The effect of data encoding on the expressive power of variational quantum machine learning models

September 21, 2023

Summarize article [1] by : Yaniv Hajaj, Asaf Ben Or, Almog Sharoni and Shir Cohen

Abstract:

By treating parametric quantum circuits as models that map data input to predictions, quantum computers have shown a potential mode for supervised learning.

The expressive power of parametric quantum circuits as function approximations is directly affected by how the data is encoded in the model. It seems that a quantum model can naturally be written as a partial Fourier series in the data, where the accessible frequencies are determined by the nature of the data encoding gates in the circuit. By repeating simple data encoding gates multiple times, these quantum models can access a richer frequency spectrum and, under certain conditions, can be used as universal function approximations.

The study shows that there are quantum models that can realize all possible sets of Fourier coefficients, therefore, if the accessible frequency spectrum is asymptotically rich enough, such models are approximations of universal functions.

Overall, the study provides insights into the relationship between data encoding strategies and the expressive power of quantum models in machine learning. It discusses the limitations and capabilities of quantum circuits for function approximation and offers a theoretical perspective on quantum machine learning.

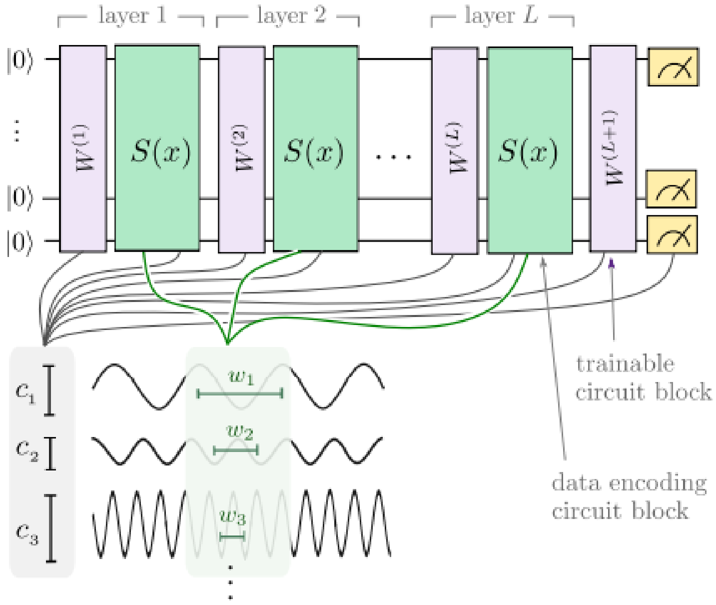
Introduction:

In quantum machine learning, quantum gates are used to encode data inputs $x = (x_1 \dots x_N)$ as well as trainable weights $\theta = (\theta_1 \dots \theta_M)$.

The circle is measured several times to estimate the expectation of the observed part, and the result is interpreted as a prediction. The overall computation implements a "quantum model function" $f_\theta(x)$, a machine learning model based on quantum computation.

Quantum models consisting of layers of trainable circuit blocks W=W(θ) and blocks of data-encoded circuits S(x) can be written as a weighting sum $\sum_\omega c_\omega \, e^{i\omega x}$.

as shown in the main result of the paper shown for one-dimensional inputs x :



The frequencies ω are determined by the data coding circuit, while the coefficients c_ω are determined by the rest of the circuit architecture. When the frequencies ω are integer-valued (or multiples of the fundamental frequency ω_0), the sum becomes a partial Fourier series, which allows us to systematically learn properties of the class of functions that a given quantum model can learn.

In this article we consider standard models from the literature consisting of several "circuit layers". Each layer consists of a data encoding block and a trainable block, and it is assumed that the input features $x \in \mathbb{R}$ are encoded by gates of the form e^{ixH}, where H is an arbitrary Hamiltonian. The main tool is the natural representation of quantum models such as a Fourier-type sum $f_\theta(x) = \sum_{\omega \in \Omega} c_\omega \, e^{i\omega x}$, where ωx is the inner product.

