

Universal approximation of continuous functions with minimal quantum circuits

Adrián Pérez-Salinas,^{1,2} Mahtab Yaghobi Rad,^{1,3} Alice Barthe,^{1,4} and Vedran Dunjko^{1,3}

¹ *$\langle aQa \rangle^L$ 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, the input becomes continuous and the output represents real-valued functions. While the universality of discrete quantum computers is well understood, basic questions remained open in the continuous case. We focus on universality on multivariate functions. Current approaches require either a number of qubits scaling linearly with the dimension of the input for fixed encodings, or a tunable encoding procedure in single-qubit circuits. The question of whether universality can be reached with a fixed encoding and sub-linearly many qubits remained open for the last five years. In this paper, we answer this question in the affirmative for arbitrary multivariate functions. We provide two methods: (i) a single-qubit circuit where each coordinate of the arguments to the function to represent is input independently, and (ii) a multi-qubit approach where all coordinates are input in one step, with number of qubits scaling logarithmically with the dimension of the argument of the function of interest. We view the first result of inherent and fundamental interest, whereas the second result opens the path towards representing functions whose arguments are densely encoded in a unitary operation, possibly encoding for instance quantum processes.

I. Introduction — Quantum computing extends the paradigm of classical circuits by adding new elements to the set of available logical operations, as allowed by the principles of quantum mechanics. The conventional paradigm of quantum computing is discrete and relies on a discrete set of gates to realize probabilistic bitstring-to-bitstring mappings. Some of these mappings are believed to be intractable for classical computers [1]. Motivated by near-term quantum computing and the limitations of their experimental realizations, recently there has been substantial interest in settings where quantum computers are used to realize *continuous-valued functions depending on continuous 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 of these cases, the question of the set of functions such a family can represent – i.e. the universality of such a function family – is central.

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[9]. In the classical case, Boolean universality can be achieved in constant-width settings [10, 11]. Quantum computing can surprisingly do better by representing arbitrary boolean functions over n bits using just a single-qubit wire [12]. For the continuous classical case, much is known about the representation power of many models such as neural networks [13, 14], or via e.g. the Kolmogorov-Arnold representation theorems [15], but for the quantum analogous fundamental open problems persisted.

Various methods to represent continuous functions with parametrized quantum circuits exist, mainly depending on how the arguments \mathbf{x} of the functions are encoded [8, 16–19]. We consider these encoding gates to be interleaved with parameter-dependent gates to construct a larger quantum circuit $U(\boldsymbol{\theta}, \mathbf{x})$ [20]. The type of encoding has direct impact in the number of parameters needed to guarantee universality in the functions that a quantum circuit can represent. For m -variate functions, fixed encoding gates require the number of tunable parameters to scale as $\mathcal{O}(\exp(m))$ and the number of qubits as $\mathcal{O}(m)$ [16, 17, 21]. In contrast, single-qubit circuits suffice to guarantee universality if the coordinates of \mathbf{x} admit pre-processing through arbitrary linear combinations before being input to the encoding gates [22]. The existence of universality guarantees in the scenario where \mathbf{x} is densely encoded, i. e. maximally using the size of the Hilbert space through a fixed $\mathcal{O}(\log m)$ -qubit encoding gate, remains an open question. Note that pre-processing is not available in this scenario due to the large number of coordinates in \mathbf{x} .

In this work, we address the minimal qubit requirements to represent multivariate arbitrary functions with fixed encoding gates, and obtain universality guarantees for $\mathcal{O}(\log m)$ -qubit fixed encoding gates, and provide a comprehensive comparison to previous results; see Table I. We consider two scenarios. If each coordinate of \mathbf{x} is input independently from others, single-qubit quantum circuits are universal for multivariate functions of arbitrary dimension. In the dense scenario where a m -dimensional \mathbf{x} is densely encoded in a fixed gate, universality holds for only $\mathcal{O}(\log m)$ -qubit circuits. These findings bridge the gap between the existing results and further reduce the qubit

	Reference	Pre-process	# qubits	# parameters	Functions	Norm
Foundational results	Fourier [23], Equation (3)	NO	-	$\mathcal{O}(L^m)$	L_2 -integrable	$\ \cdot\ _2$
	UAT [13], Equation (4),	YES	-	$\mathcal{O}(Lm)$	Continuous	$\ \cdot\ _\infty$
Quantum Circuits	QSP (for $m = 1$) [18], Theorem 5 and Yu&Chen's PQC [16], Theorem 7	NO	$\mathcal{O}(m)$	$\mathcal{O}(L^m)$	Continuous	$\ \cdot\ _\infty$
	Data re-uploading [22], Theorem 8	YES	1	$\mathcal{O}(Lm)$	Continuous	$\ \cdot\ _\infty$
	This work, Theorem 1	NO	1	$\mathcal{O}(Lm)$	Continuous	$\ \cdot\ _\infty$
	This work, Theorem 2	NO	$\mathcal{O}(\log m)$	$\mathcal{O}(Lm)$	Continuous	$\ \cdot\ _\infty$

TABLE I. Summary of existing results for universality of different function families and quantum circuits. The results in this work improve state-of-the-art requirements by exponential scalings. Theorems AX are available in the supplemental material.

requirements from the previous $\mathcal{O}(m)$.

II. Background on universality — Parameterizing arbitrary functions allows us to approximate them through controllable function families. The expressivity of these function families quantifies the range of implementable maps. Maximal expressivity is commonly referred to as universality, guaranteeing that arbitrary functions can be approximated. The formal definition of universality is as follows.

Definition 1 (Universality). *Let $\mathcal{G}_L = \{g\}$ be a sequence of function families, $\mathcal{G}_L \subseteq \mathcal{G}_{L+1}$. Let $\mathcal{F} = \{f\}$ be a reference function family, and $g, f : \mathcal{X} \subset \mathbb{R}^m \rightarrow \mathbb{C}$. Let p define a function norm as*

$$\|f\|_p = \left(\int_{\mathbf{x} \in \mathcal{X}} |f(\mathbf{x})|^p d\mathbf{x} \right)^{1/p}. \quad (1)$$

The sequence \mathcal{G}_L is universal with respect to \mathcal{F} for the p -norm if

$$\forall f \in \mathcal{F}, \forall \varepsilon > 0, \exists L, g \in \mathcal{G}_L \text{ s.t. } \|f - g\|_p \leq \varepsilon. \quad (2)$$

Next, we cover some universality results. We remark that these results will be illustrative for our findings, but the list is by no means an exhaustive review. The first result is the known Fourier's theorem, relying on functions of the form

$$g_L^F(\mathbf{x}) = \sum_{\substack{\mathbf{n} \\ \|\mathbf{n}\|_1 \leq L}} c_{\mathbf{n}} e^{i2\pi \mathbf{n} \cdot \mathbf{x}}, \quad (3)$$

for $\mathbf{x} \in \mathbb{R}^m, \mathbf{n} \in \mathbb{Z}^m$. These functions are universal in the $\|\cdot\|_2$ norm for square-integrable periodic functions. This result makes apparent that universality for functions of the form of $g_L^F(\mathbf{x})$ demands $\mathcal{O}(L^m)$ degrees of freedom. Alternatively, the universal approximation theorem (UAT) guarantees universality for functions of the form

$$g_L^{\text{UAT}}(\mathbf{x}) = \sum_{n=1}^L \alpha_n \exp(i(\mathbf{w}_n \cdot \mathbf{x} + \phi_n)), \quad (4)$$

with $\mathbf{w}_n \in \mathbb{R}^m, \alpha_n, \phi_n \in \mathbb{R}$, for continuous functions in the $\|\cdot\|_\infty$ norm. This theorem requires only $\mathcal{O}(Lm)$ degrees of freedom. Detailed statements of both theorems are available in Theorem 3 and Theorem 4 respectively.

