The effect of the processing and measurement operators on the expressive power of quantum models

Aikaterini (Katerina) Gratsea · Patrick Huembeli

Received: 11 November 2022 / Accepted: 14 March 2023 © The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract
There is an increasing interest in Quantum Machine Learning (QML) models, how they work and for which applications they could be useful. There have been many different proposals on how classical data can be encoded and what circuit ansätze and measurement operators should be used to process the encoded data and measure the output state of an ansatz. The choice of the aforementioned operators plays a determinant role in the expressive power of the QML model. In this work we investigate how certain changes in the circuit structure change this expressivity. We introduce both numerical and analytical tools to explore the effect that these operators have in the overall performance of the QML model. These tools are based on previous work on the teacher-student scheme, the partial Fourier series and the averaged operator size. We focus our analysis on simple QML models with two and three qubits and observe that increasing the number of parameterized and entangling gates leads to a more expressive model for certain circuit structures. Also, on which qubit the measurement is performed affects the type of functions that QML models could learn. This work sketches the determinant role that the processing and measurement operators have on the expressive power of simple quantum circuits.

Keywords Quantum neural networks · Machine learning · Variational quantum circuits

1 Introduction
In recent years, there is increasing interest in the capabilities of Quantum Machine Learning (QML) models and their potential applications Gili et al. 2022; Pérez-Salinas et al. 2020; Peters and Schuld 2022; Huang et al. 2020; Heimann et al. 2022. To better understand for what tasks a QML model could be potentially used for Dawid et al. (2022), we first need to understand where its strengths and weaknesses lie.

Instead of focusing on the data encoding, other works study the importance of the trainable part of the quantum circuits (consisting of data independent but parameterized gates) to which we refer to as the processing part of the circuit. For example, in Ostaszewski et al. (2021), the authors perform a structure optimization that aims to find a more hardware efficient architecture for the desired problem.

Finally, recent works focus on benchmarking different circuit structures both numerically (Du et al. 2020; Sim et al. 2019; Moll et al. 2018; Gratsea and Huembeli 2021 and theoretically Wu et al. (2021); Abbas et al. (2021); Bishop et al. (2017); Gratsea et al. (2021); Lewenstein et al. (2021)). How to choose an ansatz for a supervised QML task is still an open question. Circuit ansätze that are used to find ground states in quantum chemistry (Cao et al. 2019) or many-body physics are often inspired by the problem itself (Kandala et al. 2017; Meyer et al. 2022; Cai 2020) and some understanding of the optimization problem at hand can help to find a well-suited ansatz (Ostaszewski et al. 2021; Grimsley et al. 2019). For QML applications this does not apply in general (Peruzzo et al. 2014; Mitarai et al. 2018; Schuld et al. 2020a; Tang et al. 2021) which might partially be because we still do not fully understand how specific parts of the processing and measurement operators affect the performance of the QML model. Moreover, even for small quantum circuits with a few number of qubits, there are still some open questions: “How many parameterized gates should be used?”, “How many entangling gates and where?”, “Where should the circuit be measured?” and “Do ancilla qubits help improve the expressivity of the circuit?”.

1 ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain
2 San Francisco, USA
Inspired by these questions, we aim to study the effect of the processing part of QML architectures. We believe that the structure of the processing and measurement operators play a determinant role in the QML model performance. Here the term performance characterizes the model’s ability to learn through training (Benedetti et al. 2019), the type of functions that it can express (Schuld et al. 2020b) and its scrambling capability (Wu et al. 2021; Shen et al. 2020), i.e. extracting information from the input state to the readout qubit.

We use both analytical and numerical tools to explore how the processing unitaries and measurement affect the properties of a QML model. We focus our analysis on simple quantum models with just 2–3 qubits, which help us better understand how different parts of the processing and measurement operators affect the performance. This analysis allows us to understand how the processing architecture affects the expressivity of the whole model, at least for circuits with small amounts of qubits, with the hope to draw insights on how specific parts will affect the performance of more complicated architectures with more qubits. We observe that increasing the number of parameterized and entangling gates leads to a more expressive model for certain circuit structures and that ancillary qubits help increasing the expressive power of the model. Finally, applying the measurement on certain qubits changes the classes of functions that QML models can learn.

The paper is structured as follows. In Section 2, we introduce the analytical and numerical tools discussed in this work. In Section 3, we apply these tools to analyze the performance of two simple quantum models. In Section 3.3, we discuss that regarding the processing and measurement operator as a single unit could give insights on the performance of the QML models. Finally, in Section 5, we summarize the importance of the non-encoding unitaries and the effect they have on the expressive power of quantum models.

2 Tools for analysing QML models

We study quantum circuits that consist of three parts: the encoding unitaries $S(x)$, the processing gates $U_{th}$ (consisting of data independent but parameterized gates) and the measurement operator $M$. The building blocks of such arbitrary quantum circuits are shown in Fig. 1.

Throughout this work, we focus on supervised learning tasks with a given data set $\mathcal{D} = \{(x^k, y^k)\}$, where the input data $x^k$ are encoded by $S(x^k)$ and $y^k$ are the labels given by the measurement outcome. The initial state is $|0\rangle \otimes^N$, where $N$ is the number of qubits. As shown in Fig. 1, we first apply the encoding unitary $S(x)$ (yellow), then the processing $U_{th}$ (blue) and finally the measurement operator $M$ (green).