The frequency spectrum $\Omega \in \mathbb{R}^N$ is solely determined by the eigenvalues of the data encoding Hamiltonian while the design of the entire circuit controls the coefficients C_ω that a quantum model can realize.

Research Question & Answer:

Can quantum models express any function on the input x, or are they limited to a certain class of functions?

The basic idea of writing quantum models as partial Fourier series:

We define a quantum (univariate) model $f_\theta(x)$ as the expectation value of some observable with respect to a state prepared by means of a parametric quantum circuit, i.e

$$(1) \quad f_\theta(x) = \langle 0|U^\dagger(x,\theta) M U(x,\theta)|0\rangle$$

where $|0\rangle$ is some initial state of the quantum computer, $U(x,\theta)$ is a quantum circuit depending on the input x and a (possibly empty) set of parameters θ, and M is an observable part.

We want to focus on the role of the data encoding and avoid further assumptions about how the trainable circuit blocks are parameterized, so we view the trainable circuit blocks as arbitrary unitary operations, $W(\theta) = W$, and drop the f_θ subscript from here on. Under this assumption, the total quantum circuit has the form

$$(2) \quad U(x) = W^{(L+1)}S(x)W^{(L)} \ldots W^{(2)}S(x)W^{(1)}$$

Our goal is to write f as a partial Fourier series $f(x) = \sum_{n\in\Omega} c_n\, e^{inx}$ with integer valued frequencies.

The first step is to note that one can always find an eigenvalue decomposition of generator Hamiltonian $H = V^\dagger \Sigma V$ where Σ is a diagonal operator containing H's eigenvalues $\lambda_1 \ldots \lambda_d$ on its diagonal. The data encoding unitary becomes $S(x) = V^\dagger \mathrm{e}^{-\mathrm{i}x\Sigma} V$, and we can "absorb" $VV^\dagger$ into the arbitrary unitaries $W_0 = VWV^\dagger$. Hence, without loss of generality we will assume that H is diagonal which allows us to separate the data-dependent expressions from the rest of the circuit in each component i of the quantum state $U(x)|0\rangle$,

$$(3) \quad [U(x)|0\rangle]_i = \sum_{j_1 \ldots j_L = 1}^{d} e^{-i(\lambda_{j_1} + \cdots + \lambda_{j_L})x} \times W_{ij_L}^{(L+1)} \ldots W_{j_2 j_1}^{(2)} W_{j_1 1}^{(1)}$$

To make the notation easier we introduce the multi-index $j = \{j_1 \ldots j_L\} \in [d]^L$, where $[d]^L$ denotes the set of any L integers between $1, \ldots, d$. Then we can denote the sum of eigenvalues for a given j by $\Lambda_j \;=\; \lambda_{j1} \;+\cdot\cdot\cdot + \lambda_{jL}$, and write

$$(4) \quad [U(x)|0\rangle]_i = \sum_{j\in[d]^L} e^{-i\,\Lambda_j x} \times W_{ij_L}^{(L+1)} \ldots W_{j_2 j_1}^{(2)} W_{j_1 1}^{(1)}$$

To account for the full quantum model we must take into account the complex conjugation of this expression as well as the measurement, and get

$$(5) \quad f(x) = \sum_{k,j\in[d]^L} e^{-i(\Lambda_k - \Lambda_j)x} a_{k,j}$$

where the $a_{k,j}$ contain the terms stemming from the arbitrary unitaries and measurement,

$$(6) \quad a_{k,j} = \sum_{i,i`} (W^*)_{ik_1}^{(1)} (W^*)_{j_1 j_2}^{(2)} \ldots (W^*)_{j_L i}^{(L+1)} M_{i,i`} \times W_{i` j_L}^{(L+1)} \ldots W_{j_2 j_1}^{(2)} W_{j_1 1}^{(1)}$$

