Universal approximation of continuous functions with minimal quantum circuits

Adrián Pérez-Salinas,[1,2] Mahtab Yaghubi Rad,[1,3] Alice Barthe,[1,4] and Vedran Dunjko[1,3]

[1]⟨aQa⟩^{L} *Applied Quantum Algorithms, Universiteit Leiden*
[2]*Instituut-Lorentz, Universiteit Leiden, Niels Bohrweg 2, 2333 CA Leiden, Netherlands*
[3]*LIACS, Universiteit Leiden, Niels Bohrweg 1, 2333 CA Leiden, Netherlands*
[4]*CERN, Espl. des Particules 1, 1211 Geneva, Switzerland*

The conventional paradigm of quantum computing is discrete: it utilizes discrete sets of gates to realize bitstring-to-bitstring mappings, some of them arguably intractable for classical computers. In parameterized quantum approaches, widely used in quantum optimization and quantum machine learning, the input becomes continuous and the output represents real-valued functions. Various strategies exist to encode the input into a quantum circuit. While the bitstring-to-bitstring universality of quantum computers is quite well understood, basic questions remained open in the continuous case. For example, it was proven that full multivariate function universality requires either (i) a fixed encoding procedure with a number of qubits scaling as the dimension of the input or (ii) a tunable encoding procedure in single-qubit circuits. This reveals a trade-off between the complexity of the data encoding and the qubit requirements. The question of whether universality can be reached with a fixed encoding and constantly many qubits has been open for the last five years. In this paper, we answer this remaining fundamental question in the affirmative. We provide a constructive method to approximate arbitrary multivariate functions using just a single qubit and a fixed-generator parametrization, at the expense of increasing the depth. We also prove universality for a few of alternative fixed encoding strategies which may have independent interest. Our results rely on a combination of techniques from harmonic analysis and quantum signal processing.

a. Introduction — Quantum computing extends the paradigm of classical circuits by adding to the set of available logical operations new elements, as allowed by the principles of quantum mechanics. The conventional paradigm of quantum computing is still discrete and relies on a discrete set of gates to realize bitstring-to-bitstring mappings. Some of these mappings can be arguably intractable for classical computers [1]. Motivated by near-term quantum computing limitations and common implementations, recently there has been substantial interest in settings where quantum computers are used to realize continuous-valued functions depending on continuous tunable parameters, which we refer to as variational quantum computing. Such variational approaches may be utilized in a plethora of contexts, including quantum optimization [2–4] or quantum machine learning [5–8], among others. In all cases, an arguably fundamental question that can be asked about function representation frameworks is that of its universality.

For discrete computation, universality is well understood, both in classical and quantum cases. Both classical and quantum computing provide universality in terms of the capacity to represent (or compute) arbitrary boolean functions[1]. In the classical case, Boolean universality can be achieved with just NAND gates and FANOUT [9] and even in constant-width settings [10], by querying n-bit inputs to compute n-bit boolean functions. Quantum computing can do even better by representing arbitrary boolean functions over n bits using just a single-qubit wire [11]. For the continuous classical case, much is known about the representation power of many models such as neural networks [12, 13]. However, for the quantum parametrized circuit representation of continuous functions, certain fundamental questions have remained open.

Various methods to represent continuous functions with quantum circuits have been introduced in literature [5, 8], commonly relying on the application of so-called parametrized gates. In this setting, a parametrized gate is defined by choosing a Hermitian generator H, and a real-valued parameter t. Then, the parametrized gate $\exp(iHt)$ is applied, in analogy to time evolution. Gates of this type are composed to construct a larger quantum circuit of some *architecture* $U(\boldsymbol{\theta}, \boldsymbol{x})$, combining tunable parameters $\boldsymbol{\theta}$ with data x. A family of functions $h_{\boldsymbol{\theta}}(x)$ is determined by applying the circuit to an initial state, and then measuring the expectation value of a fixed observable. Universality for this function family was proven in two settings, namely (i) for fixed generators and the number of qubits scaling with the dimension of the problem [14, 15], and (ii) for tunable-weighted generators wH, for w a real number, and single-qubit wires [16]. These results show a trade-off between the complexity of the encoding strategy and the cost in qubit numbers, which has practical and theoretical repercussions. These proof techniques differ in nature, and there seems to be no direct way to adapt the strategies of one or the other. The fundamental question of whether it is possible to achieve both, i.e. fixed-generator universality, with just a single qubit, in analogy to the discrete case [11] remained

[1] In the quantum case, we also have the universality of (approximately) representing arbitrary unitary transformations, which is also guaranteed for any so-called universal gate set, but this is tangential to our discussion

open.

In this work, we resolve the aforementioned open question in the affirmative and show that single-qubit quantum parametrized circuits with fixed generators are universal for all multivariate continuous functions. See Figure 1 for a graphical description. To achieve universality under these conditions, our construction demands deeper circuits. Our proof techniques rely on approximating the tunable-weighted generators with fixed encodings to arbitrary precision and then combining these approximations for universality. Such approximation is possible via a combination of techniques from quantum signal processing and harmonic analysis.

b. Background on universality — Universality has been a central concept in function analysis Parameterizing arbitrary functions allows us to construct controllable function families with the purpose of implementing arbitrary functions. Consider a family of parameter-dependent maps $\mathcal{G}_\theta$ acting on a domain space $\mathcal{F}_\theta : \mathcal{X} \to \mathcal{Y}$. The challenge is to find a set of parameters θ^* such that the map $g_{\theta^*} \in \mathcal{G}_\theta$, when applied to any point in $\mathcal{X}$ fulfills $g_{\theta^*}(x) \approx f(x)$, for arbitrary $f(x)$. Expressivity of the function families quantifies the range of maps that the family $\mathcal{G}_\theta$ is capable of implementing. Maximal expressivity is commonly referred to as universality. This property guarantees that $\mathcal{G}_\theta$ is capable of capturing arbitrary functions, assuming perfect optimization of $\boldsymbol{\theta}$[2]. The formal definition of universality is as follows.

Definition 1 (Universality). *Let $\mathcal{G} = \{g(\boldsymbol{x})\}, \mathcal{F} = \{f(\boldsymbol{x})\}$, $g, f : \mathbb{R}^m \to \mathbb{C}$ be two sets of functions. Then $\mathcal{G}$ is universal with respect to $\mathcal{F}$ if*

$$\forall f(\boldsymbol{x}) \in \mathcal{F} \quad \exists g(\boldsymbol{x}) \in \mathcal{G} \text{ s.t. } \quad \|f - g\|_p \leq \varepsilon, \tag{1}$$

for arbitrarily small $\varepsilon > 0$, where $\|\cdot\|_p$ is the p-norm for L_p functions.

A central result on the universality of functions is the Fourier theorem [17], lying at the core of foundational results in functional analysis.

Theorem 1 (Fourier theorem [17]). *Let $\mathcal{G}_N^F = \{g_N^F(x)\}$ be the set of functions $g_N^F : [0,1]^m \to \mathbb{C}$ of the form*

$$g_N^F(\boldsymbol{x}) = \sum_{\boldsymbol{n}} c_{\boldsymbol{n}} e^{i2\pi \boldsymbol{n}\cdot\boldsymbol{x}} \tag{2}$$

for $\boldsymbol{n}$ integers, and $c_{\boldsymbol{n}}$ complex values. $\mathcal{G}$ is universal with respect to square-integrable periodic functions in the norm $\|\cdot\|_2$.

The universal approximation theorem (UAT) is another universality theorem, proving that single-layer neural networks are universal. UAT provides robustness to the field of ML.

[2] In this manuscript, we will not discuss optimization strategies or problems.