In this section, we focus on two theoretical tools for characterizing the performance of QML models, i.e. the averaged operator size (Wu et al. 2021) and the Fourier representation (Schuld et al. 2020b). These methods give us quantitative results on the effect of the processing and measurement operators (blue parameterized unitary and green operator in Fig. 1). We also employ an extended version of the teacher-student scheme discussed in Gratsea and Huembeli (2021) to quantify differences in the prediction maps numerically, i.e. the density plots of the model predictions or labels $y^k$ for the input data $x^k$.

2.1 Averaged operator size

In the works Wu et al. (2021); Nahum et al. (2018); Roberts et al. (2018), the authors introduce related quantities to characterize quantum information scrambling, i.e. extracting information from the input state $S(x)|0\rangle$ to the readout qubit. In the work of Wu et al. (2021), the authors propose the averaged operator size - a quantity that depends on the circuit architecture, the processing and the measurement operators. Moreover, its value is positively correlated with the learning efficiency of the QNN architecture.

To calculate the averaged operator size of an arbitrary operator $\hat{O}$, we decompose it into a summation of a Pauli strings, which can be done for any Hermitian operator (Nielsen and Chuang 2011). Such a decomposition has the form

$$\hat{O} = \sum_{\alpha} c_{\alpha} \hat{\sigma}_{\alpha_1}^1 \otimes \hat{\sigma}_{\alpha_2}^2 \cdots \otimes \hat{\sigma}_{\alpha_n}^n,$$

where $\hat{\sigma}_{\alpha_i}^i$ for $\alpha_i \in \{0, 1, 2, 3\}$ define the Pauli operators including the identity acting on qubit i. The coefficients $c_{\alpha}$ can be computed by

$$c_{\alpha} = \frac{1}{2\pi} \text{Tr}\left( \hat{O} \hat{\sigma}_{\alpha_1}^1 \otimes \hat{\sigma}_{\alpha_2}^2 \cdots \otimes \hat{\sigma}_{\alpha_n}^n \right),$$
The operator size of a Hermitian operator $\hat{O}$ is given by

$$\text{Size}(\hat{O}) = \sum_{\alpha} |c_{\alpha}|^2 l(\alpha),$$  \hspace{1cm} (3)

where $c_{\alpha}$ are the coefficients of decomposition in Eq. (1) and $l(\alpha)$ counts the number of non-identity matrices in each Kronecker product in the summation given by the same equation. In our case, the hermitian operator $\hat{O} = U^\dagger_\theta M U_\theta$ depends only on the processing unitary $U_\theta$ and the measurement $M$ operators.

To obtain the averaged operator size for a given circuit architecture $U_\theta$ and a fixed measurement $M$, we take the average over Haar random unitaries $U_\theta$

$$\text{Size} = \int dU_\theta \text{Size} \left( U_\theta^\dagger M U_\theta \right).$$  \hspace{1cm} (4)

A larger value of the average operator size suggests a more expressive circuit structure (Wu et al. 2021).

### 2.2 Map differences from teacher-student scheme

The teacher student scheme employed in Gratsea and Huembeli (2021) can be used to compare the expressivity of different QML circuit architectures. The main idea is that alternately one circuit plays the role of the teacher and the other one of the student. A randomly initialized teacher circuit maps the input data $x^k$ to labels $y^k$ that are to be learned by the student, and then, the roles are reversed. To compare their performances in the original work (Gratsea and Huembeli 2021), three different quantitative scores were considered: the accuracy score, the (average) loss and the relative entropy and the prediction maps for a more qualitative overview. Here, we add another score to the TS scheme that directly compares differences between prediction maps which allows us to better quantify them. Specifically, we compute the average of the differences between the prediction maps of the teacher and student

$$\Delta \tilde{y} = \frac{1}{p} \sum_{k} |y^k_T - y^k_S|,$$  \hspace{1cm} (5)

where $\{y^k_T\}$ are the generated labels from the teacher and $\{y^k_S\}$ the learned labels of the student where $k$ runs over all points $p$ of a given input dataset $\{x^k\}$. Also, $y^k_T$, $y^k_S$ take continuous values in $[-1, 1]$, but for the calculation of the $\Delta \tilde{y}$ we re-scale them to be in $[0, 1]$. We can present the $\Delta \tilde{y}$ as a percentage difference between the two studied models. This allows us directly to compare how similar the prediction maps of the two models are. It is a quantitative measure of how well the student can learn the teacher, i.e. a $\Delta \tilde{y}$ equal to zero suggests that the student learns the teacher perfectly, while high values for $\Delta \tilde{y}$ indicate that the student is not able to learn the labeling provided by the teacher.

### 2.3 The representation of quantum models with partial Fourier series

In the work of Schuld et al. (2020b), the authors explore how the data encoding influences the class of functions that a quantum model can learn. Circuit architectures employed in supervised tasks with multiple encoding unitaries on different qubit can be mapped to a partial multivariate Fourier series

$$f_\theta(x) = \sum_{j} \sum_{k} c_{jk} e^{i x \cdot (\lambda_k - \lambda_j)},$$  \hspace{1cm} (6)

where $j, k \in [2d]^N$ with $N$ the number of qubits and $d$ is the dimension of an encoding gate, i.e. $d = 1$ if it is a single qubit gate. As it is explained in the original work, the frequency spectrum determines the functions that the quantum model could express, while the coefficients $c_{jk}$ determine how the accessible functions can be combined.