The second step consists of grouping all the terms in the sum (5) whose basis function $e^{i(\Lambda_k - \Lambda_j)x}$ have the same frequency $\omega = \Lambda_k - \Lambda_j$. All frequencies accessible to the quantum model are included in its frequency spectrum

$$.(7) \quad \Omega = \{\Lambda_k - \Lambda_j, k, j \in [d]^L\}$$

This yields,

$$(8) \quad f(x) = \sum_{\omega\in\Omega} c_\omega\, e^{i\omega x}$$

where the coefficients are obtained by summing over all $a_{k,j}$ contributing to the same frequency

$$(9) \quad c_\omega = \sum_{\substack{k,j \in [d]^L \\ \Lambda_k - \Lambda_j = \omega}} a_{k,j}$$

The frequency spectrum Ω has the following important properties: $0 \in \Omega$, and for every frequency $\omega \in \Omega$, we also have $-\omega \in \Omega$. In addition, since $c_\omega = c^*_{-\omega}$, equation (8) realizes a real valued function.

<u>The expressivity of quantum models</u>

We will continue to use the Fourier series formalism to investigate the expressivity of quantum models. We begin with an analysis of the popular strategy of using single Pauli rotations in the S(x) encoding subroutine to demonstrate the practical value of the approach.

As a "warm-up" application of the Fourier series formalism, we begin by considering a simple quantum model with L = 1, where we use a single-qubit gate $\mathcal{G}(x) = e^{-ixH}$ to encode the input x into the circuit,

$$.(10) \quad U(x) = W^{(2)}\, \mathcal{G}(x) W^{(1)}$$

H has two distinct eigenvalues (λ_1, λ_2) as a single qubit gate generator. It is possible without loss of generality to always rescale the energy spectrum to $(-\gamma, \gamma)$ since the global phase is unobservable. We note that the class of such coding gates includes Pauli rotations, where $H = (1/2)\sigma$ for $\sigma \in \{\sigma x, \sigma y, \sigma z\}$, for which $\gamma = \frac{1}{2}$.

We aim to show that models of the type (10) always lead to functions of the form $f(x) = A\, sin(2\gamma x + B) + C$ where A, B, C are constants determined by the uncoded part of the variational circuit, which reproduces the previous observation. A sine function can be described by a truncated Fourier series of degree 1, we will see how the degree can be systematically increased by repeating the encoding gate.

First, we can assume without loss of generality that the eigenvalues of H are always $\lambda_1 = -1, \lambda_2 = 1$, since we can absorb the factor γ into the data input by rescaling it using $\tilde{x} = \gamma x$. From equation (7) we can see that the spectrum of the quantum model is given by $\Omega = \{-2,0,2\}$ (since the possible differences $\lambda_{k_1} - \lambda_{j_1}$ for $\lambda_{k_1}, \lambda_{j_1} = \in \{-1,1\}$ are $-1 - (1\,), -1-(-1), 1-(1), 1\, -\, (-1)$). The Fourier coefficients in equation (9) become

$$(11) \quad c_0 = \sum_{i,i'} M_{i,i'} (W^*)^{(1)}_{12} (W^*)^{(2)}_{2i} (W)^{(2)}_{i'1} (W)^{(1)}_{11}$$

$$(12) \quad c_2 = \sum_{i,i'} M_{i,i'} (W^*)^{(1)}_{11} (W^*)^{(2)}_{1i} (W)^{(2)}_{i'2} (W)^{(1)}_{21}$$

$$(13) \quad c_{-2} = c_2^*$$

and the frequency spectrum of the quantum model consists of a single non-zero frequency:

