Pseudo-dimension of quantum circuits

Matthias C. Caro[1] · Ishaun Datta[1,2]

Abstract

We characterize the expressive power of quantum circuits with the pseudo-dimension, a measure of complexity for probabilistic concept classes. We prove pseudo-dimension bounds on the output probability distributions of quantum circuits; the upper bounds are polynomial in circuit depth and number of gates. Using these bounds, we exhibit a class of circuit output states out of which at least one has exponential gate complexity of state preparation, and moreover demonstrate that quantum circuits of known polynomial size and depth are PAC-learnable.

Keywords Quantum computing · Computational learning theory · Complexity theory

1 Introduction

An important line of research in classical learning theory is characterizing the expressive power of function classes using complexity measures. Such complexity bounds can in turn be used to bound the size of training data required for learning. Among the most prominent of these are the Vapnik-Chervonenkis (VC) dimension introduced by Vapnik and Chervonenkis (1971). Other well-known measures are the pseudo-dimension due to Pollard (1984), the fat-shattering dimension due to Alon et al. (1997), the Rademacher complexities (see Bartlett and Mendelson 2002), and more generally covering numbers in metric spaces.

The goal of characterizing an object's expressive power also appears in different guises throughout quantum information. A well-known example is quantum state tomography. Aaronson (2007) related a variant of state tomography to a classical learning task whose fat-shattering dimension can be bounded using a particular function class

related to the set of quantum states. Associated with this is a corresponding upper bound on sample complexity.

Aaronson (2007) observes that there is no analogous theorem for general quantum process tomography, but leaves as an open question whether there are restricted classes of operations that are information-efficiently learnable. We answer this question in the affirmative. In particular, we show that for quantum circuits with depth and size polynomial in the number of qubits, quantum process tomography is possible using only polynomially many examples.

Gate complexity of unitary implementation and state preparation are yet another example of how one may capture the richness of a function class that corresponds to a quantum computational process (see, e.g., Aaronson 2016). For unitary complexity, the challenge is to determine, e.g., how many two-qubit unitaries (i.e., two-qubit logical gates, in a computational setting) are required to implement a certain multi-qubit unitary (i.e., a quantum circuit). For the gate complexity of state preparation, it is to determine how many unitaries produce a certain multi-qubit state. An alternative perspective, adopted in this work, is to consider the expressive power of a set of circuits with a fixed number of unitaries.

In this work we describe a new way of applying complexity measures from classical learning, specifically pseudo-dimension, to quantum information. We associate with a quantum circuit a natural probabilistic function class describing the outcome probabilities of measurements performed on the circuit output. In this way, a function class corresponding to a quantum circuit can be studied with the classical tool of pseudo-dimension. Here, we show that the pseudo-dimension of such a class can be bounded in

✉ Matthias C. Caro
caro@ma.tum.de

Ishaun Datta
idatta@stanford.edu

[1] Department of Mathematics, Technical University of Munich, Munich, Germany

[2] Institute for Computational and Mathematical Engineering, Stanford University, Stanford, USA

terms of a polynomial of the circuit depth and size. We also give two applications of these bounds, one for the gate complexity of quantum state preparation, the other in learnability of quantum circuits.

These findings are noteworthy not only because of the results themselves, but because we demonstrate the power of pseudo-dimension to gain insight into quantum computation. We hope that these tools may be applied to other problems in quantum computing in future work.

1.1 Related work

Aaronson (2007) showed that using the framework of PAC learning, one can introduce a variant of quantum state tomography and prove an upper bound on the required number of copies of the unknown state. This idea was developed further in Aaronson et al. (2018) and Aaronson (2018).

Motivated by Aaronson's work, Cheng et al. (2016) use pseudo-dimension and fat-shattering dimension to characterize the learnability of measurements, as a dual problem to learning the state. We apply this mathematical framework to study the problem of learning the circuit itself, in particular by offering a natural function class corresponding to a quantum circuit.