The number of Fourier basis functions is solely determined by the eigenvalues of the data-encoding Hamiltonians. This means that repeated data encoding gives rise to a larger frequency spectrum $\Omega$ and more complicated function classes. The processing and the measurement of the circuit determine the coefficients, and therefore, how the accessible functions can be combined. Inspired by the work of Schuld et al. (2020b), in Section 3.3, we study how the processing and measurement operators of simple quantum circuits affect the coefficients, and therefore, the final function classes that these quantum circuits “has access to”. This analysis sheds light on how individual elements of the processing and measurement architecture affect the function classes that the quantum model can express.

### 3 Simple quantum models

In this section, we apply the tools described in the Section 2 on two simple quantum models shown in Fig. 2. We use a two dimensional input data set with $x^k = (x^k_1, x^k_2)$ and encode it to the circuit with the gates $S(x) = R_z(x_1) \otimes R_z(x_2)$ applied on two distinct qubits, where $R_z(\phi) = \exp(-i\phi\sigma_z/2)$ is the single qubit $X$ rotation gate. These data encoding gates are depicted in light and dark orange colors respectively in the circuit diagrams. The blue gates represent the parameterized single qubit rotations $Rot(\phi, \theta, \omega) = R_z(\omega) R_y(\theta) R_z(\phi)$. The predictions take continuous values $y^k \in [-1, 1]$ given by the outcome of the measurement $\langle \psi^k | Z | \psi^k \rangle$, where $| \psi^k \rangle = U_\theta S(x) | 0 \rangle^\otimes N$.  

© Springer
Fig. 2 A 2-qubit and 3-qubit simple circuit architecture. The light and dark orange colors represent the data encoding gates, while the blue the parameterized gates. A CNOT gate is applied before the measurement operator for the 2-qubit circuit, while a Toffoli gate for the 3-qubit circuit.

3.1 Averaged operator size of simple quantum models

We compute the averaged operator size of Eq. (4) for the 2-qubit and 3-qubit circuits of Fig. 2 by taking the average over the Haar random unitaries for the parameterized single qubit unitaries depicted with blue color. For the numerical simulation we use a Monte-Carlo integration (McClean 2023) introduced in the Appendix A. In Fig. 3, we plot the averaged operator size as a function of the number of parameterized gates. The error bars are the standard deviation from the mean value. We increase the number of parameterized gates by adding a layer $L$ with two arbitrary single qubit gates (blue color) and an a CNOT entangling gate, $L = CNOT Rot(\phi, \theta, \omega) \otimes Rot(\phi', \theta', \omega')$. We stress here that the 3-qubit circuit always has a Toffoli gate before the measurement operator.

The average operator size for the 2-qubit case has approximately the same value around 1.60 independent of the number of gates used (see blue solid line of Fig. 3). The drop observed from two to four parameterized gates is small while the error is large. Therefore, this suggests that for the 2-qubit case the performance is approximately the same as we increase the number of parameterized gates. Moreover, we see that on average the random unitaries are not as expressive as the ones that correspond to the maximum value, which is why it does not converge to the maximum value and also results in large errors. On the contrary, for the 3-qubit case, the averaged operator size significantly increases once we move from two to four parameterized gates, and afterwards, reaches a plateau. These results already suggest that the 3-qubit case performs better than the 2-qubit case, i.e. the circuit structure of the processing and measurement operator is more expressive.
Next, we focus on the maximum value of the operator size, which corresponds to specific values of the parameterized angles that give the best possible performance of this circuit structure (represented with the dashed line in Fig. 3). From the dashed lines in Fig. 3, we see that for both circuits the maximum value of the operator size is equal to 2 for the circuits with two parameterized gates. As we increase the number of parameterized (and entangling gates), the maximum value of the operator size is increased for the 3-qubit, but quickly reaches a plateau. Once again, these results suggest that the 3-qubit circuit structure is more expressive than the 2-qubit.

### 3.2 Teacher-student scheme

To further analyze the expressive power of the two models in Fig. 2, we employ the Teacher-student scheme introduced in Section 2.2. The analysis of the differences $\Delta y$ of the prediction maps validates the results from the average operator size in the previous section.

We compute the $\Delta y$ defined in Eq. 5 and see that on average both students are approximately 30% off from the desired target distributions of their teachers. An example of the prediction maps are shown in Fig. 10 in the Appendix B, which suggests that the student learns pretty much uncorrelated labeling. These results are in accordance with the results from the averaged operator size of Fig. 3 for two parameterized gates in the sense that both of them have similar values for their averaged operator size, i.e. around 1.7 and 2 for the 2-qubit and 3-qubit students, respectively.

Next, we increase the number of parameterized (and entangling gates) by adding a layer $L$ and present the results in Fig. 4. We see a significant improvement in performance by increasing the parameterized gates from 2 to 4 for the 3-qubit student. This means that it can learn more reliably the labeling provided by the 2-qubit teacher. On the contrary, the 2-qubit student has almost a constant $\Delta y$ for any amount of parameterized gates which suggests that it can’t improve and learn more reliably the 3-qubit teacher’s outputs.