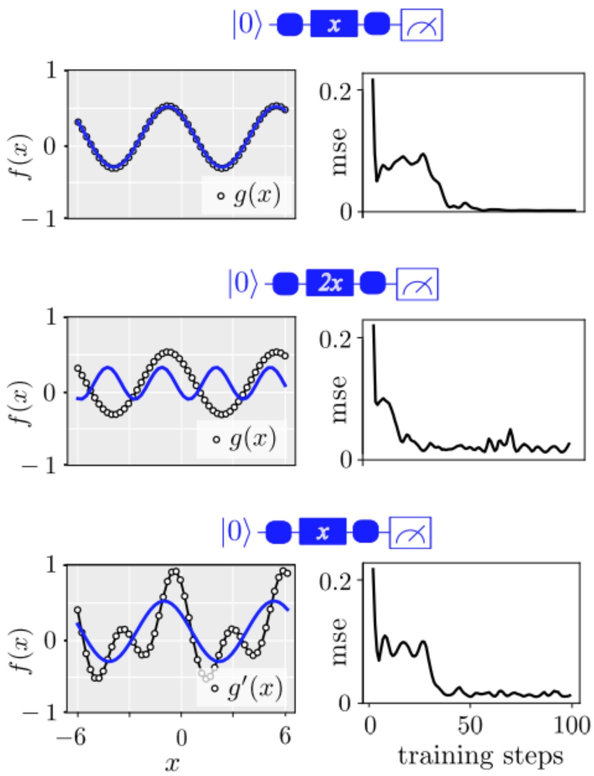
$$f(x) = c_{-2} e^{i2\tilde{x}} + c_0 + c_2 e^{-i2\tilde{x}} = c_0 + 2|c_2| \cos\,(2\tilde{x} - \arg(c_2))$$

where $arg(c_2)$ is the complex phase of c_2. For Pauli rotations, we have $\tilde{x} = \gamma x = \frac{x}{2}$.

And with $A = 2|c2|,\ B = -\frac{\pi}{2} - arg(c2),\ C = c0$ we can calculate equation (4).

Note that we have not assumed anything about the number of qubits, the nature of the unitaries W, or the measurement M. This ilustrates a very important point: even with the ability to implement very wide and deep quantum circuits (which may even be classically intractable to simulate), the expressivity of the appropriate quantum model fundamentally limited by the data encoding strategy.

The following figure (FIG 3 in the article [1], we also reconstruct it in our code) supports this finding and presents numerical evidence: Encoding data using a Pauli-X rotation results in a quantum model that can only learn to fit a Fourier series of a single frequency—and only if that frequency is tuned exactly the way the data is tuned.



We consider here a parametric quantum model trained with data samples to fit a target function $g(x) = \sum_{n=-1}^{1} c_n e^{-nix}$ or $g'(x) = \sum_{n=-2}^{2} c_n e^{-nix}$ with coefficients $c_0 = 0.1, c_1 = c_2 = 0.15 - 0.15i$.

The variational circuit is of form of $f(x) = \langle 0|U^{\dagger}(x)\sigma_z U(x)|0\rangle$ where $|0\rangle$ is a single qubit, and $U = W^{(2)}R_x(x)W^{(1)}$.

Let's explain, the left panels show the quantum model function $f(x)$ and the target function $g(x), g'(x)$, while the right panels show the mean squared error between the data sampled from g and f during a typical training run. Given the input x as is, the quantum model easily fits a target of rank 1. When the inputs are rescaled $x \rightarrow 2x$ this causes a frequency mismatch, and the model cannot learn the target anymore. However, even when the scale is correct, the variational circuit cannot fit the objective function of degree 2.

The complex phase of C(2) is given by arg(C2) For Pauli rotations $x_{\sim} = \gamma x = \frac{x}{2}$

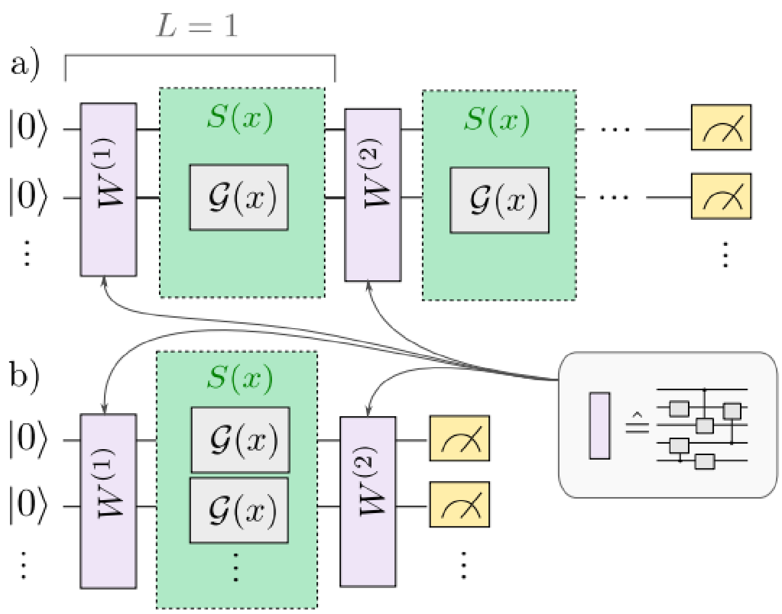
This aligns with the results from another reference which is brang in the article [2] where