Rocchetto (2017) proved that stabilizer states, prevalent in error correction, are *computationally* efficiently learnable, establishing a connection between efficient classical simulability and computationally efficient learnability. This was realized experimentally for small optical systems in Rocchetto et al. (2019). Similarly, in Section 5 we pose as an open problem whether there are quantum operations that can be PAC-learned with modest computation, which could then in principle be demonstrated in an experiment.

In Chung and Lin (2018), the authors study the problem of PAC learning classes of functions with computational basis states as input and quantum output, possibly mixed. We highlight two main differences: first, whereas we assume the training data to be measurement statistics, Chung and Lin (2018) consider examples given as classical-quantum states. Thus, the two scenarios are not directly comparable. Our learning result yields a semi-classical strategy for the problem described in Chung and Lin (2018), though it is possibly suboptimal. Second, the learnability result of Chung and Lin (2018) is only for finite concept classes, whereas our result does not have this restriction. While Chung and Lin (2018) show learnability of quantum circuits with a finite gate set, we allow for arbitrary 2-qudit gates, i.e., a continuous gate set. Note that our corresponding notions of learnability differ.

While we take a formal approach to learning quantum circuits, others have studied learning unitaries numerically, e.g., with heuristics such as gradient descent (Kiani et al.

2020). Practical machine learning algorithms have also been used for state tomography by Torlai et al. (2018), and similar techniques could be applied to restricted classes of process tomography.

Another branch of quantum learning deals with whether quantum examples can decrease the information-theoretic complexity of learning a classical function. There are different flavors of this question, e.g., depending on whether learning is distribution-specific or distribution-independent. Arunachalam and de Wolf (2017) gives an overview of some of these aspects of quantum learning.

In classical learning theory, bounding the complexity measures of function classes (based on complexity-theoretic assumptions) has been studied widely. Goldberg and Jerrum (1995) derived an upper bound on the VC-dimension of a function class in terms of the runtime required by an algorithm implementing the elements of that class. Karpinski and Macintyre (1997) established an analogous bound for the function class implemented by a neural network (for various activation functions) in terms of the number of nodes and the number of programmable parameters of the network. Koiran (1996) demonstrated that by bounding the complexity of function classes implemented on a given architecture, one can lower bound the size of an architecture implementing a specific "hard" function.

1.2 Overview of results

We consider the general scenario in which one measures the output state of a 2-local qu*d*it quantum circuit, generating a probability distribution. We do not assume *geometric* locality, i.e., we do not assume that 2-qu*d*it unitaries act on neighboring qu*d*its. We show an upper bound on the pseudo-dimension of the distributions arising from these quantum circuits. By doing so, we provide insight into the complexity or "hardness" of the circuit and the output state that gives rise to the probability distribution. Below, we provide informal statements of the key results.

Theorem (Pseudo-dimension bounds, Informal) *Consider quantum circuits with fixed architecture, namely those for which the input qudits of the 2-qudit gates are specified, but the gates may vary subject to this constraint. That is, we allow for arbitrary 2-qudit unitaries, and in particular we do not restrict ourselves to a finite gate library.*

Parameterize a quantum circuit $\mathcal{N}$ by its qudit dimension d, depth δ, and number of gates or size γ.

Theorem 2: *For a suitable function class $\mathscr{F}_{\mathcal{N}}$ corresponding to the possible probability distributions formed by product measurements in the computational basis on the circuit output, $Pdim(\mathscr{F}_{\mathcal{N}}) \leq \mathcal{O}(d^4 \cdot \gamma \log \gamma)$.*

Consider quantum circuits with variable architecture, i.e., those for which the input qudits of the gates are not specified. For such circuits of depth δ and number of gates or size γ, one may similarly define function classes $\mathscr{F}_{\delta,\gamma}$ for circuits whose gates are unitaries, and $\mathscr{G}_{\delta,\gamma}$ for circuits whose gates are quantum operations, which describe the possible probability distributions formed by product measurements on the circuit output. Then,

Theorem 3: $Pdim(\mathscr{F}_{\delta,\gamma}) \leq \mathscr{O}(\delta \cdot d^4 \cdot \gamma^2 \log \gamma)$.
Theorem 4: $Pdim(\mathscr{G}_{\delta,\gamma}) \leq \mathscr{O}(\delta \cdot d^8 \cdot \gamma^2 \log \gamma)$.