To obtain these results we used a dataset of 500 points on a 2D grid with $x_l \in [-\pi, \pi]$ on a grid (which essentially affects only the Fourier basis functions that can be expressed) and we generated 100 different labelings via different random initializations of the teacher. The students are as well randomly initialized and trained until convergence. These results are in agreement with the calculations from the averaged operator size of the previous Subsection 3.1. In both cases, for the 3-qubit circuit, we observe an improvement in performance as we increase the number of parameterized gates. Finally, we quickly reach a plateau in performance as we increase further the number of gates as it is also observed in the analysis of the averaged operator size.

### 3.3 The type of functions that simple quantum models learn

The results from the previous two subsections [3.1 3.2] suggest that simple changes in the structure of the circuit highly affects the average operator size and $\Delta y$, i.e., the learning capability and expressive power of the models. Therefore, the number of parameterized (and entangling) gates, the circuit architecture and the number of (ancilla) qubits play an important role in QML model’s performance. To further analyze the effect of the processing and measurement operators on the expressive power of the models, we use the representation of quantum models as partial Fourier series introduced in Subsection 2.3.

The Fourier coefficients defined in Eq. 6 are determined by the eigenvalues of the data-encoding Hamiltonians. For example, the data-encoding Hamiltonian for each qubit from the circuits of Fig. 2 is $H = \frac{1}{2} \sigma_x$ and we can assume without loss of generality that it has two distinct eigenvalues, i.e. $\lambda_1 = -1, \lambda_2 = 1$. Importantly, since we have the same type of encoding gate for each qubit, they have the same frequency spectrum $\Omega$. As it is explained in detail in the work of Schuld et al. (2020b), we can derive the frequency spectrum for each qubit $\Omega = \{-1, 0, 1\}$ from the possible differences $n_{\text{qubit}} = \lambda_j^{\text{qubit}} - \lambda_j^{\text{qubit}}$ for $\lambda_j^{\text{qubit}}, \lambda_j^{\text{qubit}} \in \{-1, 1\}$. Then, we have the Fourier coefficients as $c_{n_1n_2}$ and Eq. (6) can be written as

$$f(x) = \sum_{n_1 \in \Omega} \sum_{n_2 \in \Omega} c_{n_1n_2} e^{-in_1x_1} e^{-in_2x_2}. \quad (7)$$

Both circuit models in Fig. 2 have the same number of Fourier basis functions given by Eq. 7, since they have the
same number of encoding gates. This fundamentally limits their learning ability. But importantly, the structure of the processing and measurement operator affect the distribution of the coefficients.

In Fig. 5, we plot the real and imaginary parts of the Fourier coefficients for the 2-qubit (blue color) and 3-qubit (orange color) explicit circuits of Fig. 2 with two parameterized gates each for a random point from the dataset used in Section 3.2.

The coefficients \(c_{00}, c_{01}, c_{10}, c_{0,-1}\) from Fig. 2 for the 2-qubit circuit are all zero (blue color), while they are non-zero for the 3-qubit (orange color). But what happens with the coefficients \(c_{11}, c_{-1,-1}, c_{-1,-1}, c_{1,-1}\)? Interestingly, computing the exact analytical formula for the distributions of the coefficient \(c_{1,1}\) shows that they are different. We present these analytical formulas in Appendix C. In summary, even though the two quantum models of Fig. 2 have access to the same Fourier basis functions, the different distribution of their coefficients results in a different combination of these basis functions, and hence, give rise to different function classes.

This is also in accordance with the results from the previous two subsections (3.1 3.2). The 2-qubit student (Fig. 2 left) has difficulties learning the prediction map of the 3-qubit teacher (Fig. 2 right), since it has four coefficients strictly equal to zero which does not help approximating the more complicated distribution of the 3-qubit teacher. But reversing the roles, the 3-qubit student has also difficulties learning the 2-qubit teacher. Even though, the student could learn the zero coefficients, the coefficients \(c_{11}, c_{-1,-1}, c_{-1,-1}, c_{1,-1}\) of the teacher belong to a different function class, i.e. they draw their values from a different distribution. This could be seen from the larger spread of the coefficients in Fig. 5. These results are in accordance with the averaged operator size (Fig. 3) and \(\overline{\Delta y}\) (Fig. 4), since they also suggest that for just two parameterized gates both the 2-qubit and 3-qubit circuit have similar performance.

![Fig. 5](image-url) The real and imaginary part of the Fourier coefficients for the 2-qubit (blue color) and 3-qubit (orange color) explicit circuits of Fig. 2 with two parameterized gates each
Most importantly, once we add an extra layer $L$ to the 3-qubit circuit, the spread of the Fourier coefficients in Fig. 5 (in orange) increases and overlaps with the coefficients of the 2-qubit circuit (in blue) almost completely. These results are shown in light blue color at Fig. 12 in the Appendix D). This is in accordance with the averaged operator size and $\Delta y$ from the teacher-student scheme where a better performance is observed for the 3-qubit case by adding an extra layer $L$. On the contrary, adding an extra layer in the 2-qubit case does not improve the performance of the student as suggested as well from Figs. 3 and 4, i.e. the distribution of the coefficients shown in Fig. 5 in blue stays the same.

But what more can we say for the distribution of these coefficients? Could we understand which circuit elements determine the exact zero terms? We tackle these questions in the next section.

### 4 Variational measurement

We can describe the expectation value of any circuit of Fig. 1 with the representation of matrix blocks shown in Fig. 6a.