$A = 2|c2|, B = -\frac{\pi}{2} - \arg(c2), fc0.$

The article [1] doesn't make any assumptions about the number of qubits, the properties of the unitaries W or the measurement M.

A key point emphasized is that the potential of quantum circuits, no matter how extensive, is constrained by the data encoding approach. Supporting this, the figure above reveals that using a Pauli-X rotation for data encoding allows a quantum model to only fit a Fourier series of one frequency, which is contingent on the precise data scaling.

Repeated Pauli encodings linearly extend the frequency spectrum The next step is to extend the accessible frequency spectrum of a quantum model The writer of the article [1] are doing that in two deferment ways (as you can see in the picture below):



The initial method (referenced as "a" in the image) involves employing multi-layered models where L is greater than 1. In this setup, the encoding gate is sequentially replicated r = L times. By doing this, one can progressively augment the truncated Fourier series degree to r. The alternative method (indicated as "b" in the image) applies a singular-layer model where L is equal to 1, but the encoding gate is duplicated r times simultaneously. We will begin by delving into the details of the latter method (denoted as "b" in the image). This method is a specific instance of the foundational model depicted in the Eq. (1) (equation 3 from the article [1])

and with L = 1

$$S(x) = e^{-i\frac{x}{2}\sigma_r} \otimes \ldots \otimes e^{-i\frac{x}{2}\sigma_1}, \quad := e^{-ixH} \tag{15}$$

Where $\sigma j \in \{\sigma x, \sigma y, \sigma z\}$

Understanding that all rotation gates operate on distinct qubits and therefore commute, lets us diagonalize each rotation gate on its own. This, in turn, enables us to diagonalize H. This brings us to:

$$\begin{aligned} S(x) &= V_r e^{-i\frac{x}{2}\sigma_z} V_r^\dagger \otimes \ldots \otimes V_1 e^{-i\frac{x}{2}\sigma_z} V_1^\dagger, \\ &= V \exp\left(-i\frac{x}{2}\sum_{q=1}^{r}\sigma_z^{(q)}\right) V^\dagger, \\ &:= V e^{-ix\Sigma} V^\dagger, \end{aligned} \tag{16}$$

Where $\sigma z^{(q)}$ It's the diagonal r-qubit operator that functions non-trivially, demonstrated by: σ_z solely on the q'th qubit. Determining this provides us with: Σ = diag (λ_1,…,λ_2^r),

with the unique entries numbering r + 1

$$\lambda_p = \left(\frac{p}{2} - \frac{r-p}{2}\right) = p - \frac{r}{2},\ p \in \{0, \ldots, r\}, \tag{17}$$

comprising all potential combinations of r values either +1/2 or -1/2. Considering the subsequent Eq. (7) (equation 10 from the article [1]), For L = 1, the frequency spectrum includes the differences between any pair of these eigenvalues, leading us to:

$$\begin{aligned} \Omega_{\mathrm{par}} &= \{\lambda_{k_1} - \lambda_{j_1} \,|\, k_1, j_1 \in \{1, \ldots, 2^r\}\} \\ &= \Big\{\left(p - \frac{r}{2}\right) - \left(p' - \frac{r}{2}\right) \Big| \\ &\qquad\qquad p, p' \in \{0, \ldots, r\}\Big\} \\ &= \{p - p' \,|\, p, p' \in \{0, \ldots, r\}\} \\ &= \{-r, -(r-1), \ldots, 0, \ldots, r-1, r\}. \end{aligned} \tag{18}$$