Theorem 2 (Universal approximation theorem, adapted from [12]). *Let $\mathcal{G}_N^{\rm UAT} = \{g_N^{\rm UAT}(x)\}$ be the set of functions $g_N^{\rm UAT} : [0,1]^m \to \mathbb{C}$ of the form*

$$g_N^{\rm UAT}(x) = \sum_{n=1}^{N} \alpha_n \exp\left(i(\boldsymbol{w}_n \cdot \boldsymbol{x} + \phi_n)\right), \tag{3}$$

for $\boldsymbol{w}_n, \alpha_n, \phi_n$ reals. $\mathcal{G}$ is universal with respect to continuous functions in the domain $[0,1]^m$ in the norm $\|\cdot\|_\infty$.

We remark that these theorems will be useful for our results, but the list is by no means an exhaustive representation of existing results.

c. Universality in parameterized quantum circuits — Quantum circuits can also represent continuous functions when using the proper ansatzes, usually discussed in the context of QML. The data $\boldsymbol{x}$ is introduced into the quantum processing pipeline, together with tunable unitary gates, yielding a global operation $U(\boldsymbol{\theta}, \boldsymbol{x})$. The hypothesis functions defined by the quantum model are

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \langle 0| U_L(\boldsymbol{\theta}, \boldsymbol{x}) |0\rangle\,, \tag{4}$$

Notice that the outputs of the functions are overlaps, which can only be approximately computed through measurements. In most practical implementations, a measurement with respect to an observable is considered. The definition above is however more general and will be used in this manuscript.

We call a quantum circuit universal if the family of hypothesis functions $\mathcal{H} = \{h_{\boldsymbol{\theta}}(\boldsymbol{x})\}$ is universal, in the conditions of Definition 1. To provide universality to the quantum model, data is introduced through repeated calls to a data-dependent gate. The data re-uploading framework [8] comprises all circuits of the latter kind, described as follows.

$$U_L(\boldsymbol{\theta}; \boldsymbol{x}) = \prod_{j=1}^{L} R(\theta_j) V(\boldsymbol{x}), \tag{5}$$

where $R(\cdot)$ is a parameterised gate and $V(\boldsymbol{x})$ is a possibly parameter-dependent encoding gate.

We cover now two relevant results on universality for data re-uploading models.

Theorem 3 (Re-uploading: fixed encoding gates, adapted from [18, 19]). *Consider the single-qubit circuit*

$$U_L^{\rm f}(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; x) = \left(\prod_{j=1}^{L} e^{(i\sigma_z\theta_j)} e^{(i\sigma_y\phi_j)} e^{(i\sigma_z 2\pi x)}\right) e^{(i\sigma_z\theta_0)} e^{(i\sigma_y\phi_0)} e^{(i\sigma_z\lambda)}, \tag{6}$$

with $\sigma_{\{y,z\}}$ being Pauli matrices, and $x \in [0,1]$. Then $h_L(x) = \langle 0| U_L^{\rm f}(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; x) |0\rangle$ is universal with respect to square-integrable functions in the norm $\|\cdot\|_2$.

The proof can be found in Appendix 3.

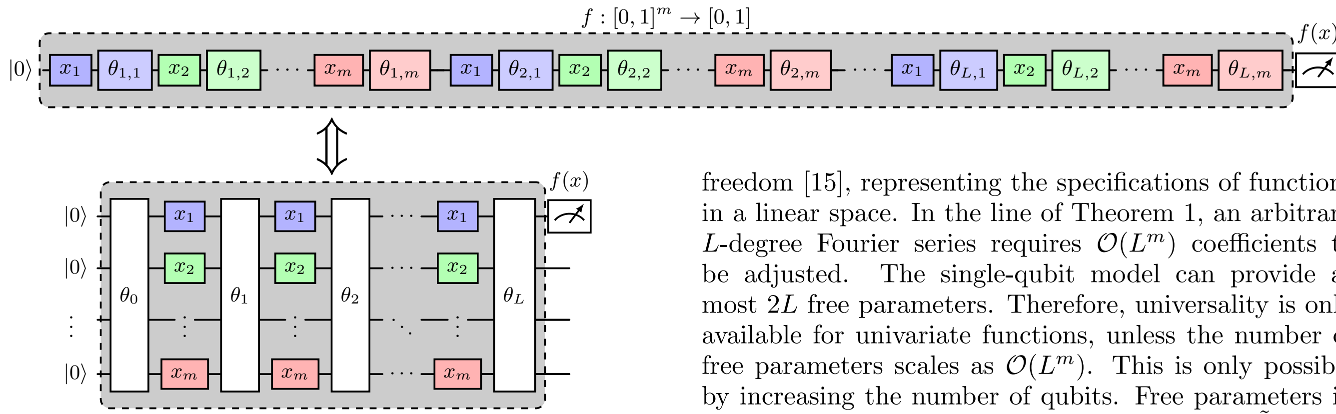

FIG. 1. Quantum circuits with parameters and fixed encoding gates can represent multivariate continuous functions with multiple qubits. In this paper, we show that single-qubit quantum circuits retain universal representation capabilities.

Theorem 4 (Re-uploading: tunable encoding gates [16])**.** *Consider the single-qubit circuit*

$$U_L^{\mathrm{w}}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \boldsymbol{x}) = \prod_{j=1}^{L} e^{(i\sigma_z \theta_j)} e^{(i\sigma_y \phi_j)} e^{(i\sigma_z \boldsymbol{w}_j \cdot \boldsymbol{x})} e^{(i\sigma_z \theta_0)} e^{(i\sigma_y \phi_0)} e^{(i\sigma_z \lambda)}, \quad (7)$$

with $\sigma_{\{y,z\}}$ being Pauli matrices, and $\boldsymbol{x} \in [0,1]^m$. Then $h_L(x) = \langle 0| U_L^{\mathrm{w}}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \boldsymbol{x}) |0\rangle$ is universal with respect to continuous functions in the norm $\|\cdot\|_\infty$.

Both results in Theorem 3 and Theorem 4 provide universality under different conditions, with different implications. It is worth it to investigate these differences in detail. These observations will motivate the main question of the present manuscript: *is it possible to guarantee universality with fixed encoding gates?*, which we answer affirmatively in Section *d*.

First of all, Theorem 3 is at the core based on Theorem 1, while Theorem 4 relies on Theorem 2. This difference triggers many other technical yet important differences between the two results, such as $\|\cdot\|_2$ or $\|\cdot\|_\infty$ distances between functions. Note that $\|f - g\|_2 \leq \|f - g\|_\infty$, and therefore Theorem 2 imposes stronger constraints in convergence than Theorem 1. Another important component is the periodicity of the functions. Theorem 3 is by construction representing periodic functions. Hence, it cannot accurately capture the functions at the limits of the domain $x \in [0,1]$ in general, unless the target function is continuous everywhere, in particular, $f(0) = f(1)$ [20]. This is extensible to continuous functions by symmetrizing the function and appropriately adapting the domain.

A prominent difference is the dimensionality of the functions. Theorem 3 provides universality *only* for univariate functions, while Theorem 4 does so for multivariate functions. To understand this phenomenon, it is possible to argue through counting degrees of freedom [15], representing the specifications of functions in a linear space. In the line of Theorem 1, an arbitrary L-degree Fourier series requires $\mathcal{O}(L^m)$ coefficients to be adjusted. The single-qubit model can provide at most $2L$ free parameters. Therefore, universality is only available for univariate functions, unless the number of free parameters scales as $\mathcal{O}(L^m)$. This is only possible by increasing the number of qubits. Free parameters in n-qubit systems increase as $2^n L$, imposing $n \in \tilde{\mathcal{O}}(m)$, in agreement with existing works [15]. On the other hand, Theorem 2 guarantees universality for m-dimensional functions with only $\mathcal{O}(mL)$ free parameters, matching the reported resource requirements in Theorem 4. The reason for this difference is deeply connected to the presence of tunable weights, which allow to cover the space of functions much more efficiently as compared to fixed weights. We refer the interested reader to Theorem 8 and Theorem 9 in the Appendix for more details.

d. Universality with fixed encoding — Universality for multivariate functions with fixed encoding gates is possible, albeit in a less direct way than the results given up until now. To demonstrate it, we argue that it is possible to approximate encoding gates with tunable weights to arbitrary accuracy making use only of fixed encoding gates. This first result is leveraged to show that the distance $\|\cdot\|_\infty$ between hypothesis functions with tunable and encoding gates can be made arbitrarily small, thus reaching universality.