Following Schuld et al. (2020b), the data encoding unitary can be written as $S(x) = V^† e^{-ix\Sigma} V$ via an eigenvalue decomposition, where $\Sigma$ is a diagonal matrix of eigenvalues and $V$ is the unitary formed with the eigenvectors as columns. We absorb $V$, $V^†$ into the initial state $|\psi\rangle = V |\psi\rangle$ and into the processing part of the circuit $U' = V U V^†$. For consistency, we transform the measurement operator with $M' = VMV^†$. As a result, the encoding unitary is simply a diagonal matrix $\exp(-i\Sigma \delta)$. The transformed matrix blocks are shown in Fig. 6b.

In Fig. 7, we schematically illustrate the mapping between the Fourier coefficients and the frequencies instead of explicitly writing the full matrix given by the inner product of the orange, blue and green matrix blocks of Fig. 6b. The Fourier coefficients depend only on the processing and measurement operators, therefore the matrix elements of $\overline{M} = U^{\dagger} M' U'$, while the frequencies on the encoding operator, i.e. the diagonal matrix of the encoding Hamiltonian as explained in detail in Schuld et al. (2020b). As introduced in Eq. 7, the Fourier coefficients are defined as $c_{n1,n2}$. For example, $c_{00}$ corresponds to the frequencies with $n_1 = n_2 = 0$ for both qubits, which after explicitly calculating the expectation value in Eq. 7 in turn correspond to the diagonal matrix elements of $\overline{M}$. Following the same procedure, $c_{01}$ is the sum of elements $\overline{M}_{12}$ and $\overline{M}_{14}$, $c_{10}$ is the sum of elements $\overline{M}_{13}$ and $\overline{M}_{24}$, $c_{11}$ is the element $\overline{M}_{14}$ and $c_{1,-1}$ is the element $\overline{M}_{23}$. Finally, $c_{0,-1}$, $c_{-1,0}$, $c_{-1,-1}$ and $c_{-1,1}$ are the complex conjugates of $c_{01}$, $c_{10}$, $c_{11}$ and $c_{1,-1}$, respectively.

We now discuss the simple example of the 2-qubit case from Fig. 2 with the Pauli Z operator applied to the 2nd qubit ($M'_1 = I$, $M'_2 = Z$). The transformed measurement operator $M' = M'_1 \otimes M'_2 = I \otimes (VZV^†)$ of the data-encoding Hamiltonian becomes

$$ M' = \begin{bmatrix} 0 & h & 0 & 0 \\ k & 0 & 0 & 0 \\ 0 & 0 & 0 & h^* \\ 0 & 0 & k^* & 0 \end{bmatrix}, \quad (8) $$

where $h, k$ are complex trigonometric functions that determine the distribution of the coefficients. But the mapping between the zero elements and the coefficients that are zero does not match. For example, $c_{11}$ which is determined by
has a non zero distribution as shown in Fig. 5 (blue color), while eq. (8) suggests that \( \overline{M}_{14} \) it should be strictly zero.

To resolve this discrepancy, we need to consider the processing and measurement operators as one matrix block \( \overline{M} = U^\dagger M' U' \) (shown in green in Fig. 6c). This is in accordance with recent works (Schuld 2021a; Jerbi et al. 2021), where the processing and measurement operators are regarded as one part in the circuit structure.

For the 2-qubit case, the combined matrix \( M \) (green blocks in Fig. 6c) becomes:

\[
M = \begin{bmatrix}
  0 & 0 & f \\
  0 & g & 0 \\
  f^* & 0 & 0
\end{bmatrix},
\]

where \( f \), \( g \) are complex trigonometric functions that determine the distribution of the coefficients shown in Fig. 5. Following the mapping between the coefficients and the elements of the combined matrix \( M \), we can immediately see which coefficients are zero. The only non-zero coefficients are the \( c_{11}, c_{1-1} \) and their complex conjugates determined by \( \overline{M}_{14}, \overline{M}_{23} \) and \( \overline{M}_{41}, \overline{M}_{32} \), respectively.

This analysis suggests that in some cases replacing the usual segregation of a quantum circuit into three parts: encoding, processing and measurement with just two parts: encoding and variational measurement could reveal further insights on the performance and properties of QML models.

### 4.1 Toy application

To further emphasize the importance of the measurement operators of a circuit, we present a simple quantum model with two qubits (Fig. 8). The Fourier coefficients of this circuit are represented in Fig. 9, where the coefficients in purple represent the measurement of the 1st qubit and the Fourier coefficients in green represent the measurement on the 2nd qubit. As we see, the distribution of the Fourier coefficients differ greatly. Therefore, with the same circuit structure, we could create two different functions classes depending on where we measure. This might have a more relevant applications, but it already suggests that once we better understand how simple elements on our circuit structure affect the overall performance it is more natural to think of applications.

### 5 Conclusions

The data encoding is fundamentally limiting the performance of a circuit, but the processing and measurement can also significantly affect the capabilities of the model. We show that different parts of processing and measurement opera-
tors affect the performance of the QML model. As expected, increasing the number of layers \( L \) has a direct effect on the model’s performance, but a plateau in the performance seems to be reached. We also observed that the circuit with the ancilla has overall a better performance in accordance with the recent work of Du et al. (2020).