All upper bounds are polynomial in the dimension d, the depth δ, and the size γ.

In Section 4.1, we demonstrate how to apply these complexity upper bounds to explicitly construct, for each $n \in \mathbb{N}$, a finite-but-large set of n-qubit quantum states, out of which at least one cannot be implemented by a 2-local qudit circuit of subexponential depth or size.

Theorem (Gate Complexity of State Preparation, Informal)

For any subset $C \subseteq \{|x0\rangle\}_{x \in \{0,1\}^n}$, define

$$|\psi_C\rangle = \begin{cases} \frac{1}{\sqrt{|C|}} \sum\limits_{|x0\rangle \in C} |x0\rangle & \text{if } C \neq \emptyset \\ |0\rangle^{\otimes n} \otimes |1\rangle & \text{if } C = \emptyset. \end{cases}$$

If each state in $\{|\psi_C\rangle\}_C$ can be generated from the input state $|0\rangle^{\otimes(n+1)}$ by some circuit of depth δ and size γ, then $2^n \leq \mathscr{O}\left(\delta \cdot \gamma^2 \log \gamma\right)$. As a corollary, there exists at least one such C so that $|\psi_C\rangle$ requires a circuit exponential in depth and size.

Analogously to Aaronson (2007), in Section 4.2 we use our pseudo-dimension bounds to prove a relaxed variant of quantum process tomography, which following Aaronson's terminology can be called *pretty-good circuit tomography*:

Theorem (Learnability, Informal) *Given a circuit with depth Δ and size Γ, both polynomial in the number of qudits and known in advance to the learner, polynomially-many training examples, each a triple of input state, output measurement, and corresponding probability, suffice to learn the quantum operation implemented by a 2-local quantum circuit of depth Δ and size Γ.*

That is, for confidence δ, accuracy ε, and error margins α and β, all in $(0, 1)$, a candidate circuit of depth Δ and size Γ that performs sufficiently well (in a sense made rigorous in Section 4.2*) on*

$$\mathscr{O}\left(\frac{1}{\varepsilon}\left(\Delta d^8 \Gamma^2 \log \Gamma \log^2\left(\frac{\Delta d^8 \Gamma^2 \log \Gamma}{(\beta - \alpha)\varepsilon}\right) + \log \frac{1}{\delta}\right)\right)$$

many samples will with probability at least $1 - \delta$ approximate the actual circuit from which the samples are drawn.

In this framework, each training example is a three-tuple of the input state, the observed measurement outcome, and the corresponding measurement probability. Alternately, one may take each training example as a two-tuple of the input state and the measurement outcome, whose probability is the corresponding measurement probability (see Aaronson 2007, Appendix 8).

We review the basics of quantum information, quantum computation, and classical learning theory in Section 2. We also discuss prior classical results as motivation. Section 3 contains our main results on the pseudo-dimension of quantum circuits and the respective proofs. In Section 4, we apply these results to fin lower bounds on the gate complexity of quantum state preparation and to a learning problem for quantum operations. We conclude with open questions in Section 5.

2 Preliminaries

As our readership includes both physicists and computer scientists, in this section we review the mathematical frameworks of quantum information theory and learning theory. Further details appear in the reference texts (Heinosaari and Ziman 2013; Nielsen and Chuang 2010).

2.1 Quantum information and computation

The most general descriptor of a d-level quantum system or statistical ensemble thereof is a density matrix, an element of

$$\mathscr{S}\left(\mathbb{C}^d\right) := \{\rho \in \mathbb{C}^{d \times d} \mid \rho \geq 0, \ \text{tr}[\rho] = 1\}.$$

Here, $\rho \geq 0$ means that the matrix ρ is Hermitian and all its eigenvalues are non-negative. An important subset of density matrices is the set of pure states, which are one-dimensional projections. Following Dirac notation, we denote the projector onto the subspace spanned by a unit vector $|\psi\rangle \in \mathbb{C}^d$ by $|\psi\rangle\langle\psi|$. By the spectral theorem, every quantum state can be written as a convex combination of pure states, though this decomposition is not unique in general.