We begin by defining a multivariate single-qubit re-uploading model with fixed encoding gates.

Definition 2 (Multivariate fixed re-uploading)**.** *A single-qubit multivariate data re-uploading circuit with fixed encoding gates is given by*

$$U_L^{\mathrm{mf}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; \boldsymbol{x}) = \left(\prod_{j=1}^{L} \prod_{k=1}^{m} \prod_{n=1}^{N} e^{i\theta_{j,k,n}\sigma_z} e^{i\phi_{j,k,n}\sigma_y} e^{i\pi x_k \sigma_z} \right) e^{i\theta_0 \sigma_z} e^{i\phi_0 \sigma_y} e^{i\lambda \sigma_z}, \quad (8)$$

where $\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda$ are real and tunable.

We will refer to the output function of the circuit in Definition 2 as

$$h_L^{\mathrm{f}}(\boldsymbol{x}) = \langle 0| U_L^{\mathrm{mf}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; \boldsymbol{x}) |0\rangle . \quad (9)$$

The road towards proving universality is to show that circuits as given in Equation (8) can approximate circuits of the form given in Equation (7), with arbitrary precision. In other words, one may approximate any

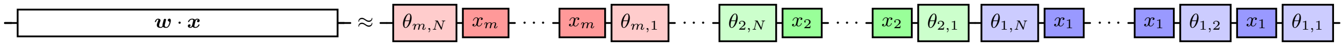

FIG. 2. Quantum circuit allowing the universality of quantum re-uploading models, with fixed encoding gates (in this figure represented by $\{x_1, x_2, \ldots, x_m\}$). By Theorem 5, we show that gates of the form $e^{iwx\sigma_z}$ can be approximated to arbitrary accuracy by a re-uploading circuit with fixed encoding gates. The successive application of these approximations allow for an approximation of $e^{i\boldsymbol{w}\cdot\boldsymbol{x}\sigma_z}$, where $\boldsymbol{x}$ is now multidimensional.

tunable encoding gate with enough layers of fixed encoding gates.

The first step is to approximate gates of the form $e^{iwx\sigma_z}$, where x is the input and w is a tunable real weight. Inspection of Definition 2 allows us to identify the condition

$$R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) \equiv \prod_{n=1}^{N} e^{i\theta_{j,k,n}\sigma_z} e^{i\phi_{j,k,n}\sigma_y} e^{i\pi x_k \sigma_z} \approx e^{iw_k x_k \sigma_z}, \quad (10)$$

as a requirement for the approximation. In the previous equation, the parameters $\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}$ will be chosen to match the right hand for any $w_k \in \mathbb{R}$ to a precision ε. This intuition is supported by the following result.

Theorem 5. *For a gate $R(\boldsymbol{\theta}, \boldsymbol{\phi}, x)$ as given in Equation (10), for any $w \in \mathbb{R}$ and for any $\varepsilon > 0$, there exists a value N such that*

$$\exists\, (\boldsymbol{\theta}, \boldsymbol{\phi}) \text{ s.t.} \sup_{x\in[0,1]} \|R(\boldsymbol{\theta}, \boldsymbol{\phi}, x) - \exp(iwx\sigma_z)\|_F \leq \varepsilon, \quad (11)$$

where $\|\cdot\|_F$ is the Frobenius norm.

The proof can be found in Appendix 4.

Corollary 1. *In the assumptions of Theorem 5, the gate $R(\boldsymbol{\theta}, \boldsymbol{\phi}, x)$ is realizable in depth*

$$N \in \tilde{\mathcal{O}}\left(w + \varepsilon^{-2}\right). \quad (12)$$

The proof can be found in Appendix 5.

Subsequent applications of approximations as given by Theorem 5 allow us to find the parameters $\boldsymbol{\theta}', \boldsymbol{\phi}'$ such that, for given $W, \boldsymbol{\theta}, \boldsymbol{\phi}$

$$U^{\mathrm{mf}}_{L'}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \boldsymbol{x}) \approx \left(\prod_{j=1}^{L} e^{i\theta'_j\sigma_z} e^{i\phi'_j\sigma_y} e^{i\sum_{k=1}^{m} w_{j,k}x_k\sigma_z}\right) e^{i\theta'_0\sigma_z} e^{i\phi'_0\sigma_y} e^{i\lambda'} = U^{\mathrm{w}}_{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \boldsymbol{x}). \quad (13)$$

For this next step, we will leverage Theorem 5 to approximate with arbitrary accuracy the output of a quantum circuit as in Equation (7). In a nutshell, we take our previous result to approximate a gate of the form $e^{iw_k x_k\sigma_z}$. Then, we can subsequently apply these gates to approximately obtain $\prod_k e^{ix_k w_k \sigma_z} = \exp(i\sum_k x_k w_k \sigma_z)$, which are tunable weights. This step bridges the gap with models relying on tunable weights.

Lemma 1. *Consider the re-uploading models $U^{\mathrm{w}}_{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \boldsymbol{x})$ from Equation (7) and $U^{\mathrm{mf}}_{L'}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \boldsymbol{x})$ from Equation (8). Then for any $\varepsilon > 0$, there exists a value $L' \geq L$ such that*

$$\forall (\boldsymbol{\theta}, \boldsymbol{\phi}) \quad \exists (\boldsymbol{\theta}', \boldsymbol{\phi}') \text{ s.t.} \left\| \mathrm{Tr}\left(\left(U^{\mathrm{w}}_{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \boldsymbol{x}) - U^{\mathrm{mf}}_{L'}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \boldsymbol{x}) \right) |0\rangle\langle 0| \right) \right\|_{\infty} \leq \varepsilon \quad (14)$$

The proof is available in Appendix 6. We refer the reader to Figure 3 for a graphical intuition of this proof.

For the final step, we just need to consider Lemma 1 and the triangular inequality. A model with tunable weights can approximate any function (with certain technical requirements) with arbitrary accuracy, and a fixed-weight model can approximate a model with tunable weights as well. Then, single-qubit fixed-weight models are universal approximators, at the expense of an overhead in depth.

Theorem 6. *Consider the model $U^{\mathrm{mf}}_{L'}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \boldsymbol{x})$ from Equation (8). The set of output functions $\mathcal{H} = \{h^{\mathrm{f}}_{L'}\}$, with*

$$h^{\mathrm{f}}_{L'}(\boldsymbol{x}) = \langle 0| U^{\mathrm{mf}}_{L'}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \boldsymbol{x}) |0\rangle \quad (15)$$

is universal with respect to multivariate continuous functions $f : [0,1]^m \to \mathbb{C}$ with the constraint $|f(\boldsymbol{x})|^2 \leq 1$, in the norm $\|\cdot\|_\infty$.