The differences between the universality properties of both function families are rooted in the presence of tunable parameters \mathbf{w}_n . In short, Equation (3) decomposes a function into a chosen basis of the space of functions. However, Equation (4) decomposes a function into terms with non-zero overlaps with all the (infinitely many) elements of any basis for the function space. As a consequence, the function space is densely covered. We refer the interested reader to [13] and the proof of UAT for an in-depth discussion. The constraints on $\|\cdot\|_{\{2,\infty\}}$ norms or continuous vs. square-integrable can be overcome by choosing different function families, for instance Fejér's trigonometric or Bernstein's polynomial approximations [24, 25]. Another component is the periodicity of the functions, which can be overcome by symmetrizing the function and appropriately adapting the domain [26].

III. Universality in parameterized quantum circuits — In this work, we consider quantum circuits to encode a given function. These circuits combine arguments \mathbf{x} and tunable unitary gates in a global operation $U_L(\boldsymbol{\theta}, \mathbf{x})$, where L is a parameter to scale the size and depth of the circuit. We define the hypothesis functions given by the quantum model as

$$h_L(\mathbf{x}) = \langle 0 | U_L(\boldsymbol{\theta}, \mathbf{x}) | 0 \rangle, \quad (5)$$

Notice that h_L represents a complex number which cannot be retrieved through measurements. The choice above will be used in this manuscript as a mathematical formulation. The reduction to a measurable quantity can be achieved, for instance, by taking the modulus of $h_L(\mathbf{x})$, among other choices. We call a family of quantum circuits universal if the sequence of function families $\mathcal{H}_L = \{h_L\}$ is universal, under the conditions of

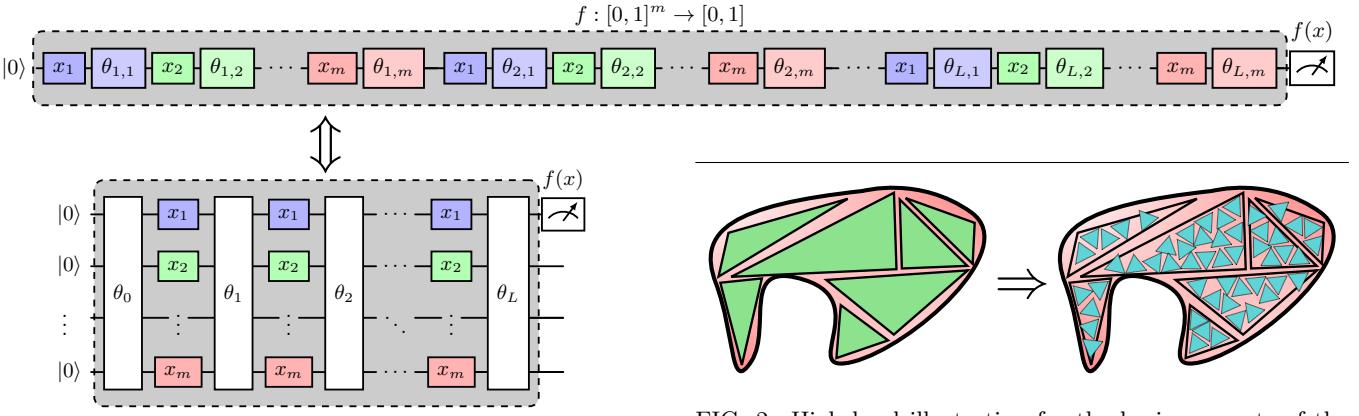


FIG. 1. Quantum circuits with parameters and fixed encoding gates can represent multivariate continuous functions with multiple qubits. In this paper, we show that single-qubit quantum circuits retain universal representation capabilities.

Definition 1.

We cover now relevant results on universality for quantum circuits with tunable parameters. The arguments \mathbf{x} are input to the quantum circuit through encoding gates, which in certain cases admit tunable parameters. Analogously to the discussion in Section II, the presence of tunable encodings allows for the distinction of two cases.

First, we focus on fixed encodings, that is, cases in which \mathbf{x} is introduced through queries to unitary operations which *only* depend on the data. For univariate functions ($m = 1$), it is possible to use a method based on QSP [27] to reach arbitrary polynomials up to degree L with single-qubit circuits and $\mathcal{O}(L)$ degrees of freedom, see Theorem 5 [18]. Depending on the polynomials to be encoded, this circuit provides universality with different conditions. For multivariate functions of arbitrary m , there exist constructions for arbitrary m -dimensional (Bernstein's) polynomials to degree L with $\mathcal{O}(m \log L)$ qubits and $\mathcal{O}(mL^m)$ parameters, see Theorem 7 [16]. This circuit family provides universality for continuous functions. If we drop the condition for fixed encoding gates, it is possible to achieve universality for multivariate functions with single-qubit circuits and $\mathcal{O}(Lm)$ parameters, for continuous functions in the $\|\cdot\|_\infty$, see Theorem 8 [22]. The different conditions for the considered scenarios are analogous to those discussed in Section II. The observations there given can be easily transported to results on quantum circuits.

To question the impact of quantum circuits, it is interesting to consider how the arguments \mathbf{x} are encoded. The only operation that allows such encoding is querying a \mathbf{x} -dependent unitary operation. In the context of quantum circuits, one can then argue that having fixed encoding gates implies that $\mathcal{O}(\exp(m))$ parameters are required to achieve universality. This is a direct consequence of the models considered for fixed encodings, where each coordinate of the arguments $x_i, i \in$

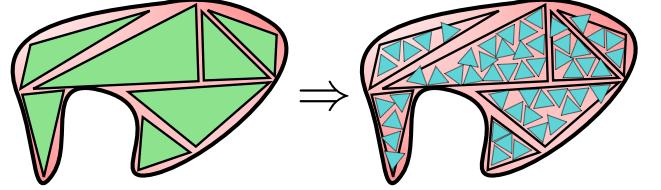


FIG. 2. High level illustration for the basic concepts of the proof for universality. Consider the curved blob to be a representation of a set of functions. Triangles are components specified by different gates in the circuit. The re-uploading model with tunable parameters (left) is represented by triangles 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.

$\{1, 2, \dots, m\}$ is independently input to the quantum circuits, for instance through unitaries composed of tensor products. On the other hand, tunable encodings seem to be required to achieve universality with $\mathcal{O}(m)$ many parameters. The trade-off between controllability of the encoding or number of degrees of freedom in principle hinders multidimensional analogues to QSP, that is, scenarios where m coordinates are simultaneously and densely encoded to a quantum circuit without ancillary qubits. Surprisingly, we show in the next section that there exists a construction to circumvent this trade-off.

IV. Results — Universality for multivariate functions with fixed encoding gates is possible. 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*

$$U_L^{\text{mf}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; \mathbf{x}) = \left(\prod_{j=1}^L \prod_{k=1}^m \prod_{n=1}^N e^{i\theta_{j,k,n}\hat{\sigma}_z} e^{i\phi_{j,k,n}\hat{\sigma}_y} e^{i\pi x_k \hat{\sigma}_z} \right) e^{i\theta_0 \hat{\sigma}_z} e^{i\phi_0 \hat{\sigma}_y} e^{i\lambda \hat{\sigma}_z}, \quad (6)$$

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

The first main result of this manuscript holds as follows.