Central to the framework of quantum mechanics is the measurement, the mechanism by which one may observe properties of a quantum system. These are typically described by so-called positive-operator valued measures (POVMs). As we focus on measurements with a finite set of outcomes $\{i\}$, it suffices to think of measurements as collections of so-called effect operators $\{E_i\}_{i=1}^m$ with

$E_i \in \mathbb{C}^{d \times d}, 0 \leq E_i \leq \mathbb{1}_d$, and $\sum_{i=1}^{m} E_i = \mathbb{1}_d$. We denote the set of effect operators by

$$\mathscr{E}\left(\mathbb{C}^d\right) := \{E \in \mathbb{C}^{d \times d} \mid 0 \leq E_i \leq \mathbb{1}_d\}.$$

Again, we highlight a special case: if we take an orthonormal basis $\{|\psi_i\rangle\}_{i=1}^{d}$ of $\mathbb{C}^d$, then the set $\{E_i = |\psi_i\rangle\langle\psi_i|\}_{i=1}^{d}$ is called a projective measurement.

Born's rule connects measurements to measurement outcomes: given a state characterized by a density operator, the effect operator has a corresponding probability $p_i = \mathrm{tr}[\rho E_i]$. Thus the requirement that the effect operators sum to the identity can be seen as probabilities summing to one. In the special case of pure state $\rho = |\psi\rangle\langle\psi|$ and projective measurement $\{E_i = |\psi_i\rangle\langle\psi_i|\}_{i=1}^{d}$, the probability of outcome i is $p_i = \mathrm{tr}[\rho E_i] = |\langle\psi|\psi_i\rangle|^2$.

So far we have described the components of static quantum theory. The dynamics of quantum states are described by so-called quantum operations, which we denote by

$$\mathscr{T}\left(\mathbb{C}^d\right) := \{ \qquad T : \mathbb{C}^{d \times d} \to \mathbb{C}^{d \times d} \mid T \text{ is linear,}$$
$$\text{completely positive, and trace-non-increasing}\}.$$

Here, a map T is completely positive if $T \otimes Id_n$ is positivity-preserving for every $n \in \mathbb{N}$. If $T \in \mathscr{T}\left(\mathbb{C}^d\right)$ is trace-preserving, we call T a quantum channel. An important example is the unitary quantum channel, $T(\rho) = U\rho U^*$ for some unitary $U \in \mathbb{C}^{d \times d}$.

Note that any element of $\mathscr{T}\left(\mathbb{C}^d\right)$ is a linear map between vector spaces of dimension d^2 and can thus be understood as a $d^2 \times d^2$ matrix.

2.2 Classical learning theory and complexity measures

Next we describe the "probably approximately correct" (PAC) model of learning, introduced and formalized by Vapnik and Chervonenkis (1971) and Valiant (1984). In (realizable) PAC learning for spaces X, Y and a concept class $\mathscr{F} \subseteq Y^X$, a learning algorithm receives as input labeled training data $\{(x_i, f(x_i))\}_{i=1}^{m}$ for some $f \in \mathscr{F}$, where the samples x_i are drawn independently according to some unknown probability distribution D on X that is unknown to the learner. Given the training examples, the goal of the learner is to approximate the unknown function f by a hypothesis function h, with high probability.

We can formalize this as follows: first, we introduce a loss function $\ell : Y \times Y \to \mathbb{R}_+$ to quantify the discrepancy between the hypothesis h and the function f. We call a concept class $\mathscr{F}$ PAC-learnable if there exists a learning algorithm $\mathscr{A}$ such that for every probability distribution D on X, $f \in \mathscr{F}$ and $\delta, \varepsilon \in (0, 1)$, running $\mathscr{A}$ on training data drawn according to D and f yields a hypothesis h

such that $\mathbb{E}_{x \sim D}[\ell(h(x), f(x))] \leq \varepsilon$ with probability $\geq 1 - \delta$ (with regard to the choice of training data). Moreover, we quantify the minimum amount of training data that an algorithm $\mathscr{A}$ needs to meet the above conditions by a map $m_{\mathscr{F}} : (0, 1) \times (0, 1) \to \mathbb{N}$, $(\delta, \varepsilon) \mapsto m(\delta, \varepsilon)$, the so-called sample complexity of $\mathscr{F}$. We focus on proper learning, in which the learning algorithm must output as its hypothesis an element of the concept class, i.e., we require $h \in \mathscr{F}$.