This explains why a univariate quantum model featuring r concurrent Pauli-rotation encodings can be represented as a truncated Fourier series up to degree r. Reflecting on the initial method we discussed — (a in the illustration) — which is a single-qubit Pauli rotation encoding applied in layers, it exhibits a similar scaling impact. Examine the quantum model in the above Eq. (1) and Eq. (2) For a given value of L, where L=r-1 layers, where "r" is a specified number of layers and S(x) = exp($-i$ (x/2) σj) It's a solitary qubit Pauli rotation that operates on the identical qubit within every layer.

The circuit in Eq. (2) is transformed to be:

$$U(x) = W^{(L+1)} e^{-i\frac{x}{2}\sigma_L} W^{(L)} \ldots W^{(2)} e^{-i\frac{x}{2}\sigma_1} W^{(1)}. \tag{19}$$

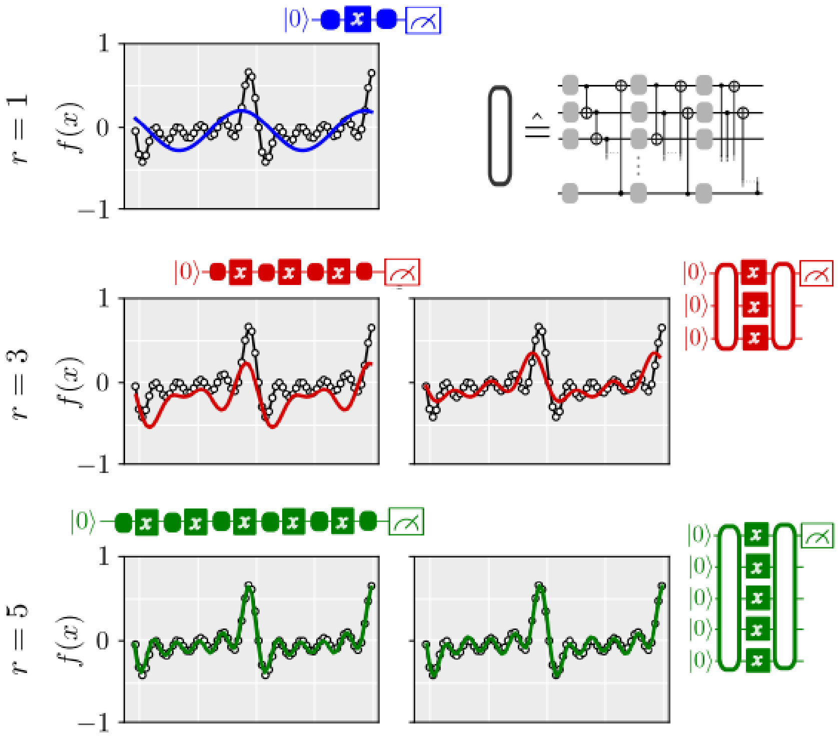
Similar to the previous situation, diagonalizing the Pauli rotations will provide us with:
Σ = (1/2) σz Across all encoding layers.
If we examine Eqs. (7) The frequency spectrum derived from it is a combination of 2r terms, each having a value of either +1/2 or -1/2:

$$\Omega_{\text{seq}} = \{(\lambda_{k_1} + \cdots + \lambda_{k_r}) - (\lambda_{j_1} + \cdots + \lambda_{j_r}) \mid k_1, \ldots, k_r, j_1, \ldots, j_r \in \{1, 2\}\}. \tag{20}$$

Upon calculation, we will observe that Ωseq = Ωpar.

Using r sequential repetitions of single-qubit Pauli encoding in a quantum model equates to a truncated Fourier series of degree r.

