy of parameters varies strongly depending on the configuration of the PQC as well as the type of gates. The redundancy could be systematically reduced by choosing appropriate single-qubit rotations or by using methods of [35] combined with the QFI.

The effective quantum dimension G_C shows the expressive power of a PQC by local variations around a specific parameter set. We find that depending on the entangling gates, G_C can scale widely different with number of qubits, with the largest value found for $\sqrt{\mathrm{iSWAP}}$ gates. While we only studied the case $\theta = 0$, PQCs with correlated parameters could feature similar behavior [18]. Tuning the parameters of a PQC from zero to a random set of parameters yields a crossover from large gradients and small G_C to vanishing gradients and large G_C. For the PQCs investigated, we can find a range of parameters that combines large gradients with a nearly maximal G_C, which could be an optimal starting point for gradient based optimisation. Trade-offs between the expressibility of a circuit and the magnitude of its gradients are a key challenge in finding good initialization strategies [31].

When increasing the number of layers p to a value p_c, a transition occurs in the QFI when D_C reaches its maximal possible value. The transition is characterized by a

disappearance of small eigenvalues of the QFI and a peak in the variance of the logarithm of eigenvalues. This transition may be related to a phase transition in the optimization landscape of control theory. When the number of parameters reaches a threshold, the optimization landscape changes from being spin-glass like with many near-degenerate minima to one with many degenerate global minima [36, 37]. For deep circuits $p > p_c$, the transition leads to a decay of the QNG as small eigenvalues are suppressed. For shallow circuits $p < p_c$, the QNG can be orders of magnitude larger in value compared to the regular gradient, however our numerical results suggest that both regular gradient and QNG decrease exponentially with number of qubits. Thus, the QNG most likely cannot help to solve the barren plateau problem. This contrasts the natural gradient in classical machine learning, which is known to be able to overcome the plateau phenomena that leads to a slow down of optimization [32].

Imaginary-time evolution and variational quantum simulation use a matrix related to the QFI to update the parameters of the PQC [28, 38]. The effective quantum dimension G_C could give major insights on the convergence properties of these algorithms. Recent proposals for adaptively generating ansätze could benefit from the QFI by taking the geometry of the PQC into account when designing PQCs [20].

While cumbersome, we note that the QFI can be determined via measurement of overlaps on the quantum processor [28, 39, 40]. However, in order to evaluate a type of PQC, it is often sufficient to study circuits of a few qubits via classical simulation [41], and extrapolate the results.

During the training of a hardware efficient PQC, the eigenvalue spectrum of the QFI can gain specific features, as has been shown for restricted Boltzmann machines [42]. We show that the PQCs have a characteristic eigenvalue spectra depending on their configuration (see also supplemental materials). The eigenvalues hold important information about the trainability and generalization of a model. For example, a model that generalizes well is known to have a low effective dimension in classical machine learning [26]. It would be interesting to study in what way these statements translate to quantum machine learning. Further, connections to complementary measures of capacity based on classical Fisher information [43] and memory capacity [44] respectively could be explored.

It would be straightforward to extend the concepts of quantum geometry to evaluate the capacity and trainability of noisy PQCs [45], convolutional PQCs [46], optimal control [47], quantum metrology [48] and programmable analog quantum simulators [49].

Python code for the numerical calculations performed in this work are available at [50].

Acknowledgements— This work is supported by a Samsung GRP project and the UK Hub in Quantum Computing and Simulation, part of the UK National Quantum Technologies Programme with funding from UKRI EPSRC grant EP/T001062/1. We are grateful to the National Research Foundation and the Ministry of Education, Singapore for financial support.