A standard approach to assessing learnability is to characterize the complexity of the respective concept class $\mathscr{F}$. Many such complexity measures are used, the most common being the VC-dimension for binary-valued function classes $\mathscr{F} \subseteq \{0, 1\}^X$, named after its progenitors (Vapnik and Chervonenkis 1971). This combinatorial parameter can be shown to fully characterize the learnability: a concept class $\mathscr{F} \subseteq \{0, 1\}^X$ is PAC-learnable (w.r.t. the 0-1-loss) if and only if the VC-dimension of $\mathscr{F}$ is finite. Moreover, the sample complexity of PAC learning $\mathscr{F}$ can be expressed in terms of its VC-dimension (see Blumer et al. 1989; Hanneke 2016).

In this work, we employ a widely used extension of the VC-dimension to real-valued concept classes:

Definition 1 *(Pseudo-dimension* (Pollard 1984)*)* Let $\mathscr{F} \subseteq \mathbb{R}^X$ be a real-valued concept class. A set $\{x_1, ..., x_k\} \subseteq X$ is pseudo-shattered by $\mathscr{F}$ if there are $y_1, ..., y_k \in \mathbb{R}$ such that for any $C \subseteq \{1, ..., k\}$ there is an $f_C \in \mathscr{F}$ such that for all $1 \leq i \leq k$, $i \in C$ if and only if $f_C(x_i) \geq y_i$.

The pseudo-dimension of $\mathscr{F}$ is defined to be

$$\mathrm{Pdim}(\mathscr{F}) := \sup\{n \in \mathbb{N}_0 \mid \quad \exists S \subseteq X \text{ s.t. } |S| = n \text{ and}$$
$$S \text{ is pseudo-shattered by } \mathscr{F}\}.$$

Alternatively, one can express the pseudo-dimension in terms of the VC-dimension. Namely,

$$\mathrm{Pdim}(\mathscr{F}) = \mathrm{VC}(\{X \times \mathbb{R} \ni (x, y) \mapsto \mathrm{sgn}(f(x) - y) \mid f \in \mathscr{F}\}).$$

Here, the VC-dimension for a function class $\mathscr{H} \subseteq \{\pm 1\}^Z$ is defined as

$$\mathrm{VC}(\mathscr{H}) := \sup\{n \in \mathbb{N}_0 \mid \exists z_1, \ldots, z_n \in Z \text{ s.t. } \forall b \in \{\pm 1\}^n$$
$$\exists h_b \in \mathscr{H} \text{ s.t. } \forall i : h_b(z_i) = b_i\}.$$

There is also a scale-sensitive version of the pseudo-dimension:

Definition 2 *(Fat-Shattering Dimension* (Alon et al. 1997)*)* Let $\mathscr{F}$ be a real-valued concept class and let $\alpha > 0$. A set $\{x_1, ..., x_k\} \subseteq X$ is α-fat-shattered by $\mathscr{F}$ if there are $y_1, ..., y_k \in \mathbb{R}$ such that for any $C \subseteq \{1, ..., k\}$ there is an $f_C \in \mathscr{F}$ such that for all $1 \leq i \leq k$:

1. $i \notin C \Rightarrow f_C(x_i) \leq y_i - \alpha$ and

2.　$i \in C \Rightarrow f_C(x_i) \geq y_i + \alpha$.

The α-fat-shattering dimension of $\mathscr{F}$ is defined to be

$$fat_{\mathscr{F}}(\alpha) := \sup\{n \in \mathbb{N}_0 | \quad \exists S \subseteq X \text{ s.t. } |S| = n \wedge \text{S}$$
$$\text{is } \alpha\text{-fat-shattered by } \mathscr{F}\}.$$

Note that, trivially, $fat_{\mathscr{F}}(\alpha) \leq \text{Pdim}(\mathscr{F})$ holds for every $\alpha > 0$ and for every real-valued function class $\mathscr{F}$.