Theorem 1. *Consider the model $U_L^{\text{mf}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; \mathbf{x})$ from Equation (6). The sequence of output function*

families $\mathcal{H} = \{h_L^f\}$, with

$$h_L(\mathbf{x}) = \langle 0 | U_L^{\text{mf}}(\boldsymbol{\theta}, \phi, \lambda; \mathbf{x}) | 0 \rangle \quad (7)$$

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

We give a sketch of the proof, to be detailed in Appendix 3 and references therein. As discussed in the previous section, the key element for universality with a small number of parameters is by enhancing the set of attainable functions, that is transforming elements of the Fourier series from integer to real indices. Our quantum circuit does not allow for that freedom, since we can only apply the encoding gates an integer number of times. However, we can show that it is possible to approximate $\exp(i\hat{\sigma}_z w x)$ to arbitrary precision with a sequence of the form given QSP [18]. This is achieved in Lemma 1, and it is possible because we will approximate only a univariate function. Next, we need to show that the aforementioned approximations can be stacked to obtain another approximation for a multivariate gate of the form $\exp(i\hat{\sigma}_z \mathbf{w} \cdot \mathbf{x})$. The last step is to apply a triangle inequality. If such circuit can approximate any continuous function, and this circuit can be approximated with fixed encoding gates, then the circuit in Definition 2 can approximate any function to arbitrary precision. See Figure 2 for a graphical interpretation of this proof.

Theorem 1 shows that it is possible to approximate arbitrary continuous multivariate functions with single-qubit circuits and fixed encoding gates of the form $\exp(i\hat{\sigma}_z \pi x_k)$. The circuits to achieve this require to query the m different encoding gates, (one per coordinate) on demand. This result implies a reduction in the requirements of number of qubits from linear in m to just one. This result is analogous to Barrington's theorem for classical circuits [11, 12].

For completeness, we can extend the obtained result to a more general theorem of functional analysis, independent from quantum circuits.

Corollary 1. Let $\mathcal{G}_{L,N} = \{g_{L,N}\}$ be the sequence of sets of functions $g_N : [0, 1]^m \rightarrow \mathbb{C}$ of the form

$$g_{L,N}(\mathbf{x}) = \sum_{j=1}^L \gamma_j \prod_{k=1}^m \sum_{n=-N}^N c_{j,k,n} e^{i\pi n x_k}, \quad (8)$$

for $\mathbf{x} \in [0, 1]^m$, $\mathbf{w}_n \in \mathbb{R}^m$, $\alpha_n, \phi_n \in \mathbb{R}$. Then, $\mathcal{G}_{L,N}$ is universal with respect to the set of all continuous functions $f : [0, 1]^m \rightarrow \mathbb{C}$, in the norm $\|\cdot\|_\infty$.

The proof can be found in Appendix 4.

For the second main result of this manuscript we consider a generalization of the univariate encoding $e^{iwx\hat{\sigma}_z}$ to a multivariate case. The motivation to study these cases stems from considering the arguments input to the system as the query of a unitary gate

carrying all relevant information. We consider diagonal encoding, which needs $\mathcal{O}(\log m)$ qubits to carry $\mathcal{O}(m)$ real numbers. Specifically, we consider the case where the m -dimensional input is given by a matrix of the form

$$V_m(\mathbf{x}) = \sum_{k=1}^m |k\rangle\langle k| \otimes e^{i\pi x_k \hat{\sigma}_z / 2}. \quad (9)$$

This scenario contrasts with other approaches where the arguments are carried by a tensor product of m matrices of size 2×2 [17]. In Appendix 7, we consider an even more restricted additional case when the encoding matrix contains each coordinate of the input as a diagonal element.

Definition 3 (Re-uploading: fixed encoding gates and arbitrary dimensions). A multi-qubit multivariate circuit with fixed encoding gates is given by

$$U_L(\Theta, \Phi, \boldsymbol{\lambda}; \mathbf{x}) = \left(\prod_{j=1}^L R(\boldsymbol{\theta}_j, \phi_j) V_m(\mathbf{x}) \right) R(\boldsymbol{\theta}_0, \phi_0) V_m(\boldsymbol{\lambda}), \quad (10)$$

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

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 $SU(2m)$. However, this choice does not hinder the generality of our arguments. Any change in the relative phase can be given by an affine transformation of \mathbf{x} . Following the rationale of QSP (Theorem 5), we note that $V_m(\mathbf{x})$ plays the role specific angles in the Euler decomposition of $SU(2m)$. The extension to Euler decomposition of arbitrary dimensions [28] shows that $R(\boldsymbol{\theta}, \phi)$ specifies any unitary matrix in $SU(2m)$, except for the relative phases among columns, fixed by \mathbf{x} . We refer the reader to Appendix 5 for a detailed description.

We can now state the final result.

Theorem 2. Let $U_L(\Theta, \Phi, \boldsymbol{\lambda}; \mathbf{x})$ be the gate defined in Definition 3, with $\mathcal{O}(\log m)$ many qubits. The sequence of output function families $\mathcal{H} = \{h_L\}$, with

$$h_L(\mathbf{x}) = \langle 0 | U_L(\Theta, \Phi, \boldsymbol{\lambda}; \mathbf{x}) | 0 \rangle \quad (11)$$

is universal for all continuous complex functions $f(\mathbf{x})$, $\|f(\mathbf{x})\|_\infty \leq 1$, for $\mathbf{x} \in [0, 1]^m$, with respect to the $\|\cdot\|_\infty$ norm.

The proof can be found in Appendix 6. This model has a constant overhead in depth as compared to the single-qubit case from Theorem 1. The sketch of the proof is as follows. There exists a choice of Θ, Φ such that the concatenation of layers of the form $R(\boldsymbol{\theta}_j, \phi_j)V_m(\mathbf{x})$ provides a matrix in a block form, such that the first block is $e^{i\pi x_k \hat{\sigma}_z}$, 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 from Theorem 1 provides universality. In Appendix 7, we consider an even more condensed additional case when the encoding matrix contains each coordinate of the input as a diagonal element.

The results here described open an avenue towards representing functions in the case where the arguments are accessible via query operations and exponentially many coordinates are simultaneously available, thus more explicitly utilizing the exponential size of the Hilbert space. An example of such operations on demand are time evolution dynamics, interleaved with tunable gates. Other existing methodologies encode argument coordinates individually, for example, by means of encoding gates composed of tensor products [16, 17], thus severely limiting for instance the applicability of learning processes to scenarios with classical data. In addition to this advantage, our construction requires a qubit scaling of $\mathcal{O}(\log m)$, improving previous linear scalings. Consequently, the requirement for the number of parameters required to specify these models is reduced from $\exp(\mathcal{O}(m))$ to just $\mathcal{O}(\text{poly}(m))$. One can interpret this as an efficient generalization of QSP in which the signal is densely encoded in the form of independent quantities inside a unitary gate of dimension 2^m . These properties are minimal requirements for the existence of quantum advantages in variational learning problems, to be further investigated.

V. 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 multivariate functions with fixed encodings on multi-qubit, and multivariate functions with tunable encodings in single-qubit circuits.

As a second result, we show that universality on

multivariate functions is also attainable if the arguments are input by a diagonal unitary matrix in $\Theta(\log m)$ qubits, where m is the dimension of the function to represent. This finding opens a path towards m -dimensional generalizations of QSP where the argument \mathbf{x} is given as the entries of a unitary matrices, for instance, time evolution, and opens the path for broader quantum machine learning algorithms.

The question of whether a trade-off between tunability of encoding and the number of qubits is needed is now closed. The results here depicted can be interpreted as analogous to width-depth trade-offs in boolean classical and quantum computing. Our results support universality by showing the existence of circuits with arbitrarily accurate approximations to functions. The performance of those models in practice, for instance related to trainability or (non-)dequantizability is a valuable future research direction.