To quantify the model’s performance, we introduced three different tools: the averaged operator size, the \( \overline{A}_y \) and the coefficient from the partial Fourier series. By applying the Fourier representation, we see that two different circuits with the same number of parameters could give rise to different function classes. Finally, we find that the segmentation of the processing and measurement operator might not be ideal when interpreting the parts of the circuit. Instead when considered as one, i.e. the variational measurement, could give insights on the model’s performance. The idea of the variational measurement has also been presented in the context of quantum kernels (Schuld 2021b). This might bring the relation between the Fourier representation and quantum kernels one step closer.

Once we better understand how QML models work, we can naturally start thinking about applications. For example, we presented a trivial implementation that arises from where we measure. Specifically, measuring either one of the two qubits on a 2-qubit circuit gives rise to fundamentally different classes of functions that this model could express. Therefore, with the same circuit structure one could learn or express two different function classes. This can be seen from the numerical calculation of the real and imaginary part of the Fourier coefficients in Fig. 8.

We understand that computing the analytical trigonometric expression for more qubits, or even gates, is similar to opening Pandora’s box. Instead we suggest to focus on the other two tools proposed, i.e. the averaged operator size and the map differences from the teacher-student scheme. In this work, we focused on simple quantum models and verified that the tools presented here are robust. The analytical and numerical results are in agreement, and therefore, these tools could be further used to test the performance of more complex circuit structures.

In conclusion, applying the averaged operator size and teacher-student scheme to other simple circuit structures could further help understand how QML models work. It

---

**Fig. 8** A 2-qubit circuit with 4 parameterized and two entangling gates where we could either measure on the 1st or 2nd qubit.
would also be interesting to exploit the effect of measurement for a more interesting use-case application. But to do so, a careful study on how the measurement position affects the model’s performance should be undertaken, i.e. by increasing the number of qubits and circuit complexity (Haferkamp et al. 2022). The recent work (Casas and Cervera-Lierta 2023) is a step towards this direction. This study could also further strengthen the belief that ancillary qubits have better performance overall (Du et al. 2020).

Now, we have a better understanding on the basic questions posed earlier in the introduction: “How many parameterized gates to use?”, “How many entangling gates and where?”, “Do ancilla qubits help?” and “Where to measure?”. We understand that the processing and measurement operators strongly affect the performance of the QML model. Therefore, focusing the research directions on how exactly they affect it is of great importance. If we want to find the great applications that QML promises (Alcazar and Perdomo-Ortiz 2021; Schuld and Killoran 2022), we first need to understand how QML models work.

**Appendix A: Monte-Carlo integration**

For the numerical simulation we use Monte-Carlo integration (McClean 2023) and approximate the integral as

\[
\int_{U^{(N)}} f(U) dU \approx \frac{1}{p} \sum_{i=1}^{p} f(U_i),
\]

\[ \text{(A1)} \]
where \( p \) is the total number of random unitaries used and \( U_i \) is a randomly drawn unitary according to the Haar measure. To define such unitary numerically, we start by calling a \( N \times N \) matrix with Gaussian values. Then, we perform a QR decomposition on this matrix which gives two matrices \( Q \) and \( R \). Next, we define the diagonal matrix \( D \) from the diagonal elements of the matrix \( R \), i.e. \( D_{ii} = R_{ii} / |R_{ii}| \). Finally, the unitary random matrix according to the Haar measure is defined as \( U_i = QD \) (McClean 2023).

**Appendix B: Prediction maps**

Fig. 10 shows an example of the prediction maps for the 2-qubit circuit as a teacher and the 3-qubit circuit as a student along with the reverse roles for each circuit. In both cases, the student deviates from the data distribution of its teacher.

**Appendix C: Coefficient formula**

Here, we present the explicit formula of the real and imaginary parts of the coefficient \( c_{11} \) for the 2-qubit and 3-qubit circuits in Fig. 2 in Eq. (C1),(C2) and Eq. (C3),(C4), respectively for the real and imaginary parts.

\[
\text{Re} (c_{11}) = \sin \left( \frac{\phi_1}{2} \right) \sin \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) \cos \left( \frac{\phi_2}{2} \right) \sin \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) - \sin \left( \frac{\phi_1}{2} \right) \sin \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) \cos \left( \frac{\phi_2}{2} \right) \cos \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) + \frac{1}{4} \left( -4 \cos \left( \frac{\phi_2}{2} \right)^2 - 2 \right) \cos \left( \frac{\phi_2}{2} \right)^2 + \frac{1}{2} \cos \left( \frac{\theta_1}{2} \right)^2 - \frac{1}{2}
\]

(C1)

\[
\text{Im} (c_{11}) = -\cos \left( \frac{\phi_1}{2} \right)^2 - \frac{1}{2} \sin \left( \frac{\phi_1}{2} \right) \cos \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) - \sin \left( \frac{\phi_1}{2} \right) \sin \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) \cos \left( \frac{\phi_2}{2} \right) \cos \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) + \frac{1}{2} \cos \left( \frac{\phi_2}{2} \right)^2.
\]

(C2)

\[
\text{Re} (c_{11}) = -\frac{1}{8} - \frac{1}{8} \left( -2 + 4 \cos \left( \frac{\phi_2}{2} \right)^2 \right) \cos \left( \frac{\theta_1}{2} \right)^2 + \frac{1}{2} \sin (\phi_2) \sin (\theta_1) \sin \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) + \frac{1}{2} \cos \left( \frac{\theta_1}{2} \right)^2.
\]