Sample complexity upper bounds for $[0, 1]$-valued function classes in terms of the fat-shattering dimension have been proved in Bartlett and Long (1998) and Anthony and Bartlett (2000).

3 Pseudo-dimension bounds for quantum circuits

We now formulate how to characterize the expressive power of quantum circuits. In particular, we consider circuits with n input registers of qu*d*its, size (i.e., number of gates) γ, and depth (i.e., number of layers) δ. More precisely, we consider circuits composed of two-qudit unitaries, i.e., logical gates with two inputs. Note that two-qudit gates include one-qudit gates. We assume that gates in the same layer and acting on disjoint pairs of qudits can act in parallel. Additionally, we assume that each qu*d*it is acted upon by at least one gate, else it effectively does not participate in the circuit.

In this section, we assign function classes to quantum circuits and then derive bounds on the pseudo-dimension of these function classes, in terms of the number of qudits and the size and depth of the circuits. First, we fix quantum circuit structure and inputs, varying only the entries of the unitary gates and thereby the resulting function. Then, we broaden our scope to variable circuit architectures, variable inputs, and circuits whose "gates" are general quantum operations.

An important tool that will recur throughout our work is the following result on polynomial sign assignments, used in Goldberg and Jerrum (1995) to derive VC-dimension bounds from computational complexity.

Theorem 1 (Warren 1968, Theorem 3) *Let* $\{p_1, \ldots, p_m\}$ *be a set of real polynomials in n variables with $m \geq n$, each of degree at most $d \geq 1$. Then the number of consistent non-zero sign assignments to $\{p_1, \ldots, p_m\}$ is at most* $\left(\frac{4edm}{n}\right)^n$.

Here, e is Euler's number and a "consistent non-zero sign assignment" to a set of polynomials $\{p_1, \ldots, p_m\}$ is a vector $b \in \{\pm 1\}^m$ s.t. there exist $x_1, \ldots, x_n \in \mathbb{R}$ for which it holds that $\text{sgn}(p_i(x_1, \ldots, x_n)) = b_i$ for all $1 \leq i \leq m$.

The following implication of Theorem 1 for consistent but not necessarily non-zero sign assignments (which we define as above, but with $b \in \{-1, 0, 1\}^m$) to sets of polynomials was observed in Goldberg and Jerrum (1995, Corollary 2.1).

Corollary 1 *Let* $\{p_1, \ldots, p_m\}$ *be a set of real polynomials in n variables with $m \geq n$, each of degree at most $d \geq 1$. Then the number of consistent sign assignments to $\{p_1, \ldots, p_m\}$ is at most* $\left(\frac{8edm}{n}\right)^n$.

Proof (Sketch) This can be obtained by applying Theorem 1 to the set $\{p_1 + \varepsilon, p_1 - \varepsilon, \ldots, p_m + \varepsilon, p_m - \varepsilon\}$ with $\varepsilon > 0$ chosen sufficiently small. □

3.1 Fixed circuit structure

Suppose we fix the architecture of a quantum circuit of depth δ and size γ. Specifically, we restrict our attention to 2-local quantum circuits, i.e., circuits whose logical gates have support on two qudits, not necessarily neighboring each other (see Fig. 1). "Fixed architecture" means that we specify the positions of the two-qudit unitaries, namely their order and which qudits they act on. Though the unitaries' positions are fixed, we may vary the entries of the unitaries themselves. Here, we allow for arbitrary 2-qu*d*it unitaries. In particular, we do not restrict ourselves to a finite gate library. Can we bound the pseudo-dimension of the function class of measurement probability distributions that this circuit generates? And how does the bound depend on d (the dimensionality of the qudits), δ and γ?