The expansion of a quantum model's frequency spectrum through Pauli encodings is shown numerically in the next image:



By viewing quantum models as Fourier summations, we can quickly set expressivity bounds with L repetitions of a dimension d encoding gate.
The spectrum capacity K(L, d) of a quantum model is the total frequencies it can handle, defined as follows:

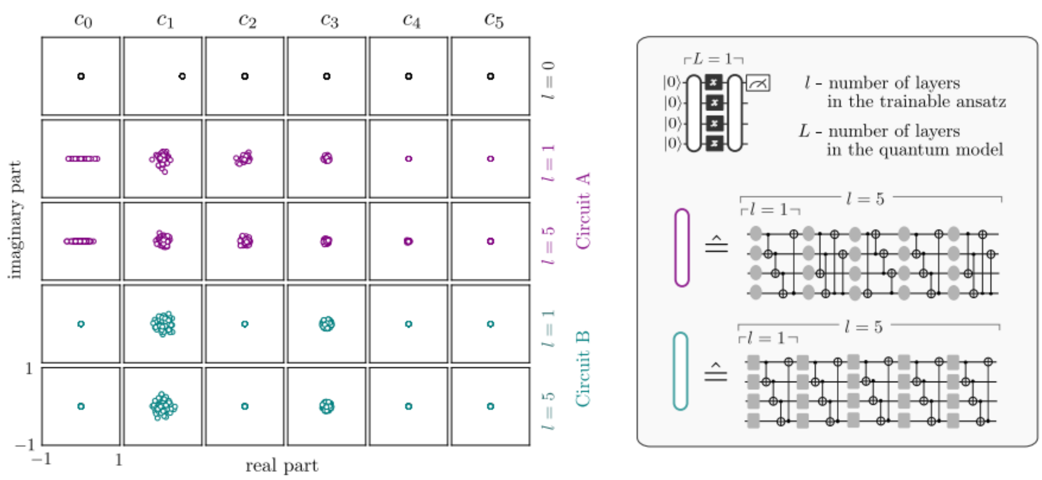
$$\Omega = \{(\lambda_{j_1} + \ldots \lambda_{k_L}) - (\lambda_{j_1} + \cdots + \lambda_{k_L})\} \tag{21}$$

The frequencies consist of 2L terms with d possible values each. Therefore, they can achieve a maximum of d^{2L}unique values, regardless of whether the eigenvalues are real or integer.
The size K represents the pairs: −ω, ω ∈ Ω including one but excluding the "zero frequency", we derive:

$$K \leq \frac{d^{2L}}{2} - 1. \tag{22}$$

Examining the coefficient flexibility is more challenging than exploring the frequency spectrum of a quantum model. Each block $W(1), \dots, W(L+1)$ contributes to every Fourier coefficient.

With the measurement observable, minor gate tweaks can drastically alter many Fourier coefficients. These coefficients, while not random, depend on the circuit's specific freedoms. Only certain structures can tap a limited set of potential Fourier coefficients cn. To control K + 1 complex coefficients, we need at least M ≥ 2K + 1 real degrees of freedom – in other words, parameters θ = (θ_1, … , W_M) – in the quantum circuit.
Simulations show that shallow circuit blocks W yield varied Fourier coefficients. However, an ansatz can inherently zero out specific coefficients, as the figure below illustrates.



An interesting further observation is that the variance of the coefficients decreases with higher orders.

<u>Quantum models are asymptotically universal:</u>

Certain quantum models can be portrayed as partial Fourier series. The frequencies accessible in these series are dictated by the spectra of Hamiltonians driving the data-encoding gates. By using Pauli rotations as an example, the models can achieve a truncated Fourier series when encodings are repeated, as determined by:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \langle 0|U^{\dagger}(\boldsymbol{\theta}, \boldsymbol{x}) M U(\boldsymbol{\theta}, \boldsymbol{x})|0\rangle, \tag{23}$$

Where :

$$U(\boldsymbol{\theta}, \boldsymbol{x}) = W^{(2)}(\boldsymbol{\theta}^{(2)}) S(\boldsymbol{x}) W^{(1)}(\boldsymbol{\theta}^{(1)}), \tag{24}$$

with θ (1) , θ (2) ⊆ θ And

$$S(\boldsymbol{x}) := e^{-ix_1 H_1} \otimes \ldots \otimes e^{-ix_N H_N}. \tag{25}$$

The above model is a natural extension of the univariate L = 1 model we explored in previous section. the quantum model's expressiveness is more than just the accessible frequency spectrum; it's about how flexible one can be when adjusting the contributions of these frequencies, i.e., setting the Fourier coefficients.