[1] J. Preskill, Quantum **2**, 79 (2018).
[2] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, arXiv:2101.08448 (2021).
[3] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. Obrien, Nature communications **5**, 4213 (2014).
[4] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, New Journal of Physics **18**, 023023 (2016).
[5] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, arXiv preprint arXiv:2012.09265 (2020).
[6] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. Sawaya, *et al.*, Chemical reviews **119**, 10856 (2019).
[7] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature communications **9**, 4812 (2018).
[8] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, arXiv:2001.00550 (2020).
[9] C. O. Marrero, M. Kieferová, and N. Wiebe, arXiv:2010.15968 (2020).
[10] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, arXiv:2007.14384 (2020).
[11] L. Bittel and M. Kliesch, arXiv:2101.07267 (2021).
[12] H.-Y. Huang, K. Bharti, and P. Rebentrost, arXiv:1909.07344 (2019).
[13] K. Bharti, arXiv:2009.11001 (2020).
[14] K. Bharti and T. Haug, arXiv:2011.06911 (2020).
[15] K. Bharti and T. Haug, arXiv:2010.05638 (2020).
[16] T. Haug and K. Bharti, arXiv:2011.14737 (2020).
[17] J. W. Z. Lau, K. Bharti, T. Haug, and L. C. Kwek, arXiv:2101.07677 (2021).
[18] T. Volkoff and P. J. Coles, arXiv preprint arXiv:2005.12200 (2020).
[19] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, Quantum **3**, 214 (2019).
[20] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, Nature communications **10**, 1 (2019).
[21] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, arXiv preprint arXiv:2006.14904 (2020).
[22] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Nature **549**, 242 (2017).
[23] K. Nakaji and N. Yamamoto, arXiv:2005.12537 (2020).
[24] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Advanced Quantum Technologies **2**, 1900070 (2019).
[25] D. J. MacKay, in *Advances in neural information processing systems* (1992) pp. 839–846.
[26] W. J. Maddox, G. Benton, and A. G. Wilson, arXiv preprint arXiv:2003.02139 (2020).

[27] N. Yamamoto, arXiv:1909.05074 (2019).
[28] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum **4**, 269 (2020).
[29] D. Wierichs, C. Gogolin, and M. Kastoryano, arXiv preprint arXiv:2004.14666 (2020).
[30] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, Phys. Rev. Res. **2**, 033125 (2020).
[31] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, arXiv preprint arXiv:2101.02138 (2021).
[32] S.-i. Amari, *Information geometry and its applications*, Vol. 194 (Springer, 2016).
[33] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, Applied Physics Reviews **6**, 021318 (2019).
[34] X.-Z. Luo, J.-G. Liu, P. Zhang, and L. Wang, Quantum **4**, 341 (2020).
[35] L. Funcke, T. Hartung, K. Jansen, S. Kühn, and P. Stornati, arXiv preprint arXiv:2011.03532 (2020).
[36] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Phys. Rev. X **8**, 031086 (2018).
[37] H. A. Rabitz, M. M. Hsieh, and C. M. Rosenthal, Science **303**, 1998 (2004).
[38] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, npj Quantum Information **5**, 1 (2019).
[39] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Quantum **3**, 191 (2019).
[40] K. Mitarai and K. Fujii, Physical Review Research **1**, 013006 (2019).
[41] T. Jones, arXiv preprint arXiv:2011.02991 (2020).
[42] C.-Y. Park and M. J. Kastoryano, Physical Review Research **2**, 023232 (2020).
[43] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, arXiv preprint arXiv:2011.00027 (2020).
[44] L. G. Wright and P. L. McMahon, in *CLEO: QELS_Fundamental Science* (Optical Society of America, 2020) pp. JM4G–5.
[45] B. Koczor and S. C. Benjamin, arXiv preprint arXiv:1912.08660 (2019).
[46] I. Cong, S. Choi, and M. D. Lukin, Nature Physics **15**, 1273 (2019).
[47] A. B. Magann, C. Arenz, M. D. Grace, T.-S. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, P R X Quantum **2**, 010101 (2020).
[48] J. J. Meyer, J. Borregaard, and J. Eisert, arXiv preprint arXiv:2006.06303 (2020).
[49] V. Bastidas, T. Haug, C. Gravel, L.-C. Kwek, W. Munro, and K. Nemoto, arXiv:2009.00823 (2020).
[50] T. Haug, "Quantum geometry of parametrized quantum circuits," https://github.com/txhaug/quantum-geometry.
[51] S.-I. Amari, Neural computation **10**, 251 (1998).

Appendix A: Variational quantum eigensolver

The core idea of Variational quantum eigensolver (VQE) is to find the ground state of a Hamiltonian H by minimizing the parameters θ of a PQC in regards to an objective function that represents the energy of a given Hamiltonian $E(\theta) = \langle 0|U^\dagger(\theta)HU(\theta)|0\rangle$ [3]. The minimisation is performed with a classical optimisation algorithm, whereas the energy is measured on a quantum device. According to the Ritz variational principle, the objective function is lower bounded by the ground state energy of H, i.e. $E(\theta) \geq E_g$, where E_g is the true ground state of H.

Appendix B: Quantum Fisher information metric

For VQE, the objective function is updated in hybrid classical-quantum algorithm in an iterative manner. At step n of the procedure, the objective function is evaluated on the quantum computer for a given θ_n. Based on the result, a classical computer selects the next choice θ_{n+1} such that it (hopefully) decreases the objective function. A common scheme to update parameters is ordinary gradient descent

$$\theta_{n+1} = \theta_n - \eta \frac{\partial E(\theta)}{\partial \theta}\,, \tag{B1}$$

where η is a small coefficient and $\partial E(\theta)/\partial\theta$ is the gradient of the objective function.