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, Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer, 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, A variational eigenvalue solver on a photonic quantum processor, 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, Supervised learning with quantum-enhanced feature spaces, Nature **567**, 209 (2019).
 - [6] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, Physical Review Letters **122**, 040504 (2019), arXiv:1803.07128.
 - [7] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum Circuit Learning, Physical Review A **98**, 032309 (2018), arXiv:1803.00745.
 - [8] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum **4**, 226 (2020).
 - [9] 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.
 - [10] H. M. Sheffer, A set of five independent postulates for Boolean algebras, with application to logical constants, Transactions of the American Mathematical Society **14**,

- 481 (1913).
- [11] D. A. Barrington, Bounded-width polynomial-size branching programs recognize exactly those languages in *NC*1, *Journal of Computer and System Sciences* **38**, 150 (1989).
 - [12] A. Cosentino, R. Kothari, and A. Paetznick, Dequantizing Read-once Quantum Formulas, in *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013) pp. 80–92.
 - [13] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals and Systems* **2**, 303 (1989).
 - [14] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4**, 251 (1991).
 - [15] V. I. Arnold, On the representation of functions of several variables as a superposition of functions of a smaller number of variables, in *Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965*, edited by V. I. Arnold, A. B. Givental, B. A. Khesin, J. E. Marsden, A. N. Varchenko, V. A. Vassiliev, O. Y. Viro, and V. M. Zakalyukin (Springer, Berlin, Heidelberg, 2009) pp. 25–46.
 - [16] Z. Yu, Q. Chen, Y. Jiao, Y. Li, X. Lu, X. Wang, and J. Z. Yang, Non-asymptotic Approximation Error Bounds of Parameterized Quantum Circuits (2024), arXiv:2310.07528.
 - [17] M. Schuld, R. Sweke, and J. J. Meyer, The effect of data encoding on the expressive power of variational quantum machine learning models, *Physical Review A* **103**, 032430 (2021), arXiv:2008.08605.
 - [18] D. Motlagh and N. Wiebe, Generalized Quantum Signal Processing (2024), arXiv:2308.01501 [quant-ph].
 - [19] T. d. L. Silva, L. Borges, and L. Aolita, Fourier-based quantum signal processing (2022), arXiv:2206.02826 [quant-ph].
 - [20] In this work, we will use the words *parameters* and *arguments* for making the difference between θ and x explicit, respectively.
 - [21] B. Casas and A. Cervera-Lierta, Multidimensional Fourier series with quantum circuits, *Physical Review A* **107**, 062612 (2023), arXiv:2302.03389 [quant-ph].
 - [22] A. Pérez-Salinas, D. López-Núñez, A. García-Sáez, P. Forn-Díaz, and J. I. Latorre, One qubit as a universal approximant, *Physical Review A* **104**, 012405 (2021).
 - [23] P. G. L. Dirichlet, Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données, arXiv:0806.1294 [math] (2008), arXiv:0806.1294 [math].
 - [24] P. Turán, Sur Les Fonctions Bornées et Intégrables, in *Leopold Fejér Gesammelte Arbeiten I*, edited by P. Turán (Birkhäuser, Basel, 1970) pp. 37–43.
 - [25] M. Arató and A. Rényi, Probabilistic proof of a theorem on the approximation of continuous functions by means of generalized Bernstein polynomials, **8**, 91.
 - [26] A. Manzano, D. Dechant, J. Tura, and V. Dunjko, Parametrized Quantum Circuits and their approximation capacities in the context of quantum machine learning (2023).
 - [27] G. H. Low and I. L. Chuang, Optimal Hamiltonian Simulation by Quantum Signal Processing, *Physical Review Letters* **118**, 010501 (2017), arXiv:1606.02685.
 - [28] T. Tilma and E. C. G. Sudarshan, Generalized Euler Angle Paramterization for SU(N), *Journal of Physics A: Mathematical and General* **35**, 10467 (2002), arXiv:math-ph/0205016.

1. Relevant theorems on universality

Theorem 3 (Fourier theorem). *Let $\mathcal{G}_L^F = \{g_L^F(x)\}$ be the sequence of sets of functions $g_L^F : [0, 1]^m \rightarrow \mathbb{C}$ of the form*

$$g_L^F(\mathbf{x}) = \sum_{\mathbf{n}} c_{\mathbf{n}} \exp(i2\pi\mathbf{n} \cdot \mathbf{x}) \quad (12)$$

for $\mathbf{n} = (n_1, \dots, n_m) \in \mathbb{N}^m$, $\|\mathbf{n}\|_{\infty} \leq L$ and $c_{\mathbf{n}} \in \mathbb{C}$. For every square-integrable function $f : [0, 1]^m \rightarrow \mathbb{C}$, $f \in \mathcal{F}$, the sequence \mathcal{G}_L^F is universal with respect to \mathcal{F} in the norm $\|\cdot\|_2$, if the coefficients are given by

$$c_{\mathbf{n}} = \int_{[0,1]^m} d\mathbf{x} f(\mathbf{x}) e^{-i2\pi\mathbf{n} \cdot \mathbf{x}} \quad (13)$$

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 $\{e^{i2\pi\mathbf{n} \cdot \mathbf{x}}\}$, with n a vector of integers. The Hilbert space is equipped with the inner product

$$\langle f(\mathbf{x}), g(\mathbf{x}) \rangle = \int_{[0,1]^m} f^*(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}, \quad (14)$$

and in particular

$$\langle e^{i2\pi\mathbf{n} \cdot \mathbf{x}}, e^{i2\pi\mathbf{l} \cdot \mathbf{x}} \rangle = \prod_{k=1}^m \delta_{n_k l_k}, \quad (15)$$

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

Theorem 4 (Universal approximation theorem, adapted from [13]). Let $\mathcal{G}_L^{\text{UAT}} = \{g_L^{\text{UAT}}(x)\}$ be the set of functions $g_L^{\text{UAT}} : [0, 1]^m \rightarrow \mathbb{C}$ of the form

$$g_L^{\text{UAT}}(x) = \sum_{n=1}^L \alpha_n \hat{\sigma}(\mathbf{w}_n \cdot \mathbf{x} + \phi_n), \quad (16)$$

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

$$\int_{[0, 1]^m} d\mu(\mathbf{x}) \hat{\sigma}(\mathbf{w} \cdot \mathbf{x} + \phi) = 0, \quad \mathbf{w} \in \mathbb{R}^m, \phi \in \mathbb{R} \Leftrightarrow \mu = 0, \quad (17)$$

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

The interpretation of this theorem can be done in terms of the Hilbert space discussed after Theorem 3. For arbitrary \mathbf{w}, ϕ , discriminatory functions satisfy that they have non-zero overlaps with all elements in the basis of functions. For exponential functions

$$e^{iwx} = \sum_{n=-\infty}^{\infty} c_n(w) e^{in2\pi x}, \quad (18)$$

$$c_n(w) = \frac{-i}{(2\pi n - w)} (e^{iw} - 1). \quad (19)$$

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. [13] for an in-depth proof.

2. Relevant theorems on universality of quantum circuits

Theorem 5 (Re-uploading: fixed encoding gates, adapted from [18, 19]). Consider the single-qubit circuits

$$U_L^f(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; x) = \left(\prod_{j=1}^L e^{i\hat{\sigma}_z \theta_j} e^{i\hat{\sigma}_y \phi_j} e^{i\hat{\sigma}_z 2\pi x} \right) e^{i\hat{\sigma}_z \theta_0} e^{i\hat{\sigma}_y \phi_0} e^{i\hat{\sigma}_z \lambda}, \quad (20)$$

where $\sigma_{\{y,z\}}$ are Pauli matrices, and $x \in [0, 1]$. The sequence of output function families $\mathcal{H}_L = \{h_L\}$, $h_L : [0, 1] \rightarrow \mathbb{C}$, $|h_L(x)| \leq 1$, $h_L(x) = \langle 0 | U_L^f(\boldsymbol{\theta}, \boldsymbol{\phi}, \lambda; x) | 0 \rangle$ is universal with respect to univariate square-integrable functions $f : [0, 1] \rightarrow \mathbb{C}$ with the constraint $|f(x)|^2 \leq 1$ in the $\|\cdot\|_2$ norm.

Proof. We recall theorem 3 in [18], formulated as follows.