Quantum models can achieve any frequency spectrum if given enough repetitions of basic data-encoding gates or sufficiently varied Hamiltonians. These models are "asymptotically universal" in that they can approximate any square-integrable function if we expand the Hilbert space dimensions indefinitely.

Moreover, if one assumes that the trainable circuits can realize any global unitary transformations, this leads to an equivalent model described by:

$$f(\boldsymbol{x}) = \langle \Gamma | S^\dagger(\boldsymbol{x}) M S(\boldsymbol{x}) | \Gamma \rangle \tag{26}$$

The true test of universality comes with the Hamiltonian family, specifically:

$$H_m = \sum_{i=1}^{m} \sigma_q^{(i)}. \tag{27}$$

This Hamiltonian family links to a series of models {fm} through:

$$f_m(\boldsymbol{x}) = \langle \Gamma | S_{H_m}^\dagger(\boldsymbol{x}) M S_{H_m}(\boldsymbol{x}) | \Gamma \rangle \tag{28}$$

For the quantum model to be universal, its Hamiltonian family must contain any integer frequency as it grows. The theorem posits that for every square-integrable function, a quantum model can be found that approximates it with high precision:

$$||f_{m'} - g||_2 \leq \epsilon \tag{29}$$

in essence, with the right conditions, quantum models can represent any function by mimicking its Fourier series.

Discussion:

The findings of this research underscore the pivotal role of data encoding in determining the expressive power of quantum models. By understanding the relationship between quantum models and partial Fourier series, we can gain insights into the function classes these models can represent. This understanding can be instrumental in guiding the design and application of quantum machine learning models.

In the realm of quantum computing, the exploration of quantum models as partial Fourier series is a significant step forward in understanding the theoretical underpinnings of quantum machine learning. The potential of quantum models to act as universal function approximators, given a rich frequency spectrum, is promising.
As quantum computing technology continues to evolve, we anticipate that the expressive power of quantum models will further expand, leading to more sophisticated and powerful machine learning models. This could potentially revolutionize fields like time-series analysis and signal processing, where the Fourier representation is already pivotal. As for the timeline, while it's challenging to pinpoint exact dates, we believe that within the next 10-15 years, we will witness substantial advancements in this domain, especially as the synergy between quantum and classical machine learning techniques becomes more pronounced and when quantum computers will become more stable and less noisy.

Authors contribution:

Shir Cohen: Conceived the general construct of our article and provided insights into the answers.

Almog Sharoni: Originated the idea behind the answers.

Asaf Benor: Drafted the discussion section, conducted a comprehensive review of the entire article, and ensured its accuracy and coherence.

Yaniv Hajaj: implemented the code for reconstruct Fig 3 in the article [1], final review of our entire article.

Shir Almog and Asaf Benor: Reviewed and validated the code to ensure its functionality and accuracy.

All authors: Collaboratively deliberated and made decisions regarding the content and direction of the article.

Code:

The code used to reproduce the figures was obtained with the help of the article [1] CODE and the GITHUB Link it provides:
"Code to reproduce the figures and explore further settings can be found in the following GitHub repository: https://github.com/XanaduAI/expressive_power_of_quantum_models. "

References:

[1] Maria Schuld, Ryan Sweke and Johannes Jakob Meyer ”The effect of data encoding on the expressive power of variational quantum machine learning models”.
[2] Mateusz Ostaszewski, Edward Grant, and Marcello Benedetti, “Quantum circuit structure learning,” arXiv preprint arXiv:1905.09692 (2019).