Proof. This corollary is an immediate consequence of Lemma 1. The considered model can approximate a multivariate function $h_L(\boldsymbol{x}) = \langle 0| U^{\mathrm{w}}_{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \boldsymbol{x}) |0\rangle$ output by a model with tunable weights. The triangular inequality implies that

$$\left\| h^{\mathrm{f}}_{L'}(\boldsymbol{x}) - f(\boldsymbol{x}) \right\| \leq \left\| h_L(\boldsymbol{x}) - h^{\mathrm{f}}_{L'}(\boldsymbol{x}) \right\| + \left\| h_L(\boldsymbol{x}) - f(\boldsymbol{x}) \right\|. \quad (16)$$

Each of the summands can be made arbitrarily small, thus guaranteeing universality. □

e. Implications for approximating functions beyond those generated by quantum circuits — We extend the obtained result to a more general theorem of functional analysis

Corollary 2. *Let $\mathcal{G}_{L,N} = \{g_{L,N}(x)\}$ be the set of functions $g_N : [0,1]^m \to \mathbb{C}$ of the form*

$$g_{L,N}(\boldsymbol{x}) = \sum_{j=1}^{L} \gamma_j \prod_{k=1}^{m} \sum_{n=-N}^{N} c_{j,k,n} e^{i\pi n x_k}, \quad (17)$$

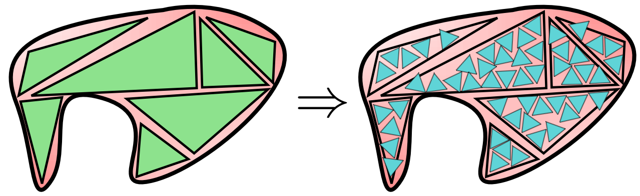

FIG. 3. Graphical description of the proof for universality. Consider the curved blob to be a representation of a set of functions. Triangles are components specified by different gates in the circuit. The re-uploading model with tunable parameters (left) is represented by triangles with tunable sizes and shapes, while fixed encoding gates (right) are represented by fixed-shaped triangles, but tunable in size. We can approximate the large tunable triangles with small fixed-shaped triangles. In the limit, any curved blob can be approximated by many small triangles.

for $\boldsymbol{x} \in [0,1]^m, \boldsymbol{w}_n \in \mathbb{R}^m, \alpha_n, \phi_n \in \mathbb{R}$. *There exists a* N *such that* $\mathcal{G}_N$ *is universal with respect to the set of all continuous functions* $f : [0,1]^m \to \mathbb{C}$, *in the norm* $\|\cdot\|_\infty$.

The proof can be found in Appendix 7.

f. Arbitrary-dimensional systems — The final result of this manuscript is a generalization of the univariate encoding $e^{iwx\sigma_z}$ to a specific and more restricted multi-variate case. The motivation to study these cases stems from quantum signal processing which computes functions of N-dimensional matrices rather than functions of the entries of the matrix. Specifically, we consider the case where the m-dimensional input is embedded into a matrix as a direct sum of the form

$$V_m(\boldsymbol{x}) = \sum_{k=1}^{m} |k\rangle\langle k| \otimes e^{i\pi x_k \sigma_z/2}, \qquad (18)$$

instead of a tensor product of m 2x2 matrices as is done in [14]. In Appendix 10, we consider an even more restricted additional case when the encoding matrix contains each feature of the input as a diagonal element.

Our quantum circuit only includes $V_m(x)$ and not any controlled version of it. Thus, we fix the relative phase with extra dimensions, ensuring the unitary is an element of in $\mathcal{SU}(2m)$. However, this choice does not hinder the generality of our arguments. Any change in the relative phase can be given by an affine transformation. If the obtained model is universal, this implies it is capable of applying arbitrary affine transformation. We choose this construction for its convenience for the proof.

We define a data re-uploading model with fixed encoding gates and arbitrary dimensionality as

Definition 3 (Re-uploading: fixed encoding gates and arbitrary dimensions)**.** *Consider the circuit*

$$U(\Theta, \Phi, \boldsymbol{\lambda}; \boldsymbol{x}) = \left(\prod_{j=1}^{L} R(\boldsymbol{\theta}_j, \boldsymbol{\phi}_j) V_m(\boldsymbol{x}) \right) R(\boldsymbol{\theta}_0, \boldsymbol{\phi}_0) V_m(\boldsymbol{\lambda}), \qquad (19)$$

where $R(\boldsymbol{\theta}, \boldsymbol{\phi})$ *are the specifications through Euler angles of arbitrary* $\mathcal{SU}(2m)$ *matrices.*

This circuit is realizable in $\log_2(\lceil 2m \rceil)$ qubits. If $\log(2m) \notin \mathbb{Z}$, the gate $V_m(\boldsymbol{x})$ can be padded with 1's for implementation in qubit-based systems. Following the rationale of Theorem 3, we identify now a building block operator that allows for universal processing, namely $R(\boldsymbol{\theta}, \boldsymbol{\phi}) V_m(\boldsymbol{x}) \in \mathcal{SU}(2m)$. From the analogy with Equation (6), we note that $V_m(\boldsymbol{x})$ plays the role of λ, which are specific angles in the Euler decomposition of $\mathcal{SU}(2)$. The extension to Euler decomposition of arbitrary dimensions [21] shows that $R(\boldsymbol{\theta}, \boldsymbol{\phi})$ is fully specified by any unitary matrix in $\mathcal{SU}(2m)$, excluding the relative phases among columns. We refer the reader to Appendix 8 for a detailed description.

Theorem 7. *Let* $U(\Theta, \Phi, \boldsymbol{\lambda}; \boldsymbol{x})$ *be the gate defined in Definition 3. The family of output functions* $\mathcal{H} = \{h_L^{(i)}\}$, *with*

$$h_L(\boldsymbol{x}) = \langle 0 | U(\Theta, \Phi, \boldsymbol{\lambda}; \boldsymbol{x}) | 0 \rangle \qquad (20)$$

is universal with respect to the norm $\|\cdot\|_\infty$ *for all continuous complex functions* $f(\boldsymbol{x}), \|f(\boldsymbol{x})\|_\infty \leq 1$, *for* $\boldsymbol{x} \in [0,1]^m$. *This model has a constant overhead in depth as compared to the single-qubit case.*

The proof can be found in Appendix 9. The sketch of the proof is as follows. There exists a choice of Θ, Φ such that the concatenation of layers of the form $R(\boldsymbol{\theta}_j, \boldsymbol{\phi}_j) V_m(\boldsymbol{x})$ provides a matrix in a block form, such that the first block is $e^{i\pi x_k \sigma_z}$, for any k, and the rest is the identity matrix. This transforms a multidimensional problem into a concatenation similar to Definition 2. Application of results in Sec. d provides universality.

g. Conclusions — In this paper, we show that it is possible to encode any continuous function of arbitrary dimensions using single-qubit quantum circuits with fixed encoding gates. This finding extends existing results of universality for univariate functions with fixed generators on multi-qubit, and multivariate functions with tunable generators in single-qubit circuits. The question of whether a trade-off between generators and the number of qubits is needed is now closed. In addition, we show that universality is also attainable if data is introduced in the form of a diagonal unitary matrix in $\Theta(\log m)$ qubits, where m is the dimension of the function to approximate. The results here depicted can be interpreted as analogous to width-depth trade-offs in boolean classical and quantum computing.

This paper opens new avenues to use quantum circuits for representing multidimensional functions, with interest for instance to generalize signal processing [22], or designing broader quantum machine learning algorithms.

ACKNOWLEDGMENTS

The authors would like to thank members of aQa Leiden for fruitful discussions. This work was supported by the Dutch National Growth Fund (NGF), as part of the Quantum Delta NL programme. This publication is part of the project Divide & Quantum (with project number 1389.20.241) of the research programme NWA-ORC which is (partly) financed by the Dutch Research Council (NWO) This work was also partially supported by the Dutch Research Council (NWO/OCW), as part of the Quantum Software Consortium programme (project number 024.003.03), and co-funded by the European Union (ERC CoG, BeMAIQuantum, 101124342). AB is supported by CERN through the CERN Quantum Technology Initiative. AB is supported by the quantum computing for earth observation (QC4EO) initiative of ESA ϕ-lab, partially funded under contract 4000135723/21/I-DT-lr, in the FutureEO program.