Theorem 6 (Generalized quantum signal processing). Consider the building block

$$R(\theta, \phi, \lambda) = \begin{pmatrix} e^{i(\phi+\lambda)} \cos(\theta) & e^{i\phi} \sin(\theta) \\ e^{i\lambda} \sin(\theta) & -\cos(\theta) \end{pmatrix}. \quad (21)$$

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

$$\left(\prod_{j=1}^L R(\theta_j, \phi_j, x) \right) R(\theta_0, \phi_0, \lambda) = \begin{pmatrix} P(e^{ix}) & \cdot \\ Q(e^{ix}) & \cdot \end{pmatrix} \quad (22)$$

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

$$P'(e^{ix}) = e^{-ikx} P(e^{ix}) \quad (23)$$

if the argument is transformed according to $x \rightarrow -x$ 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

$$\begin{pmatrix} e^{-ix} & 0 \\ 0 & 1 \end{pmatrix} = e^{-ix} X \begin{pmatrix} e^{ix} & 0 \\ 0 & 1 \end{pmatrix} X. \quad (24)$$

We identify

$$\begin{pmatrix} e^{ix} & 0 \\ 0 & 1 \end{pmatrix} = e^{ix/2} \begin{pmatrix} e^{ix/2} & 0 \\ 0 & e^{-ix/2} \end{pmatrix} = e^{ix/2} e^{ix/2\hat{\sigma}_z} \quad (25)$$

$$\begin{pmatrix} e^{-ix} & 0 \\ 0 & 1 \end{pmatrix} = e^{-ix/2} \begin{pmatrix} e^{-ix/2} & 0 \\ 0 & e^{ix/2} \end{pmatrix} = e^{-ix/2} e^{-ix/2\hat{\sigma}_z} \quad (26)$$

By querying L times the first operator and L times the second, consecutively, we can represent any polynomial by virtue of [18]. 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

$$P(e^{ix}) = \sum_{j=-L}^L c_j e^{i j x}. \quad (27)$$

These polynomials are Fourier series up to degree L , and thus this family of functions is universal as stated in Theorem 3.

Finally, we bridge the gap between this result and our statement in Theorem 5. 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 (21) is decomposable as Equation (20) up to global phases. \square

Theorem 7 (Re-uploading: fixed encoding gates, multivariate, adapted from [16]). *There exists a sequence of quantum circuits $W_L(\mathbf{x}, \boldsymbol{\theta})$ with number number of qubits $\mathcal{O}(m \log L)$, depth $\mathcal{O}(m L^m \log L)$ and number of trainable parameters $\mathcal{O}(m L^m)$, such that the sequence of output function families $\mathcal{H}_L = \{h_L\}$, $h_L(\mathbf{x}) = \langle 0 | W_L(\mathbf{x}, \boldsymbol{\theta}) | 0 \rangle$ is universal with respect to multivariate continuous functions $f : [0, 1]^m \rightarrow \mathbb{C}$ with the constraint $|f(\mathbf{x})|^2 \leq 1$, in the $\|\cdot\|_\infty$ norm.*

Theorem 8 (Re-uploading: tunable encoding gates [22]). *Consider the single-qubit circuit*

$$U_L^w(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \mathbf{x}) = \left(\prod_{j=1}^L e^{i\hat{\sigma}_z \theta_j} e^{i\hat{\sigma}_y \phi_j} e^{i\hat{\sigma}_z \mathbf{w}_j \cdot \mathbf{x}} \right) e^{i\hat{\sigma}_z \theta_0} e^{i\hat{\sigma}_y \phi_0} e^{i\hat{\sigma}_z \lambda}, \quad (28)$$

with $\sigma_{\{y,z\}}$ being Pauli matrices, and $\mathbf{x} \in [0, 1]^m$. The sequence of output function families $\mathcal{H}_L = \{h_L\}$, $h_L(x) = \langle 0 | U_L^w(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \mathbf{x}) | 0 \rangle$ is universal with respect to multivariate continuous functions $f : [0, 1]^m \rightarrow \mathbb{C}$ with the constraint $|f(\mathbf{x})|^2 \leq 1$ in the norm $\|\cdot\|_\infty$.

3. Proof of Theorem 1

The road towards proving universality is to show that circuits as given in Equation (6) can approximate circuits of the form given in Equation (28), with arbitrary precision.

The first step is to approximate gates of the form $e^{iwx\hat{\sigma}_z}$, where x is the input and w is a tunable real weight, with circuits composed of fixed encoding gates. Inspection of Definition 2 allows us to identify

$$R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) \equiv \prod_{n=1}^N e^{i\theta_{j,k,n}\hat{\sigma}_z} e^{i\phi_{j,k,n}\hat{\sigma}_y} e^{i\pi x_k \hat{\sigma}_z} \approx e^{iwx_k \hat{\sigma}_z}, \quad (29)$$

as a requirement for the approximation. We just need to show that there exist parameters $\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}$, depending on $w_k \in \mathbb{R}$, such that the above equation holds to arbitrary precision ε . This is possible by means of the following result.

Lemma 1. *For a gate $R(\boldsymbol{\theta}, \boldsymbol{\phi}, x)$ as given in Equation (29), for any $w \in \mathbb{R}$ and for any $\varepsilon > 0$, there exists a value N and parameters $(\boldsymbol{\theta}, \boldsymbol{\phi})$ such that*

$$\sup_{x \in [0, 1]} \|R(\boldsymbol{\theta}, \boldsymbol{\phi}, x) - \exp(iwx\hat{\sigma}_z)\|_F \leq \varepsilon, \quad (30)$$

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

Notice that the Frobenius norm $\|\cdot\|_F$ is a matrix norm, and not a function norm as other norms used through the text. In this and subsequent proofs, the ∞ -norm for functions is made explicit as $\sup_{x \in [0,1]^m}$ to avoid confusion with matrix norms. The Frobenius norm of a matrix is defined as

$$\|A\|_F = \sqrt{\text{Tr}(A^\dagger A)} \quad (31)$$

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 5 we know that it is possible to construct arbitrary polynomials $P(e^{ix})$ of degree at most N into a unitary matrix as

$$R_N(\boldsymbol{\theta}, \boldsymbol{\phi}, x) = \begin{pmatrix} P_N(x) & -Q_N^*(x) \\ Q_N(x) & P_N^*(x) \end{pmatrix} \approx \begin{pmatrix} e^{iwx} & 0 \\ 0 & e^{-iwx} \end{pmatrix}. \quad (32)$$

In the language of Theorem 5, $P(x)$ 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 $f : \mathbb{R} \rightarrow \mathbb{C}$ 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*

$$P_N(x) = \frac{1}{N+1} \sum_{n=-N}^N g_n^F(x), \quad (33)$$

or equivalently

$$P_N(x) = \sum_{n=-N}^N \frac{N+1-|n|}{N+1} c_n e^{in\pi x}. \quad (34)$$

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 [24]

$$K_N(x) = \frac{1}{N+1} \sum_{n=0}^N \sum_{k=-n}^n e^{ikx}. \quad (35)$$

which satisfies

$$\int_x dx K_N(x) = 1. \quad (36)$$

Thus

$$\|P_N\|_\infty = \sup_{x \in [0,1]} \left| \int dt e^{iw(x-t)} K_N(t) dt \right| \leq \sup_{x \in [0,1]} |e^{iwx}| \left| \int dt K_N(t) dt \right| = 1. \quad (37)$$

Therefore, we can implement this function within a unitary operation.

Second, we can recall Fejér's theorem [24].