The above update rule assumes that the parameter space for θ is a flat Euclidian space. However, in general this is not the case, as the underlying PQC and cost function do not have such simple forms. Recent studies have proposed the quantum natural gradient (QNG), inspired from the natural gradient in classical machine learning [51], to minimize the objective function [27, 28]. The main idea is to use information about how fast the quantum state changes when adjusting the parameter θ in a particular direction. Optimisation with the natural gradient updates the parameters according to

$$\theta_{k+1} = \theta_k - \eta_k F(\theta)^{-1} \frac{\partial E(\theta)}{\partial \theta}, \tag{B2}$$

where $\mathcal{F}(\theta)$ is the Fubini-Study metric tensor or quantum Fisher information metric (QFI)

$$\mathcal{F}_{ij} = \mathrm{Re}(\langle \partial_i\psi|\partial_j\psi\rangle - \langle \partial_i\psi|\psi\rangle\langle\psi|\partial_j\psi\rangle)\,, \tag{B3}$$

where $|\partial_i\psi\rangle = \frac{\partial}{\partial\theta_i}|\psi(\theta)\rangle$ denotes the partial derivative of $|\psi(\theta)\rangle$. One can relate $\mathcal{F}(\theta)$ to the distance in the space of pure quantum states, which is the Fubini-Study distance given by

$$\mathrm{Dist}_Q\Big(|\psi(\theta)\rangle,\ |\psi(\theta+\mathrm{d}\theta)\rangle\Big)^2 = \sum_{i,j}\mathcal{F}_{ij}(\theta)\mathrm{d}\theta_i\mathrm{d}\theta_j\,, \tag{B4}$$

where $\mathrm{Dist}_Q(x,y) = |\langle x|y\rangle|^2$.

The QNG has been demonstrated to speed up gradient based optimization techniques for PQCs [27, 28] and avoid local minimas [29]. Efficient classical methods to calculate the quantum Fisher information matrix have been proposed [41].

For the purpose of this study, we investigate PQCs of the form $|\psi(\theta)\rangle = U(\theta)|0\rangle$, where we assume $U(\theta)$ as

a variational circuit in hardware efficient manner, which consists of p layers of parametrized single-qubit rotations followed by fixed entangling gates

$$U(\theta)|0\rangle = \prod_{l=p}^{1} \left[W_l V_l(\theta_l)\right] P|0\rangle \,, \tag{B5}$$

where $V_l(\theta_l)$ consists of parametric single qubit rotations, P is the first layer of single-qubit rotations and W_l are non-parametric entangling gates.

We assume that the quantum processor has a tensored structure of N qubits with Hilbert space $\mathcal{N} = 2^N$. The quantum circuit is parametrized by θ, a vector of $M = pN$ real numbers, where we assume that θ_l are the N parameters that govern the unitary $V_l(\theta_l)$ for the l-th layer of the circuit [28]. We define the following short form for representing subcircuits between layers $l_1 \leq l_2$

$$U_{[l_1:l_2]} := W_{l_2} V_{l_2} \cdots W_{l_1} V_{l_1} \,. \tag{B6}$$

We now show how to calculate the derivative of a parameter $\theta_{l,k}$, which is the k-th parameter for a single-qubit rotations within the l-th layer. With $V_l = \prod_{n=1}^{N} \exp(-i\theta_{l,n}/2\sigma_n^\alpha)$, where σ_n^α is a Pauli operator $\alpha \in \{x, y, z\}$ acting on qubit n. We define $\partial_{l,n} = \frac{\partial}{\partial_{\theta_{l,n}}}$ as the partial derivative for the parameter $\theta_{l,n}$ which controls the single qubit rotation on the n-th qubit in the l-th layer

$$\partial_{l,n} V_l(\theta_l) = -i\frac{\sigma_n^\alpha}{2} V_l(\theta_l) \,, \tag{B7}$$

where α is a function of n and l. For the full unitary we find

$$\begin{aligned} \partial_{l,n} U(\theta)|0\rangle &= U_{(l:L]} W_l \partial_{l,n} V_l(\theta_l) U_{[1:l)} P|0\rangle \;\; , \\ &= U_{(l:L]} W_l V_l(\theta_l)(-i\frac{\sigma_n^\alpha}{2}) U_{[1:l)} P|0\rangle \;\; , \\ &= U_{[l:L]}(-i\frac{\sigma_n^\alpha}{2}) U_{[1:l)} P|0\rangle \;\; . \end{aligned} \tag{B8}$$

With Eq. (B8) inserted into Eq. (B3), we can calculate the QFI.