To formalize this question: let $n \in \mathbb{N}$ be the number of qudits, $d \in \mathbb{N}$ be their dimensionality, and $\mathscr{N}$ be a fixed quantum circuit architecture of depth δ and size γ acting on n qudits. We enumerate the positions of the two-qudit unitaries in $\mathscr{N}$ by tuples (i, j) with $1 \leq i \leq \delta$ denoting
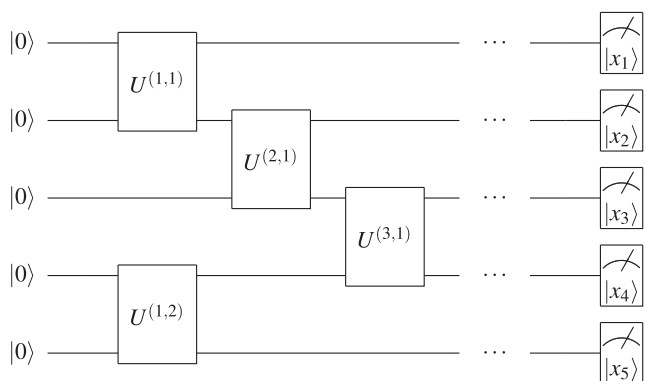
Fig. 1 An example 2-local circuit. $U^{(i,j)}$ denotes the jth 2-qudit unitary in the ith layer of the circuit

the layer and $1 \leq j \leq \gamma_i$ the position of the unitary among all the unitaries inside layer i, where w.l.o.g. we count from top to bottom and take into account only the first qudit on which a unitary acts.

Note that $\sum_{i=1}^{\delta} \gamma_i = \gamma$, and trivially $\gamma_i \leq \gamma$ and $\gamma_i \leq \frac{n}{2}$, as we assume that every qudit is acted upon by at least one gate. We write the unitary at position (i, j) as $U^{(i,j)}$. These constitute the "free parameters" which we can vary in order to make the quantum circuit perform different tasks. The overall unitary implemented by $\mathcal{N}$ when plugging in the unitaries $\{U^{(i,j)}\}_{1 \leq i \leq \delta, 1 \leq j \leq \gamma_i}$ at the respective positions we denote by $U_{\mathcal{N}|\{U^{(i,j)}\}}$. Note that $U_{\mathcal{N}|\{U^{(i,j)}\}}$ strongly depends on the two-qudit unitaries that are plugged into the architecture, but sometimes we will suppress this dependence and simply write $U_{\mathcal{N}}$ for notational ease.

The quantum circuit $\mathcal{N}$ now gives rise to the following set of output states:

$$\mathcal{S}_{\mathcal{N}}\left((\mathbb{C}^d)^{\otimes n}\right) := \left\{ U_{\mathcal{N}|\{U^{(i,j)}\}} |0\rangle^{\otimes n} \mid U^{(i,j)} \in \mathcal{U}\left((\mathbb{C}^d)^{\otimes 2}\right) \right\}.$$

These output states in turn give rise to a function class of measurement probability distributions with regard to product measurements:

$$\mathcal{F}_{\mathcal{N}} := \left\{ f : X \to [0, 1] \mid \exists |\psi\rangle \in \mathcal{S}_{\mathcal{N}}\left((\mathbb{C}^d)^{\otimes n}\right) : f(x) = |\langle x|\psi\rangle|^2 \right\},$$

where we take $X = S_d \times \ldots \times S_d$ to be the Cartesian product of n unit spheres of $\mathbb{C}^d$.

The main insight of this subsection is the following:

Theorem 2 *With the notation and assumptions from above, it holds that* $Pdim(\mathcal{F}_{\mathcal{N}}) \leq 8d^4 \cdot \gamma \cdot \log(16e \cdot \gamma)$.

Here and throughout the paper, log denotes the logarithm to base 2.

To prove this result, we provide the following.

Lemma 1 *With the notation and assumptions from above, there exists a polynomial $p_{\mathcal{N}}$ with real coefficients, in $2\gamma d^4 + 2dn$ real variables of degree $\leq 2(\gamma + n)$ such that every $f \in \mathcal{F}_{\mathcal{N}}$ can be obtained from $p_{\mathcal{N}}$ by fixing values for the first $2\gamma d^4$ variables. Moreover, in each term of p, the degree in the first $2\gamma d^4$ real variables is $\leq 2\gamma$ and the degree in the last $2dn$ real variables is $\leq 2n$.*

Notably, there is no explicit dependence on depth δ.