(C3)

\[
\text{Im} (c_{11}) = \frac{1}{2} \sin (\phi_2) \sin \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} + \frac{\theta_2}{2} \right) \left( \cos \left( \frac{\theta_1}{2} \right)^2 - \frac{1}{2} \right) \sin \left( \frac{\theta_1}{2} \right) \cos \left( \frac{\theta_1}{2} \right)
\]

(C4)

**Appendix D: 3-qubit circuit with 4 gates**

Here, we present the Fourier coefficients in Fig. 12 for the circuit with 3 qubits and four parameterized gates sketched in Fig. 11. We see that the coefficients \( c_{11} \) and \( c_{1,-1} \) now take similar values with their corresponding coefficients shown in Fig. 5 in orange color.
Fig. 12 The real and imaginary part of the Fourier coefficients for the 3-qubit (light blue color) circuit of Fig. 11

Author Contributions All authors reviewed the manuscript.

Funding ICFO group acknowledges support from: ERC AdG NOQIA; Agencia Estatal de Investigación (R&D project CEX2019-000910-S, funded by MCIN/ AEI/10.13039/501100011033; Plan National FIDEUA PID2019-106911GB-I00, FPI, QUANtera MAQS PCI2019-111828-2, QUANtera DYNAmite PCI2022-132919, Proyectos de I+D+I “Retos Colaboración” QUSPIN RTC2019-007196-7), MCIN via European Union NextGenerationEU (PRTR); Fundació Cellex; Fundació Mir-Puig; Generalitat de Catalunya through the European Social Fund FEDER and CERCA program (AGAUR Grant No. 2017 SGR 134, QuantumCAT U16-011424, co-funded by ERDF Operational Program of Catalonia 2014–2020); EU Horizon 2020 FET-OPEN OPTOlogic (Grant No 899794); National Science Centre, Poland (Symfonia Grant No. 2016/20/W/ST4/00314); European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 847517.

Availability of data and materials The simulations were made with the Pennylane library (Bergholm et al. 2018). Code to reproduce the results and explore further settings can be found in the following Github repository: https://github.com/KaterinaGratsea/Teacher-student_scheme-part-2.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abbas A, Sutter D, Zoufal C, Lucchi A, Figalli A, Woerner S (2021) The power of quantum neural networks. Nature Comput Sci 1(6):403–409. https://doi.org/10.1007/s43588-021-00084-1

Alcazar J, Perdomo-Ortiz A (2021) Enhancing combinatorial optimization with quantum generative models. https://doi.org/10.48550/ARXIV.2101.06250

Benedetti M, Garcia-Pintos D, Perdomo O, Leyton-Ortega V, Nam Y, Perdomo-Ortiz A (2019) A generative modeling approach for benchmarking and training shallow quantum circuits. NPJ Quantum Inf 5(1). https://doi.org/10.1038/s41534-019-0157-8

Bergholm V, Izaac J, Schuld M, Gogolin C, Alarn MS, Ahmed S, Arrazola JM, Blank C, Delgado A, Jahangiri S, et al (2018) Pennylane: Automatic differentiation of hybrid quantum-classical computations. arXiv preprint arXiv:1811.04968

Bishop LS, Bravyi S, Cross A, Gambetta JM, Smolin J (2017) Quantum volume. Technical report

Cai Z (2020) Resource estimation for quantum variational simulations of the hubbard model. Phys Rev Appl 14(1). https://doi.org/10.1103/physrevapplied.14.014059

Cao Y, Romero J, Olson JP, Degroote M, Johnson PD, Kieferová M, Kivlichan ID, Menke T, Peropadre B, Sawaya NPD, Sim S, Veis L, Aspuru-Guzik A (2019) Quantum chemistry in the age of quantum computing. Chem Rev 119(19):10856–10915. https://doi.org/10.1021/acs.chemrev.8b00803

Casas B, Cervera-Lierta A (2023) Multi-dimensional fourier series with quantum circuits. https://doi.org/10.48550/ARXIV.2302.03389

Dawid A, Arnold J, Requena B, Gresch A, Plodzień M, Donatella K, Nicoli K, Stornati P, Koch R, Büttner M, Okula R, Muñoz-Gil G, Vargas-Hernández RA, Cervera-Lierta A, Carrasquilla J, Dunjko V, Gabrié M, Huembeli P, van Nieuwenburg E, Vicentini F, Wang L, Wetzol SJ, Carleo G, Greplová E, Krems R, Marquardt F, Tomza M, Lewenstein M, Dauphin A (2022) Modern applications of machine learning in quantum sciences. https://doi.org/10.48550/ARXIV.2204.04198

Du Y, Hsieh MH, Liu T, Tao D (2020) Expressive power of parametrized quantum circuits. Phys Rev Research 2:033125. https://doi.org/10.1103/PhysRevResearch.2.033125