Appendix C: Further data on the PQCs

In Fig.8, we show further types of PQCs as defined in the caption. We highlight that the PQC rand(xyw) CPHASE has lower redundancy compared to rand(xyz) CPHASE. The reason is that the z rotations, which can commute with the CPHASE layer, are replaced with non-commuting $(x+y)/\sqrt{2}$ rotations. This leads to a faster decrease in the variance of the gradient as well. We also define a common type of PQC zxz CNOT, which has first been introduced in [22]. We note that while it has three rotations per qubit and layer, compared to rand(xyz) CNOT the decay of the variance of the gradient as function of p remains the same in both types of PQC. Finally, we show further examples of the transition in the QFI, visible both in the peak of the variance of the logarithm of the eigenvalues, and in the decay of the QNG.

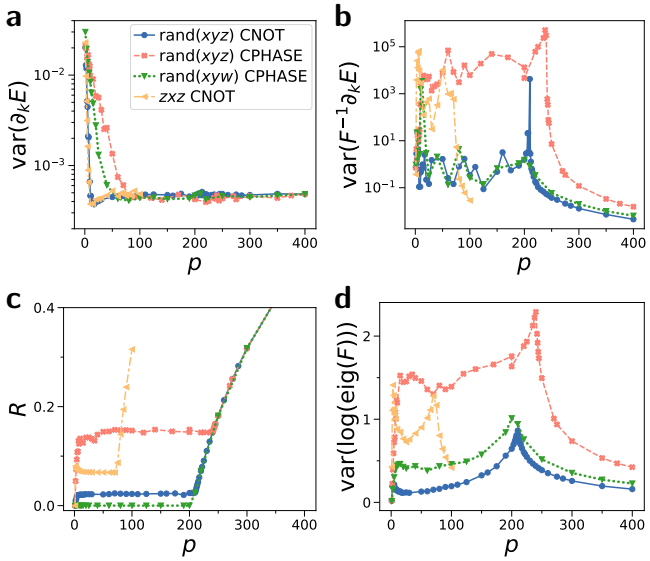

FIG. 8. Properties of further PQCs plotted against layers p for $N = 10$ qubits. PQCs have nearest-neighbor chain entangling layers. We define the type of PQCs in the legend: rand(xyz) denotes randomized single-qubit rotations around $\{x, y, z\}$ axis. rand(xyw) denotes randomized single-qubit rotations around $\{x, y, (x+y)/\sqrt{2})\}$ axis. zxz denotes that for every layer there are three single-qubit rotations, around z, x and z axis. **a)** Variance of the gradient $\mathrm{var}(\partial_k E)$ in respect to the Hamiltonian $H = \sigma_1^z \sigma_2^z$. **b)** Variance of QNG $\mathrm{var}(\mathcal{F}^{-1}\partial_k E)$. **c)** Redundancy R of parameters of the PQCs. **d)** Variance of the logarithm of the eigenvalues of the QFI.

Appendix D: Histograms of eigenvalues

In Fig.9 we show the distribution of eigenvalues for the PQCs of Fig.4 in the main text. We find that a characteristic spectrum for the different PQC types. Note that CPHASE appears to have more pronounced tails in all cases.

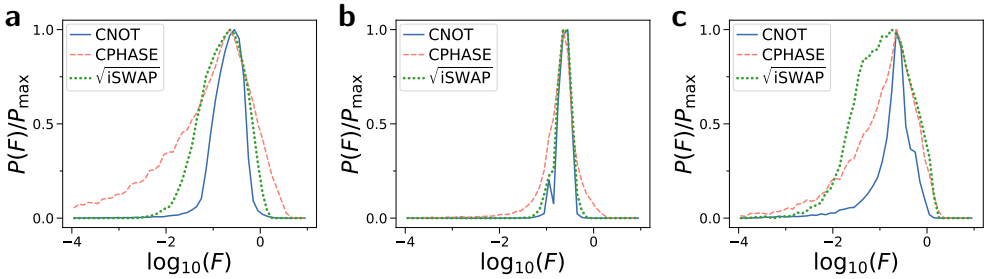

FIG. 9. Distribution of eigenvalues of QFI for PQCs shown in Fig.4 in main text. **a)** nearest-neighbor chain arrangement of entangling gates **b)** all-to-all connectivity **c)** alternating nearest-neighbor. All graphs for $N = 10$ qubits and number of layers $p = 50$.