[1] P. W. Shor, SIAM Journal on Computing **26**, 1484 (1997).
[2] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nature Communications **5**, 4213 (2014).
[3] E. Farhi, J. Goldstone, and S. Gutmann, "A Quantum Approximate Optimization Algorithm," (2014), arXiv:1411.4028 [quant-ph].
[4] A. Cervera-Lierta, J. S. Kottmann, and A. Aspuru-Guzik, "The Meta-Variational Quantum Eigensolver (Meta-VQE): Learning energy profiles of parameterized Hamiltonians for quantum simulation," (2021), arXiv:2009.13545.
[5] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019).
[6] M. Schuld and N. Killoran, Physical Review Letters **122**, 040504 (2019), arXiv:1803.07128.
[7] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Physical Review A **98**, 032309 (2018), arXiv:1803.00745.
[8] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Quantum **4**, 226 (2020).
[9] H. M. Sheffer, Transactions of the American Mathematical Society **14**, 481 (1913).
[10] D. A. Barrington, Journal of Computer and System Sciences **38**, 150 (1989).
[11] A. Cosentino, R. Kothari, and A. Paetznick, in *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013) pp. 80–92.
[12] G. Cybenko, Mathematics of Control, Signals and Systems **2**, 303 (1989).
[13] K. Hornik, Neural Networks **4**, 251 (1991).
[14] M. Schuld, R. Sweke, and J. J. Meyer, Physical Review A **103**, 032430 (2021), arXiv:2008.08605.
[15] B. Casas and A. Cervera-Lierta, Physical Review A **107**, 062612 (2023), arXiv:2302.03389 [quant-ph].
[16] A. Pérez-Salinas, D. López-Núñez, A. García-Sáez, P. Forn-Díaz, and J. I. Latorre, Physical Review A **104**, 012405 (2021).
[17] P. G. L. Dirichlet, arXiv:0806.1294 [math] (2008), arXiv:0806.1294 [math].
[18] T. d. L. Silva, L. Borges, and L. Aolita, "Fourier-based quantum signal processing," (2022), arXiv:2206.02826 [quant-ph].
[19] D. Motlagh and N. Wiebe, "Generalized Quantum Signal Processing," (2024), arXiv:2308.01501 [quant-ph].
[20] A. Manzano, D. Dechant, J. Tura, and V. Dunjko, "Parametrized Quantum Circuits and their approximation capacities in the context of quantum machine learning," (2023).
[21] T. Tilma and E. C. G. Sudarshan, Journal of Physics A: Mathematical and General **35**, 10467 (2002), arXiv:math-ph/0205016.
[22] G. H. Low and I. L. Chuang, Physical Review Letters **118**, 010501 (2017), arXiv:1606.02685.
[23] P. Turán, in *Leopold Fejér Gesammelte Arbeiten I*, edited by P. Turán (Birkhäuser, 37–43).

1. Extension of Theorem 1

Theorem 8 (Fourier theorem). *Let $\mathcal{G}_N^F = \{g_N^F(x)\}$ be the set of functions $g_N^F : [0,1]^m \to \mathbb{C}$ of the form*

$$g_N^F(\boldsymbol{x}) = \sum_{\boldsymbol{n}} c_{\boldsymbol{n}} \exp(i2\pi \boldsymbol{n} \cdot \boldsymbol{x}) \tag{21}$$

for $\boldsymbol{n} = (n_1, \ldots, n_m) \in \mathbb{N}^m$, $\|\boldsymbol{n}\|_\infty \leq N$ and $c_{\boldsymbol{n}} \in \mathbb{C}$. For every square-integrable function $f : [0,1]^m \to \mathbb{C}, f \in \mathcal{F}$ there exists a value N such that $\mathcal{G}_N^F$ is universal with respect to $\mathcal{F}$ in the norm $\|\cdot\|_2$. The optimal coefficients are

$$c_{\boldsymbol{n}} = \int_{[0,1]^m} d\boldsymbol{x} f(\boldsymbol{x}) e^{-i2\pi \boldsymbol{n}\cdot\boldsymbol{x}} \tag{22}$$

The interpretation of this theorem is direct in the language of Hilbert spaces. We consider functions as elements of an infinite-dimensional Hilbert space. In this case, the basis of the space is the set $\{e^{i2\pi \boldsymbol{n}\cdot\boldsymbol{x}}\}$, with n a vector of integers. The Hilbert space is equipped with the inner product

$$\langle f(\boldsymbol{x}), g(\boldsymbol{x})\rangle = \int_{[0,1]^m} f^*(\boldsymbol{x}) g(\boldsymbol{x}) d\boldsymbol{x}, \tag{23}$$

and in particular

$$\langle e^{i2\pi \boldsymbol{n}\cdot\boldsymbol{x}}, e^{i2\pi \boldsymbol{l}\cdot\boldsymbol{x}}\rangle = \prod_{k=1}^{m} \delta_{n_k l_k}, \tag{24}$$

thus yielding an orthonormal basis. The Fourier theorem is nothing but the transformation of functions from arbitrary forms to this Hilbert space picture.

2. Extension of Theorem 2

Theorem 9 (Universal approximation theorem, adapted from [12]). *Let $\mathcal{G}_N^{\rm UAT} = \{g_N^{\rm UAT}(x)\}$ be the set of functions $g_N^{\rm UAT} : [0,1]^m \to \mathbb{C}$ of the form*

$$g_N^{\rm UAT}(x) = \sum_{n=1}^{N} \alpha_n \sigma\left(\boldsymbol{w}_n \cdot \boldsymbol{x} + \phi_n\right), \tag{25}$$

for $\boldsymbol{x} \in [0,1]^m, \boldsymbol{w}_n \in \mathbb{R}^m, \alpha_n, \phi_n \in \mathbb{R}$, and $\sigma(\cdot)$ a discriminatory function. For every continuous function $f : [0,1]^m \to \mathbb{C}, f \in \mathcal{F}$ there exists a N such that $\mathcal{G}_N^{\rm UAT}$ is universal with respect to $\mathcal{F}$ in the norm $\|\cdot\|_\infty$. Discriminatory functions are those satisfying

$$\int_{[0,1]^m} d\mu(\boldsymbol{x}) \sigma(\boldsymbol{w}\cdot\boldsymbol{x} + \phi) = 0,\ \boldsymbol{w} \in \mathbb{R}^m, \phi \in \mathbb{R} \Leftrightarrow \mu = 0, \tag{26}$$

for μ being Borel measures. In particular, $e^{i\ \cdot}$ is a discriminatory function.

The interpretation of this theorem can be done in terms of the Hilbert space discussed in Theorem 8. For arbitrary $\boldsymbol{w}, \phi$, discriminatory functions satisfy that they have non-zero overlaps with all elements in the basis of functions. For exponential functions

$$e^{iwx} = \sum_{n=-\infty}^{\infty} c_n(w) e^{in2\pi x}, \tag{27}$$

$$c_n(w) = \frac{-i}{(2\pi n - w)} \left(e^{iw)} - 1\right). \tag{28}$$

This property, together with the Hahn-Banach theorem, guarantees dense covering of the space of continuous functions, hence universality. We refer the reader to the original statement of the UAT in Ref. [12] for an in-depth proof.

3. Proof of Theorem 3

Proof. We recall theorem 3 in [19], formulated as follows.

Theorem 10 (Generalized quantum signal processing)**.** *Consider the building block*

$$R(\theta, \phi, \lambda) = \begin{pmatrix} e^{i(\phi+\lambda)}\cos(\theta) & e^{i\phi}\sin(\theta) \\ e^{i\lambda}\sin(\theta) & -\cos(\theta) \end{pmatrix}. \tag{29}$$

Then, $\forall L \in \mathbb{N}$, $\exists \boldsymbol{\theta}, \boldsymbol{\phi} \in \mathbb{R}^{L+1}, \lambda \in \mathbb{R}$ *such that:*

$$\left(\prod_{j=1}^{L} R(\theta_j, \phi_j, x) \right) R(\theta_0, \phi_0, \lambda) = \begin{pmatrix} P(e^{ix}) & \cdot \\ Q(e^{ix}) & \cdot, \end{pmatrix} \tag{30}$$

with $P(e^{ix}), Q(e^{ix})$ *being polynomials of degree* L *subject to the constraint* $|P(e^{ix})|^2 + |Q(e^{ix})|^2 = 1$, $\forall x$*. The gate* R *is defined as*

In the same reference, corollary 5 demonstrates that $P(e^{ix})$ can be chosen arbitrarily. Additionally, theorem 6 in the same reference implies that it is possible to obtain arbitrary polynomials of the form

$$P'(e^{ix}) = e^{-ikx} P(e^{ix}) \tag{31}$$

if the data is transformed according to $x \to -x$ in the last k layers.