Gili K, Mauri M, Perdomo-Ortiz A (2022) Evaluating generalization in classical and quantum generative models. https://doi.org/10.48550/ARXIV.2201.08770
Gratsea A, Huembeli P (2021) Exploring quantum perceptron and quantum neural network structures with a teacher-student scheme. https://doi.org/10.48550/ARXIV.2105.01477
Gratsea A, Kasper V, Lewenstein M (2021) Storage properties of a quantum perceptron. https://doi.org/10.48550/ARXIV.2111.08414
Grimsley HR, Economou SE, Barnes E, Mayhall NJ (2019) An adaptive variational algorithm for exact molecular simulations on a quantum computer. Nat Commun 10(1). https://doi.org/10.1038/s41467-019-10988-2
Hafkerkamp J, Faist P, Kothakonda NBT, Eisert J, Halpern NY (2022) Linear growth of quantum circuit complexity. Nat Phys 18(5):528–532. https://doi.org/10.1038/s41567-022-01539-6
Heimann D, Schönhoff G, Kirchner F (2022) Learning capability of parametrized quantum circuits. https://doi.org/10.48550/ARXIV.2209.10345
Huang HY, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H, McClean JR (2020) Power of data in quantum machine learning. arXiv:2011.01938
Jerbi S, Fiderer LJ, Nautrup HP, Kübler JM, Briegel HJ, Dunjko V (2022) Supervised quantum machine learning models are kernel methods. arXiv:2101.111020
Shul M (2021a) Quantum machine learning beyond kernel methods. https://doi.org/10.48550/ARXIV.2101.13162
Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow JM, Gambetta JM (2017) Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. Nature 549(7671):242–246. https://doi.org/10.1038/nature23879
Lewenstein M, Gratsea A, Riera-Campeny A, Aloy A, Kasper V, Sanpera A (2021) Storage capacity and learning capability of quantum neural networks. Quantum Sci Technol 6(4):045002. https://doi.org/10.1088/2058-9565/ac070f
McClean (2023) Integrating over the unitary group. https://jarrodmclean.com/integrating-over-the-unitary-group/. Accessed 14 Oct 2015
Meyer JJ, Mulsrski M, Gil-Fuster E, Mele AA, Arzani F, Wilms A, Eisert J (2022) Exploiting symmetry in variational quantum learning. https://doi.org/10.48550/ARXIV.2205.06217
Mitarai K, Negoro M, Kitagawa M, Fujii K (2018) Quantum circuit learning. Phys Rev A 98:032309. https://doi.org/10.1103/PhysRevA.98.032309
Moll N, Barkoutsos P, Bishop LS, Chow JM, Cross A, Egger DJ, Filipp S, Fuhrer A, Gambetta JM, Ganzhorn M, Kandala A, Mezzacapo A, Müller P, Riess W, Sales G, Smolin J, Tavernelli I, Temme K (2018) Quantum optimization using variational algorithms on near-term quantum devices. Quantum Sci Technol 3(3):030503. https://doi.org/10.1088/2058-9565/aab822
Nahum A, Vijay S, Haah J (2018) Operator spreading in random unitary circuits. Phys Rev X 8(2). https://doi.org/10.1103/physrevx.8.021014
Nielsen MA, Chuang IL (2000) Quantum computation and quantum information. Cambridge
Ostaszewski M, Grant E, Benedetti M (2021) Structure optimization for parameterized quantum circuits. Quantum 5,391. https://doi.org/10.22331/q-2021-01-28-391
Peruzzo A, McClean J, Shadbolt P, Yung MH, Zhou XQ, Love PJ, Aspuru-Guzik A, O’Brien JL (2014) A variational eigenvalue solver on a photonic quantum processor. Nat Commun 5(1). https://doi.org/10.1038/ncomms5213
Peters E, Schuld M (2022) Generalization despite overfitting in quantum machine learning models. https://doi.org/10.48550/ARXIV.2209.05523
Pérez-Salinas-Salinas A, Cervera-Lierta A, Gil-Fuster E, Latorre JI (2020) Data re-uploading for a universal quantum classifier. Quantum 4:226. https://doi.org/10.22331/q-2020-02-06-226
Roberts DA, Stanford D, Streicher A (2018) Operator growth in the SYK model. J High Energy Phys 2018(6). https://doi.org/10.1007/jhep06(2018)122
Schuld M (2021a) Quantum machine learning models are kernel methods. https://doi.org/10.48550/ARXIV.2101.111020
Schuld M, Killoran N (2022) Is quantum advantage the right goal for quantum machine learning? https://doi.org/10.48550/ARXIV.2203.01340
Schuld M, Bocharov A, Svore KM, Wiebe N (2020) Circuit-centric quantum classifiers. Phys Rev A 101:032308. https://doi.org/10.1103/PhysRevA.101.032308
Schuld M, Sweke R, Meyer J (2020b) The effect of data encoding on the expressive power of variational quantum machine learning models. arXiv:2008.08605
Shen H, Zhang P, You YZ, Zhai H (2020) Information scrambling in quantum neural networks. Phys Rev Lett 124:200–504. https://doi.org/10.1103/PhysRevLett.124.200504
Sim S, Johnson PD, Aspuru-Guzik A (2019) Expressibility and entanglement capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Adv Quant Technol 2(12):1900070. https://doi.org/10.1002/qute.201900070
Tang HL, Shkolnikov V, Barron GS, Grimsley HR, Mayhall NJ, Barnes E, Economou SE (2021) Qubit-ADAPT-VQE: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor. PRX Quantum 2(2). https://doi.org/10.1103/prxquantum.2.020310
Wu Y, Zhang P, Zhai H (2021) Scrambling ability of quantum neural network architectures. Phys Rev Res 3(3). https://doi.org/10.1103/physrevresearch.3.032057

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.