Theorem 9 (Fejér's theorem, [24]). *Let $f : \mathbb{R} \rightarrow \mathbb{C}$ 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*

$$\sup_{x \in [0,1]} |f(x) - P_N(x)|_\infty \leq \varepsilon. \quad (38)$$

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 (32). With this choice, we can now write the Frobenius norm of the difference matrix as

$$\|R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\hat{\sigma}_z}\|_F = \sqrt{2} \sqrt{|P_N(x) - e^{iwx}|^2 + |Q_N(x)|^2}. \quad (39)$$

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

$$|e^{iwx} - P_N(x)| \geq 1 - |P_N(x)| \geq \frac{1 - |P_N(x)|^2}{2} = \frac{|Q_N(x)|^2}{2}, \quad (40)$$

hence

$$\|R_{j,k}(\theta_{j,k}, \phi_{j,k}, x) - e^{iwx\hat{\sigma}_z}\|_F \leq \sqrt{2} \sqrt{|e^{iwx} - P_N(x)|^2 + 2|e^{iwx} - P_N(x)|}. \quad (41)$$

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 [26]. We define the period of this auxiliary function as

$$a(x) = \begin{cases} e^{iwx} & 0 \leq x \leq 1 \\ e^{iw(2-x)} & 1 < x < 2 \end{cases}, \quad (42)$$

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 fulfills 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 $R_{j,k}(\theta_{j,k}, \phi_{j,k}, x)$, which has $e^{inx\hat{\sigma}_z}$ in it. Thus, we can find a gate $R_{j,k}(\theta_{j,k}, \phi_{j,k}, x)$ satisfying that for all $\varepsilon > 0$ there exists a N such that

$$\sup_{x \in [0, 1]} \|R_{j,k}(\theta_{j,k}, \phi_{j,k}, x) - e^{iwx\hat{\sigma}_z}\|_F \leq \varepsilon, \quad (43)$$

and concludes the proof. \square

We can also bound the errors in Lemma 1.

Corollary 2. *The circuit $R(\theta, \phi, x)$ from Lemma 1 can be realized using N fixed-encoding gates, with*

$$N \in \tilde{\mathcal{O}}(w + \varepsilon^{-2}). \quad (44)$$

Proof. We begin with the function defined in Equation (42). This function is continuous and periodic. Its Fourier series is given by the sets of coefficients

$$c_n = \int_0^1 e^{i(w-n)x} dx + e^{i2w} \int_1^2 e^{-i(w+n)x} = -i \frac{w((-1)^n e^{iw} - 1)}{w^2 - (n\pi)^2}. \quad (45)$$

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

$$P_N(x) - e^{iwx} = \sum_{n=-N}^N c_n \frac{-|n|}{N+1} e^{inx} - \sum_{|n|=N+1}^{\infty} c_n e^{inx}, \quad (46)$$

hence by the triangle inequality

$$\sup_{x \in [0, 1]} |P_N(x) - e^{iwx}| \leq \frac{2}{N+1} \left| \sum_{n=1}^N n c_n \right| + 2 \left| \sum_{n=N+1}^{\infty} c_n \right|. \quad (47)$$

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,

$$\left| \sum_{n=1}^N n c_n \right| \leq \sum_{n=1}^N n |c_n| \leq \frac{2}{\pi} \left(\sum_{n=1}^N \frac{n}{n^2 - \frac{1}{4}} \right) \leq \frac{2}{\pi} \left(\frac{4}{3} + \int_1^N dx \frac{x}{x^2 - \frac{1}{4}} \right) = \frac{2}{\pi} \left(\frac{4}{3} + \frac{1}{2} \log \left(\frac{4N^2 - 1}{3} \right) \right). \quad (48)$$

For the second term,

$$\left| \sum_{n=N+1}^{\infty} c_n \right| \leq \sum_{n=N+1}^{\infty} |c_n| \leq \frac{2}{\pi} \int_N^{\infty} \frac{1}{n^2 - 1/4} \leq \frac{2}{\pi} \int_N^{\infty} \frac{1}{(n - 1/4)^2} = \frac{2}{\pi} \frac{1}{N - 1/4}. \quad (49)$$

We see that the first term dominates over the second term. Now, we only need to re-arrange terms from Equation (47) to show

$$\sup_{x \in [0,1]} |P_N(x) - e^{iwx}| \in \tilde{\mathcal{O}}(N^{-1}). \quad (50)$$

We recover now the Frobenius norm from Lemma 1

$$\|R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\hat{\sigma}_z}\|_F = \sqrt{2}\sqrt{|P_N(x) - e^{iwx}|^2 + |Q_N(x)|^2}. \quad (51)$$

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

$$\|R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\hat{\sigma}_z}\|_F \leq \sqrt{6|P_N(x) - e^{iwx}|}, \quad (52)$$

and subsequently

$$\sup_{x \in [0,1]} \|R_{j,k}(\boldsymbol{\theta}_{j,k}, \boldsymbol{\phi}_{j,k}, x) - e^{iwx\hat{\sigma}_z}\|_F \in \tilde{\mathcal{O}}(N^{-1/2}). \quad (53)$$

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

$$N \in \tilde{\Omega}(w + \varepsilon^{-2}) \quad (54)$$

gates. \square

For the next step, we subsequently apply Lemma 1 to show that there exist parameters $\boldsymbol{\theta}', \boldsymbol{\phi}'$ such that, for given $W, \boldsymbol{\theta}, \boldsymbol{\phi}$

$$U_{L'}^{\text{mf}}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \mathbf{x}) \approx \left(\prod_{j=1}^L e^{i\theta'_j \hat{\sigma}_z} e^{i\phi'_j \hat{\sigma}_y} e^{i \sum_{k=1}^m w_{j,k} x_k \hat{\sigma}_z} \right) e^{i\theta'_0 \hat{\sigma}_z} e^{i\phi'_0 \hat{\sigma}_y} e^{i\lambda'} = U_L^{\text{w}}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \mathbf{x}). \quad (55)$$

In a nutshell, we take our previous result to approximate a gate of the form $e^{iwx_k \hat{\sigma}_z}$. Then, we can subsequently apply these gates to approximately obtain $\prod_k e^{ix_k w_k \hat{\sigma}_z} = \exp(i \sum_k x_k w_k \hat{\sigma}_z)$, which are tunable weights. This step bridges the gap with models relying on tunable weights. We refer the reader to Figure 3 for an illustration of this step.

Lemma 2. Consider the re-uploading models $U_L^{\text{w}}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \mathbf{x})$ from Equation (28) and $U_{L'}^{\text{mf}}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \mathbf{x})$ from Equation (6). For any $\varepsilon > 0$, $L, (\boldsymbol{\theta}, \boldsymbol{\phi}, W)$ there exists value $L' \geq L$, $(\boldsymbol{\theta}', \boldsymbol{\phi}')$ satisfying

$$\sup_{\mathbf{x} \in [0,1]^m} |\text{Tr}((U_L^{\text{w}}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \mathbf{x}) - U_{L'}^{\text{mf}}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \mathbf{x})) |0\rangle \langle 0|) \leq \varepsilon \quad (56)$$

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

$$\begin{aligned} \min_{\boldsymbol{\theta}, \boldsymbol{\phi}} \sup_{\mathbf{x}} |\text{Tr}((U_L^{\text{w}}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \mathbf{x}) - U_{L'}^{\text{mf}}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \mathbf{x})) |0\rangle \langle 0|) | \leq \\ \min_{\boldsymbol{\theta}, \boldsymbol{\phi}} \sup_{\mathbf{x}} \|U_L^{\text{w}}(\boldsymbol{\theta}, \boldsymbol{\phi}, W, \lambda; \mathbf{x}) - U_{L'}^{\text{mf}}(\boldsymbol{\theta}', \boldsymbol{\phi}', \lambda'; \mathbf{x})\|_{\infty}, \end{aligned} \quad (57)$$

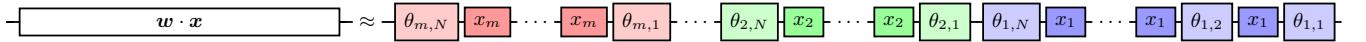


FIG. 3. Quantum circuit allowing the universality of quantum re-uploading models, with fixed encoding gates (in this figure represented by $\{x_1, x_2, \dots, x_m\}$). By Lemma 1, we show that gates of the form $e^{iwx\hat{\sigma}_z}$ can be approximated to arbitrary accuracy by a re-uploading circuit with fixed encoding gates. The successive application of these approximations allow for an approximation of $e^{i\mathbf{w} \cdot \mathbf{x} \hat{\sigma}_z}$, where \mathbf{x} is now multidimensional.