We focus in this last result. Notice that changing the sign of x is equivalent to performing an inversion, as

$$\begin{pmatrix} e^{-ix} & 0 \\ 0 & 1 \end{pmatrix} = e^{-ix} X \begin{pmatrix} e^{ix} & 0 \\ 0 & 1 \end{pmatrix} X. \tag{32}$$

We identify

$$\begin{pmatrix} e^{ix} & 0 \\ 0 & 1 \end{pmatrix} = e^{ix/2} \begin{pmatrix} e^{ix/2} & 0 \\ 0 & e^{-ix/2} \end{pmatrix} = e^{ix/2} e^{ix/2\sigma_z} \tag{33}$$

$$\begin{pmatrix} e^{-ix} & 0 \\ 0 & 1 \end{pmatrix} = e^{-ix/2} \begin{pmatrix} e^{-ix/2} & 0 \\ 0 & e^{ix/2}. \end{pmatrix} = e^{-ix/2} e^{-ix/2\sigma_z} \tag{34}$$

By querying L times the first operator and L times the second, consecutively, we can represent any polynomial by virtue of [19]. These gates are accessible with the same query, plus the unitary transformation that can be absorbed in the gates before and after. The global phases are compensated. Thus, following this recipe, we can find arbitrary polynomials of the form

$$P(e^{ix}) = \sum_{j=-L}^{L} c_j e^{i\ jx}. \tag{35}$$

These polynomials are Fourier series up to degree L, and thus this family of functions is universal as stated in Theorem 1.

Finally, we bridge the gap between this result and our statement in Theorem 3. Notice that in this case $x \in [0, 2\pi]$. To maintain consistency with our result, we impose $x \in [0, 1]$ at the expense of adding the factor 2π in the encoding. The building block in Equation (29) is decomposable as Equation (6) up to global phases. □

4. Proof of Theorem 5

Proof. We consider in this proof only approximations to $w \in [-\pi/2, \pi/2]$. Without loss of generality, this implies arbitrary approximations for $w \in \mathbb{R}$, since the integer approximation to w is attainable by just repeating the encoding layer. Notice that, by definition, $f(x) = e^{iwx}$ satisfies $|f(x)|^2 = 1$. By Theorem 3 we know that it is possible to construct arbitrary polynomials $P(e^{ix})$ of degree at most N into a unitary matrix as

$$R_N(\boldsymbol{\theta}, \boldsymbol{\phi}, x) = \begin{pmatrix} P_N(x) & -Q_N^*(x) \\ Q_N(x) & P_N^*(x) \end{pmatrix} \approx \begin{pmatrix} e^{iwx} & 0 \\ 0 & e^{-iwx} \end{pmatrix}. \tag{36}$$

In the language of Theorem 3, $P(x)$ is a Fourier-like polynomial of e^{iwx} up to degree N. In particular, we choose the polynomial to be a Cesàro mean of the function e^{iwx}.

Definition 4 (Cesàro means). *Let $f:\mathbb{R}\to\mathbb{C}$ be a continuous function with period 2π, and let $g_n^F(x)$ be its n-term Fourier series. The Cesàro mean of order N is given by*

$$P_N(x) = \frac{1}{N+1}\sum_{n=-N}^{N} g_n^F(x), \tag{37}$$

or equivalently

$$P_N(x) = \sum_{n=-N}^{N} \frac{N+1-|n|}{N+1} c_n e^{in\pi x}. \tag{38}$$

This choice of $P_N(x)$ implies several convenient properties. First, $P_N(x)$ can be understood as the convolution of the function e^{iwx} with the Féjer Kernel [23]

$$K_N(x) = \frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=-n}^{n} e^{ikx}. \tag{39}$$

which satisfies

$$\int_x dx K_N(x) = 1. \tag{40}$$

Thus

$$\|P_N\|_\infty = \sup_{x\in[0,1]} \left| \int dt e^{iw(x-t)} K_N(t) dt \right| \leq \sup_{x\in[0,1]} |e^{iwx}| \left| \int dt K_N(t) dt \right| = 1. \tag{41}$$

Therefore, we can implement this function within a unitary operation. Note that in this and subsequent proofs, the ∞-norm for functions is made explicit as $\sup_{x\in[0,1]}$ to avoid confusion with matrix norms.

Second, we can recall Fejér's theorem [23].

Theorem 11 (Fejér's theorem). *Let $f:\mathbb{R}\to\mathbb{C}$ be a continuous function with period 2π, and let $P_N(x)$ be its N-term Cesàro mean. Then P_N converges uniformly to f as N increases, that is for every $\varepsilon > 0$ there exists N satisfying*

$$\sup_{x\in[0,1]} |f(x) - P_N(x)|_\infty \leq \varepsilon. \tag{42}$$

Choosing $P_N(x)$ as the corresponding Cesàro mean, we choose $\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}$ such that $R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x)$ implements $P_N(x)$ and $Q_N(x)$ as in Equation (36). With this choice, we can now write the Frobenius norm of the difference matrix as

$$\left\| R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\sigma_z} \right\|_F = \sqrt{2}\sqrt{|P_N(x) - e^{iwx}|^2 + |Q_N(x)|^2}. \tag{43}$$

The Frobenius norm can be bounded as follows. We use first the triangular inequality and convexity of functions to reach

$$|e^{iwx} - P_N(x)| \geq 1 - |P_N(x)| \geq \frac{1 - |P_N(x)|^2}{2} = \frac{|Q_N(x)|^2}{2}, \tag{44}$$

hence

$$\left\| R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\sigma_z} \right\|_F \leq \sqrt{2}\sqrt{|e^{iwx} - P_N(x)|^2 + 2|e^{iwx} - P_N(x)|}. \tag{45}$$

By virtue of Fejér's theorem, we can make the supremum norm of this function arbitrarily small, as long as e^{iwx} satisfies the continuity assumption. We enforce this constraint by finding the Cesàro means of an auxiliary related function, following the techniques in [20]. We define the period of this auxiliary function as

$$a(x) = \begin{cases} e^{iwx} & 0 \leq x \leq 1 \\ e^{iw(2-x)} & 1 < x < 2 \end{cases}, \tag{46}$$

and compute the Cesàro means of this function. This function is equivalent to e^{iwx} in the domain $x \in [0,1]$, and it also fulfils all the requirements of Fejér's theorem. We just need to ensure that our model is capable of expressing polynomials of $e^{in\pi x}$. This follows immediately from the definition of $R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x)$, which has $e^{i\pi x\sigma_z}$ in it. Thus, we can find a gate $R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x)$ satisfying that for all $\varepsilon > 0$ there exists a N such that

$$\sup_{x\in[0,1]} \left\| R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\sigma_z} \right\|_F \leq \varepsilon, \tag{47}$$

and concludes the proof. ☐

5. Proof of Corollary 1

We begin with the function defined in Equation (46). This function is continuous and periodic. Its Fourier series is given by the sets of coefficients

$$c_n = \int_0^1 e^{i(w-n)x} dx + e^{i2w} \int_1^2 e^{-i(w+n)x} = -i \frac{w\left((-1)^n e^{iw} - 1\right)}{w^2 - (n\pi)^2}. \tag{48}$$

We bound now $\sup_{x\in[0,1]} \left|P_N(x) - e^{iwx}\right|$. The definition of $P_N(x)$ from Appendix 4 implies

$$P_N(x) - e^{iwx} = \sum_{n=-N}^{N} c_n \frac{-|n|}{N+1} e^{in\pi x} - \sum_{|n|=N+1}^{\infty} c_n e^{in\pi x}, \tag{49}$$

hence by the triangular inequality

$$\sup_{x\in[0,1]} |P_N(x) - e^{iwx}| \leq \frac{2}{N+1} \left|\sum_{n=1}^{N} n c_n\right| + 2 \left|\sum_{n=N+1}^{\infty} c_n\right|. \tag{50}$$