where the matrix norm is the ∞ -Schatten norm. Both $U_L^w(\boldsymbol{\theta}, \phi, W, \lambda; \mathbf{x})$ and $U_{L'}^{mf}(\boldsymbol{\theta}', \phi', \lambda'; \mathbf{x})$ are constructed through layers. Following the rational of Equation (29), we group the gates in $U_{L'}^{mf}(\boldsymbol{\theta}', \phi', \lambda'; \mathbf{x})$ to approximate $e^{iwf\hat{\sigma}_z}$ in $U_L^w(\boldsymbol{\theta}, \phi, W, \lambda; \mathbf{x})$. Then this is done, we can match the parameters in $(\boldsymbol{\theta}', \boldsymbol{\theta}), (\phi', \phi)$ to apply the same operations between encoding gates as in $U_L^w(\boldsymbol{\theta}, \phi, W, \lambda; \mathbf{x})$. This fixes the parameters $\boldsymbol{\theta}', \phi'$, thus giving an upper bound to the previous equation.

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

$$\|uU - vV\| = \|uU - vU + vU - vV\| \leq \|u - v\| + \|U - V\|. \quad (58)$$

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

$$\|U_L^w(\boldsymbol{\theta}, \phi, W, \lambda; \mathbf{x}) - U_{L'}^{mf}(\boldsymbol{\theta}', \phi', \lambda'; \mathbf{x})\|_\infty \leq \sum_{j=1}^{L'} \sum_{k=1}^m \sup_{x_k \in [0,1]} \|R_{j,k}(\boldsymbol{\theta}_{j,k}, \phi_{j,k}, x) - e^{iw_{j,k}x_k\hat{\sigma}_z}\|_\infty. \quad (59)$$

By virtue of Lemma 1, the Frobenius norm $\|R_{j,k}(\boldsymbol{\theta}_{j,k}, \phi_{j,k}, x) - e^{iw_{j,k}x_k\hat{\sigma}_z}\|_F$ 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. \square

For the final step, we just need to consider Lemma 2 and the triangle inequality. A model with tunable weights can approximate any function 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 10. *Consider the model $U_{L'}^{mf}(\boldsymbol{\theta}', \phi', \lambda'; \mathbf{x})$ from Equation (6). The sequence of output function families $\mathcal{H} = \{h_{L'}^f\}$, with*

$$h_{L'}^f(\mathbf{x}) = \langle 0 | U_{L'}^{mf}(\boldsymbol{\theta}', \phi', \lambda'; \mathbf{x}) | 0 \rangle \quad (60)$$

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

Proof. This corollary is an immediate consequence of Lemma 2. The considered model can approximate a multivariate function $h_L(\mathbf{x}) = \langle 0 | U_L^w(\boldsymbol{\theta}, \phi, W, \lambda; \mathbf{x}) | 0 \rangle$ output by a model with tunable weights. The triangle inequality implies

$$\|h_{L'}^f(\mathbf{x}) - f(\mathbf{x})\| \leq \|h_L(\mathbf{x}) - h_{L'}^f(\mathbf{x})\| + \|h_L(\mathbf{x}) - f(\mathbf{x})\|. \quad (61)$$

Each term can be made arbitrarily small. The first one by virtue of Lemma 1, and the second one following Theorem 8. Hence, universality is guaranteed. \square

This last theorem is analogous to the one written in the main text.

4. Proof of Corollary 1

Proof. To prove Corollary 1, we follow the same procedure 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 $g_N(\mathbf{x})$ from Theorem 4 and its discrete-weights approximation is given by

$$\sup_{\mathbf{x} \in [0,1]^m} \left| \sum_{j=1}^L \gamma_j \prod_{k=1}^m e^{iw_{j,k}x_k} - \sum_{j=1}^L \gamma_j \prod_{k=1}^m P_{N,\omega_{j,k}}(x_k) \right| \leq \sum_{j=1}^L |\gamma_j| \sup_{\mathbf{x} \in [0,1]^m} \left| \prod_{k=1}^m e^{iw_{j,k}x_k} - \prod_{k=1}^m P_{N,\omega_{j,k}}(x_k) \right|, \quad (62)$$

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

$$\sup_{\mathbf{x} \in [0,1]^m} |f(\mathbf{x})| = \sup_{\mathbf{x} \in [0,1]^m} |e^{i\alpha \cdot \mathbf{x}} f(\mathbf{x})|, \quad (63)$$

we can perform an equivalent trick to the one in Equation (58) and find

$$\sup_{x \in [0,1]^m} \left| \sum_{j=1}^L \gamma_j \prod_{k=1}^m e^{iw_{j,k}x_k} - \sum_{j=1}^L \gamma_j \prod_{k=1}^m \omega_{j,k}(x_k) \right| \leq \sum_{j=1}^L |\gamma_j| \sum_{k=1}^m \sup_{x_k \in [0,1]} |e^{iw_{j,k}x_k} - P_{N,\omega_{j,k}}(x_k)|. \quad (64)$$

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. \square

5. Euler angles for $SU(N)$

We follow the construction in [28] for Euler rotations of arbitrary dimensions. We define first the generators of the corresponding algebra $\mathfrak{su}(N)$ as generalized Gell-Mann matrices

$$\hat{g}_3 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad (65)$$

$$\hat{g}_{(k-1)^2+1} = \underbrace{\begin{pmatrix} 0 & \cdots & -i & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ i & \cdots & 0 & & \cdots 0 \\ \hline k & & & & \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & & \cdots 0 \end{pmatrix}}_{(k-1)^2+1} \quad (66)$$

$$\hat{g}_{N^2-1} = \sqrt{\frac{2}{N^2-N}} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix} \quad (67)$$

There exists an Euler decomposition of arbitrary dimension by employing a recursive construction in which matrices in $SU(N)$ are defined as matrices in $SU(N-1)$ plus extra parameters.

$$U = \prod_{2 \leq k \leq N} e^{i\hat{g}_3\theta_{2k-3}} e^{i\hat{g}_{(k^2-1)+1}\theta_{2k-2}} [SU(N-1)] e^{i\hat{g}_{N^2-1}\theta_{N^2-1}} \quad (68)$$

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 $V_m(\mathbf{x})$ from Equation (72). Such definition allows for a one-to-one identification of values \mathbf{x} and the corresponding Euler angles for the diagonal elements.