For this proof, we consider that $|w| \leq \pi/2$. The reason is that we can obtain $K\pi \approx w$ exactly with K gates, and from this point on it is only needed to approximate the remainder $|w| \leq \pi/2$ using Fejér's theorem. We focus on each term individually. The first term is a 1-norm of the vector defined by nc_n. Thus,

$$\left|\sum_{n=1}^{N} n c_n\right| \leq \sum_{n=1}^{N} n\,|c_n| \leq \frac{2}{\pi}\left(\sum_{n=1}^{N} \frac{n}{n^2 - \frac{1}{4}}\right) \leq \frac{2}{\pi}\left(\frac{4}{3} + \int_1^N dx \frac{x}{x^2 - \frac{1}{4}}\right) = \frac{2}{\pi}\left(\frac{4}{3} + \frac{1}{2}\log\left(\frac{4N^2-1}{3}\right)\right). \tag{51}$$

For the second term,

$$\left|\sum_{n=N+1}^{\infty} c_n\right| \leq \sum_{n=N+1}^{\infty} |c_n| \leq \frac{2}{\pi}\int_N^{\infty} \frac{1}{n^2 - 1/4} \leq \frac{2}{\pi}\int_N^{\infty} \frac{1}{(n-1/4)^2} = \frac{2}{\pi}\frac{1}{N - 1/4}. \tag{52}$$

We see that the first term dominates over the second term. Now, we only need to re-arrange terms from Equation (50) to show

$$\sup_{x\in[0,1]} |P_N(x) - e^{iwx}| \in \tilde{\mathcal{O}}\left(N^{-1}\right). \tag{53}$$

We recover now the Frobenius norm from Theorem 5

$$\left\|R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\sigma_z}\right\|_F = \sqrt{2}\sqrt{|P_N(x) - e^{iwx}|^2 + |Q_N(x)|^2}. \tag{54}$$

By virtue of Fejér's theorem, the absolute value can be made arbitrarily small. Hence, in particular $|P_N(x) - e^{iwx}|^2 \leq |P_N(x) - e^{iwx}| \leq 1$. It is immediate to see that

$$\left\|R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\sigma_z}\right\|_F \leq \sqrt{6|P_N(x) - e^{iwx}|}, \tag{55}$$

and subsequently

$$\sup_{x\in[0,1]} \left\|R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\sigma_z}\right\|_F \in \tilde{\mathcal{O}}\left(N^{-1/2}\right). \tag{56}$$

Therefore, we can approximate the gate $e^{iwx\sigma_z}$ with two steps. First, we apply $e^{i\pi x\sigma_z}$ a number of times N' to reach $|\omega - N'\pi \leq \pi/2$. Second, we use Fejér's theorem to approximate $e^{iwx\sigma_z}$, for $|w| \leq \pi/2$, with the errors specified in this proof. Hence, in order to approximate $e^{iwx\sigma_z}$, to ε precision in the Frobenius norm, we need to apply

$$N \in \tilde{\Omega}(w + \varepsilon^{-2}) \tag{57}$$

gates. This concludes the proof. □

6. Proof of Lemma 1

Proof. We aim to bound the difference between the output functions of two unitaries built with and without tunable weights, explicitly $U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})$ and $U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})$. First, Hölder's inequality allows us to write

$$\min_{\boldsymbol{\theta},\boldsymbol{\phi}}\sup_{\boldsymbol{x}}\left|\mathrm{Tr}\left(\left(U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})-U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})\right)|0\rangle\langle 0|\right)\right|\leq\min_{\boldsymbol{\theta},\boldsymbol{\phi}}\sup_{\boldsymbol{x}}\left\|U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})-U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})\right\|_\infty, \tag{58}$$

where the matrix norm is the ∞-Schatten norm. Both $U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})$ and $U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})$ are constructed through layers. Following the rational of Equation (10), we group the gates in $U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})$ to approximate $e^{iwf\sigma_z}$ in $U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})$. Then this is done, we can match the parameters in $(\boldsymbol{\theta}',\boldsymbol{\theta}),(\boldsymbol{\phi}',\boldsymbol{\phi})$ to apply the same operations between encoding gates as in $U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})$. This fixes the parameters $\boldsymbol{\theta}',\boldsymbol{\phi}'$, thus giving an upper bound to the previous equation.

We can now use the triangular inequality. Consider two unitaries given by uU, vV, where u,v,U and V are unitaries as well. Then

$$\|uU-vV\|=\|uU-vU+vU-vV\|\leq\|u-v\|+\|U-V\|. \tag{59}$$

This procedure is repeated telescopically to the matrix $U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})-U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})$. We tune the parameters in $U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})$ to exactly match those of $U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})$ in the gates between encoding gates, thus not contributing to the difference. For the steps involving $e^{iw_{j,k}x_k\sigma_z}$, we approximate it with gates of the form $R_{j,k}(\boldsymbol{\theta}_{j,k},\boldsymbol{\phi}_{j,k},x)$. Thus,

$$\left\|U_L^{\rm w}(\boldsymbol{\theta},\boldsymbol{\phi},W,\lambda;\boldsymbol{x})-U_{L'}^{\rm mf}(\boldsymbol{\theta}',\boldsymbol{\phi}',\lambda';\boldsymbol{x})\right\|_\infty\leq\sum_{j=1}^{L'}\sum_{k=1}^{m}\sup_{x_k\in[0,1]}\left\|R_{j,k}(\boldsymbol{\theta}_{j,k},\boldsymbol{\phi}_{j,k},x)-e^{iw_{j,k}x_k\sigma_z}\right\|_\infty. \tag{60}$$

By virtue of Theorem 5, the Frobenius norm $\|R_{j,k}(\boldsymbol{\theta}_{j,k},\boldsymbol{\phi}_{j,k},x)-e^{iw_{j,k}x_k\sigma_z}\|_F$ can be made arbitrarily small. The Frobenius norm upper bounds the ∞-norm, thus each of the components in the sum can be made arbitrarily small. □

7. Proof of Corollary 2

Proof. To prove Corollary 2, we follow the same as for the quantum re-uploading circuits. We will approximate e^{iwx} with its Cesàro mean and merge all the functions together. The distance between a function $g_N(\boldsymbol{x})$ from Theorem 2 and its discrete-weights approximation is given by

$$\sup_{x\in[0,1]^m}\left|\sum_{j=1}^{L}\gamma_j\prod_{k=1}^{m}e^{iw_{j,k}x_k}-\sum_{j=1}^{L}\gamma_j\prod_{k=1}^{m}P_{N,\omega_{j,k}}(x_k)\right|\leq\sum_{j=1}^{L}|\gamma_j|\sup_{x\in[0,1]^m}\left|\prod_{k=1}^{m}e^{iw_{j,k}x_k}-\prod_{k=1}^{m}P_{N,\omega_{j,k}}(x_k)\right|, \tag{61}$$

with $P_{N,\omega_{j,k}}$ being the Cesàro means. Since

$$\sup_{x\in[0,1]^m}|f(\boldsymbol{x})|=\sup_{x\in[0,1]^m}\left|e^{i\boldsymbol{\alpha}\cdot\boldsymbol{x}}f(\boldsymbol{x})\right|, \tag{62}$$

we can perform an equivalent trick to the one in Equation (59) and find

$$\sup_{x\in[0,1]^m}\left|\sum_{j=1}^{L}\gamma_j\prod_{k=1}^{m}e^{iw_{j,k}x_k}-\sum_{j=1}^{L}\gamma_j\prod_{k=1}^{m}\omega_{j,k}(x_k)\right|\leq\sum_{j=1}^{L}|\gamma_j|\sum_{k=1}^{m}\sup_{x_k\in[0,1]}\left|e^{iw_{j,k}x_k}-P_{N,\omega_{j,k}}(x_k)\right|. \tag{63}$$

Since $P_{N,\omega_{j,k}}(x_k)$ is the Cesàro mean of $e^{iw_{j,k}x_k}$, up to degree N, we can recall Fejér's theorem to approximate the desired function with arbitrary precision. Hence, this construction provides universal functions in the assumptions of Fejér's theorem. □

8. Euler angles for $\mathcal{SU}(N)$

We follow the construction in [21] for Euler rotations of arbitrary dimensions. We define first the generators of the corresponding algebra $\mathfrak{su}(N)$ as generalized Gell-Mann matrices

$$\lambda_3 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{64}$$

$$\lambda_{(k-1)^2+1} = \begin{pmatrix} 0 & \cdots & -i & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ \underbrace{i \quad \cdots \quad 0}_{k} & & & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \tag{65}$$

$$\lambda_{N^2-1} = \sqrt{\frac{2}{N^2-N}} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix} \tag{66}$$

One can find an Euler decomposition of arbitrary dimension by employing a recursive construction in which matrices in $\mathcal{SU}(N)$ are defined as matrices in $\mathcal{SU}(N-1)$ plus extra parameters.

$$U = \prod_{2 \leq k \leq N} A(k)[SU(N-1)] e^{i\lambda_{N^2-1}\theta_{N^2-1}} \tag{67}$$

$$A(k) = e^{i\lambda_3\theta_{2k-3}} e^{i\lambda_{(k^2-1)+1}\theta_{2k-2}}. \tag{68}$$

Notice that this construction allocates all rotations constructed as exponentiations of diagonal matrices at the rightest part of the operations. Hence, we can justify our choice $V_m(\boldsymbol{x})$ from Equation (72).

9. Proof of Theorem 7

Proof. We begin by considering the data-encoding gate

$$V_m(\boldsymbol{x}) = \begin{pmatrix} e^{ix_1/2} & 0 & \cdots & 0 & \cdots \\ 0 & e^{-ix_1/2} & \cdots & 0 & \cdots \\ 0 & 0 & e^{ix_2/2} & 0 & \vdots \\ 0 & 0 & 0 & e^{-ix_2/2} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{69}$$

We design now a permutation Π_k that exchanges the positions $2j$ and $2j+1$, except for $j = k$. An example of this permutation is given by

$$\Pi_1 V_m(\boldsymbol{x}) = \begin{pmatrix} e^{ix_1/2} & 0 & \cdots & 0 & \cdots \\ 0 & e^{-ix_1/2} & \cdots & 0 & \cdots \\ 0 & 0 & e^{-ix_2/2} & 0 & \vdots \\ 0 & 0 & 0 & e^{ix_2/2} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{70}$$

These permutations are implementable with gates $R(\boldsymbol{\theta}, \boldsymbol{\phi})$, since R specifies any unitary gate, except for the relative phases among columns [21], and no phases are needed for the permutations. It is immediate to see

$$R_1(\boldsymbol{x}) = V_m(\boldsymbol{x})\Pi V_m(\boldsymbol{x}) = \begin{pmatrix} e^{ix_1} & 0 & \cdots & 0 & \cdots \\ 0 & e^{-ix_1} & \cdots & 0 & \cdots \\ 0 & 0 & 1 & 0 & \vdots \\ 0 & 0 & 0 & 1 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \begin{bmatrix} e^{ix_1} & 0 \\ 0 & e^{-ix_1} \end{bmatrix} & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} e^{i\pi x_1 \sigma_z} & 0 \\ 0 & I \end{pmatrix}. \tag{71}$$

The construction of $R_1(\boldsymbol{x})$ requires an overhead of $\mathcal{O}(1)$ encoding gates of the form $V_m(\boldsymbol{x})$ to achieve it. The same condition holds for any other $R_k(\boldsymbol{x})$. These blocks, up to permutations that can be reabsorbed in the parameterized gates, allow us to conduct the algorithms in Theorem 3 on the 2×2 upper-left corner. This can be extended to all coordinates. Notice that the relative phases cancel each other due to the matrix being applied the same number of times. Since universality is guaranteed in Theorem 3, this construction allows for universality as well in the multi-qubit case. □

10. An extension to Theorem 7

The proof above allows us to formulate an alternative theorem considering a different encoding gate, defined as

$$V'_m(\boldsymbol{x}) = \sum_{k=1}^{m} e^{i\pi x_k/m} |k\rangle\langle k| + e^{-i\sum_{i=1}^{m} \pi x_k/m} |m+1\rangle\langle m+1| = \begin{pmatrix} e^{ix_1/m} & 0 & 0 & \cdots & 0 \\ 0 & e^{ix_2/m} & 0 & \cdots & 0 \\ 0 & 0 & e^{ix_3/m} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{-i\frac{1}{m}\sum_{k=1}^{m} x_k} \end{pmatrix} \tag{72}$$

We define the auxiliary permutations $\Pi_{k,j}$, for $j < k, k \leq m$, being circular permutations on all elements except for k, of distance j. Again, these permutations are implementable with gates $R(\boldsymbol{\theta}, \boldsymbol{\phi})$, since R specifies any unitary gate, except for the relative phases among columns [21], and no phases are needed for the permutations. An example of this permutation is

$$\Pi_{1,1} V'_m(\boldsymbol{x}) = \begin{pmatrix} e^{ix_1/m} & 0 & 0 & \cdots & 0 \\ 0 & e^{-i\frac{1}{m}\sum_{k=1}^{m} x_k} & 0 & \cdots & 0 \\ 0 & 0 & e^{ix_2/m} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{ix_{m-1}/m} \end{pmatrix}. \tag{73}$$

We apply subsequent permutations until finding

$$V'_{\Pi_1}(\boldsymbol{x}) = \prod_{j=1}^{m-1} \Pi_{1,j}(V'_m(\boldsymbol{x})) = \begin{pmatrix} e^{ix_1\frac{m-1}{m}} & 0 & 0 & \cdots & 0 \\ 0 & e^{-ix_1/m} & 0 & \cdots & 0 \\ 0 & 0 & e^{-ix_1/m} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{-ix_1/m} \end{pmatrix} = e^{-ix_1/m} \begin{pmatrix} e^{ix_1} & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{74}$$

The next step is to apply the same permutation cycle over this permuted matrix to find

$$V'_{\Pi_1}(-\boldsymbol{x}) = \prod_{j=1}^{m} \Pi_{2,j}(V'_{\Pi_1}(\boldsymbol{x})) = e^{ix_1/m} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & e^{-ix_1} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{75}$$

We can identify

$$R_1(\boldsymbol{x}) = V'_{\Pi_1}(-\boldsymbol{x})V'_{\Pi_1}(\boldsymbol{x}) = \begin{pmatrix} \begin{bmatrix} e^{ix_1} & 0 \\ 0 & e^{-ix_1} \end{bmatrix} & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} e^{i\pi x_1 \sigma_z} & 0 \\ 0 & I \end{pmatrix}. \tag{76}$$

The construction of $R_1(\boldsymbol{x})$, or equivalent $R_k(\boldsymbol{x})$ requires an overhead of $\mathcal{O}(m^2)$ encoding gates $V'_m(\boldsymbol{x})$.

From this point, we can repeat the steps in Appendix 9 and formulate the following theorem.

Theorem 12. *Let $U(\Theta, \Phi, \boldsymbol{\lambda}; \boldsymbol{x})$ be the gate defined in Definition 3, with encoding gate $V'_m(\boldsymbol{x})$. The family of output functions $\mathcal{H} = \{h_L\}$, with*

$$h_L(\boldsymbol{x}) = \langle 0| \, U(\Theta, \Phi, \boldsymbol{\lambda}; \boldsymbol{x}) \, |0\rangle \tag{77}$$

is universal with respect to the norm $\|\cdot\|_\infty$ for all continuous complex functions $f(\boldsymbol{x}), \|f(\boldsymbol{x})\|_\infty \leq 1$, for $\boldsymbol{x} \in [0,1]^m$. This model has overhead $\mathcal{O}(m^2)$ in depth as compared to the single-qubit case.

The main difference of this theorem with respect to Theorem 7 is that this case requires one qubit less. However, the overhead in depth as compared to the single-qubit case is beneficial to Theorem 7.