6. Proof of Theorem 2

Proof. We begin by considering the \mathbf{x} -encoding gate

$$V_m(\mathbf{x}) = \begin{pmatrix} e^{i\pi x_1/2} & 0 & \cdots & 0 & \cdots \\ 0 & e^{-i\pi x_1/2} & \cdots & 0 & \cdots \\ 0 & 0 & e^{i\pi x_2/2} & 0 & \vdots \\ 0 & 0 & 0 & e^{-i\pi x_2/2} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (69)$$

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

$$\Pi_1 V_m(\mathbf{x}) = \begin{pmatrix} e^{i\pi x_1/2} & 0 & \cdots & 0 & \cdots \\ 0 & e^{-i\pi x_1/2} & \cdots & 0 & \cdots \\ 0 & 0 & e^{-i\pi x_2/2} & 0 & \vdots \\ 0 & 0 & 0 & e^{i\pi x_2/2} & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (70)$$

These permutations are implementable with gates $R(\boldsymbol{\theta}, \boldsymbol{\phi})$, since R specifies any unitary gate, except for the relative phases among columns [28], and no phases are needed for the permutations. It is immediate to see

$$R_1(\mathbf{x}) = V_m(\mathbf{x}) \Pi V_m(\mathbf{x}) = \begin{pmatrix} e^{i\pi x_1} & 0 & \cdots & 0 & \cdots \\ 0 & e^{-i\pi x_1} & \cdots & 0 & \cdots \\ 0 & 0 & 1 & 0 & \vdots \\ 0 & 0 & 0 & 1 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \left[\begin{matrix} e^{ix_1} & 0 \\ 0 & e^{-ix_1} \end{matrix} \right] & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} e^{i\pi x_1 \hat{\sigma}_z} & 0 \\ 0 & I \end{pmatrix}. \quad (71)$$

The construction of $R_1(\mathbf{x})$ requires an overhead of $\mathcal{O}(1)$ encoding gates of the form $V_m(\mathbf{x})$ to achieve it. The same condition holds for any other $R_k(\mathbf{x})$. These blocks, up to permutations that can be reabsorbed in the parameterized gates, allow us to conduct the algorithms in Theorem 5 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 5, this construction allows for universality as well in the multi-qubit case. \square

7. An extension to Theorem 2

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

$$V'_m(\mathbf{x}) = \sum_{k=1}^m e^{i\pi x_k/m} |k\rangle\langle k| + e^{-i\sum_{i=1}^m \pi x_k/m} |m+1\rangle\langle m+1| = \begin{pmatrix} e^{i\pi x_1/m} & 0 & 0 & \cdots & 0 \\ 0 & e^{i\pi x_2/m} & 0 & \cdots & 0 \\ 0 & 0 & e^{i\pi x_3/m} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{-i\frac{\pi}{m} \sum_{k=1}^m x_k} \end{pmatrix} \quad (72)$$

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 [28], and no phases are needed for the permutations. An example of this permutation is

$$\Pi_{1,1} V'_m(\mathbf{x}) = \begin{pmatrix} e^{i\pi x_1/m} & 0 & 0 & \cdots & 0 \\ 0 & e^{-i\frac{\pi}{m} \sum_{k=1}^m x_k} & 0 & \cdots & 0 \\ 0 & 0 & e^{i\pi x_2/m} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{i\pi x_{m-1}/m} \end{pmatrix}. \quad (73)$$

We apply subsequent permutations until finding

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

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

$$V'_{\Pi_1}(-\mathbf{x}) = \prod_{j=1}^m \Pi_{2,j}(V'_{\Pi_1}(\mathbf{x})) = e^{i\pi x_1/m} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & e^{-i\pi x_1} & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (75)$$

We can identify

$$R_1(\mathbf{x}) = V'_{\Pi_1}(-\mathbf{x})V'_{\Pi_1}(\mathbf{x}) = \begin{pmatrix} e^{i\pi x_1} & 0 \\ 0 & e^{-i\pi x_1} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ I \end{pmatrix} = \begin{pmatrix} e^{i\pi x_1 \hat{\sigma}_z} & 0 \\ 0 & I \end{pmatrix}. \quad (76)$$

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

From this point, we can repeat the steps in Appendix 6 and formulate the following theorem.

Theorem 11. *Let $U_L(\Theta, \Phi, \lambda; \mathbf{x})$ be the gate defined in Definition 3, with encoding gate $V'_m(\mathbf{x})$. The family of output functions $\mathcal{H} = \{h_L\}$, with*

$$h_L(\mathbf{x}) = \langle 0 | U_L(\Theta, \Phi, \lambda; \mathbf{x}) | 0 \rangle \quad (77)$$

is universal with respect to the norm $\|\cdot\|_\infty$ for all continuous complex functions $f(\mathbf{x}), \|f(\mathbf{x})\|_\infty \leq 1$, for $\mathbf{x} \in [0, 1]^m$. 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 2 is that this case requires one qubit less. However, the overhead in depth as compared to the single-qubit case is beneficial to Theorem 2.

- [1] P. W. Shor, Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer, 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, A variational eigenvalue solver on a photonic quantum processor, 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. Havlfíček, A. D. Cárcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature **567**, 209 (2019).
- [6] M. Schuld and N. Killoran, Quantum machine learning in feature Hilbert spaces, Physical Review Letters **122**, 040504 (2019), arXiv:1803.07128.
- [7] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum Circuit Learning, Physical Review A **98**, 032309 (2018), arXiv:1803.00745.
- [8] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum **4**, 226 (2020).
- [9] 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.
- [10] H. M. Sheffer, A set of five independent postulates for Boolean algebras, with application to logical constants, Transactions of the American Mathematical Society **14**, 481 (1913).
- [11] D. A. Barrington, Bounded-width polynomial-size branching programs recognize exactly those languages in $NC1$, Journal of Computer and System Sciences **38**, 150 (1989).
- [12] A. Cosentino, R. Kothari, and A. Paetznick, Dequantizing Read-once Quantum Formulas, in *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013)* (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013) pp. 80–92.
- [13] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems **2**, 303 (1989).
- [14] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Networks **4**, 251 (1991).
- [15] V. I. Arnold, On the representation of functions of several variables as a superposition of functions of a smaller number of variables, in *Collected Works: Representations of Functions, Celestial Mechanics and KAM Theory, 1957–1965*, edited by V. I. Arnold, A. B. Givental, B. A. Khesin, J. E. Marsden, A. N. Varchenko, V. A. Vassiliev, O. Y. Viro, and V. M. Zakalyukin (Springer, Berlin, Heidelberg, 2009) pp. 25–46.

- [16] Z. Yu, Q. Chen, Y. Jiao, Y. Li, X. Lu, X. Wang, and J. Z. Yang, Non-asymptotic Approximation Error Bounds of Parameterized Quantum Circuits (2024), arXiv:2310.07528.
- [17] M. Schuld, R. Sweke, and J. J. Meyer, The effect of data encoding on the expressive power of variational quantum machine learning models, Physical Review A **103**, 032430 (2021), arXiv:2008.08605.
- [18] D. Motlagh and N. Wiebe, Generalized Quantum Signal Processing (2024), arXiv:2308.01501 [quant-ph].
- [19] T. d. L. Silva, L. Borges, and L. Aolita, Fourier-based quantum signal processing (2022), arXiv:2206.02826 [quant-ph].
- [20] In this work, we will use the words *parameters* and *arguments* for making the difference between θ and x explicit, respectively.
- [21] B. Casas and A. Cervera-Lierta, Multidimensional Fourier series with quantum circuits, Physical Review A **107**, 062612 (2023), arXiv:2302.03389 [quant-ph].
- [22] A. Pérez-Salinas, D. López-Núñez, A. García-Sáez, P. Forn-Díaz, and J. I. Latorre, One qubit as a universal approximant, Physical Review A **104**, 012405 (2021).
- [23] P. G. L. Dirichlet, Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données, arXiv:0806.1294 [math] (2008), arXiv:0806.1294 [math].
- [24] P. Turán, Sur Les Fonctions Bornées et Intégrables, in *Leopold Fejér Gesammelte Arbeiten I*, edited by P. Turán (Birkhäuser, Basel, 1970) pp. 37–43.
- [25] M. Arató and A. Rényi, Probabilistic proof of a theorem on the approximation of continuous functions by means of generalized Bernstein polynomials, **8**, 91.
- [26] A. Manzano, D. Dechant, J. Tura, and V. Dunjko, Parametrized Quantum Circuits and their approximation capacities in the context of quantum machine learning (2023).
- [27] G. H. Low and I. L. Chuang, Optimal Hamiltonian Simulation by Quantum Signal Processing, Physical Review Letters **118**, 010501 (2017), arXiv:1606.02685.
- [28] T. Tilma and E. C. G. Sudarshan, Generalized Euler Angle Paramterization for SU(N), Journal of Physics A: Mathematical and General **35**, 10467 (2002), arXiv:math-ph/0205016.