Proof We first observe that

$$|\langle x|U_{\mathcal{N}}|0\rangle^{\otimes n}|^2 = |\langle 0|^{\otimes n} U_{\mathcal{N}}^{\dagger}|x\rangle|^2.$$

We study this expression in a layer-wise analysis. When reading the circuit from right to left, the state that enters layer δ is transformed by the unitary $\bigotimes_{j=1}^{\gamma_\delta} U^{(\delta,j)\dagger}$ such that

each amplitude of the state after the δth layer is a linear combination of the amplitudes of $|x\rangle$, where each coefficient is a multilinear monomial of degree γ_δ in some of the $\gamma_\delta \cdot d^4$ complex entries of the $\{U^{(\delta,j)\dagger}\}_{1 \leq j \leq \gamma_\delta}$.

By iterating this reasoning, we see that the state after the $(\delta - i)$th layer has amplitudes which are given by a linear combination of the amplitudes of $|x\rangle$, where each coefficient is a multilinear polynomial of degree $\leq \sum_{k=0}^{i} \gamma_{\delta-k}$ in (some of) the entries of the unitaries $\{U^{(\delta-k,j_k)\dagger}\}_{0 \leq k \leq i, 1 \leq j_k \leq \gamma_k}$.

In particular, the $|0\rangle^{\otimes n}$-amplitude of the state $U_{\mathcal{N}}^{\dagger}|x\rangle$ can be written as a linear combination of the amplitudes of $|x\rangle$, where each coefficient is given by a multilinear polynomial $q_{\mathcal{N}}$ of degree $\leq \sum_{k=0}^{\delta} \gamma_{\delta-k} = \gamma$ in (some of) the $\gamma \cdot d^4$ complex entries of the unitaries $\{U^{(i,j_i)\dagger}\}_{0 \leq i \leq \delta, 1 \leq j_i \leq \gamma_i}$.

Recalling that the probability of observing outcome $|0\rangle^{\otimes n}$ is the square of the absolute value of the corresponding amplitude of $|x\rangle$, we obtain from the polynomial $q_{\mathcal{N}}$ a polynomial $p_{\mathcal{N}} = |q_{\mathcal{N}}|^2$ that describes the output probabilities. As $q_{\mathcal{N}}$ has degree at most γ in the $\gamma \cdot d^4$ complex parameters of the unitaries, $p_{\mathcal{N}}$ has degree at most 2γ in the corresponding $2\gamma \cdot d^4$ real parameters. Fixing these $2\gamma d^4$ parameters corresponds to fixing the circuit, and therefore one may obtain every $f \in \mathcal{F}_{\mathcal{N}}$ by fixing these parameters in $p_{\mathcal{N}}$.

Moreover, $p_{\mathcal{N}}$ is a polynomial in the $2dn$ real parameters which give rise to the amplitudes of $|x\rangle$. (Here, the assumption that $|x\rangle$ is a product state enters.) As each such amplitude has degree $\leq n$ in the $2dn$ complex parameters, the degree of $p_{\mathcal{N}}$ in these real parameters is at most $2n$. $\square$

Remark 1 We formulate the result only for measurement operators consisting of tensor products of 1-dimensional projections, and continue to do so throughout this manuscript. For $x \in X$, we can write $|x\rangle = \bigotimes_{i=1}^{n} \left(\sum_{j=0}^{d-1} \alpha_j^{(i)} |j\rangle \right)$, so we associate dn complex variables with x. That each amplitude of $|x\rangle$ can be written as a product of n complex parameters gives rise to the upper bound of n in the degree.

We could instead look at more general measurement operators consisting of 1-dimensional projections without requiring product structure, i.e., entangled measurements. In this scenario, we would write $|x\rangle = \sum_{z \in \{0,\ldots,d-1\}^n} x_z |z\rangle$, associating d^n complex variables with x. In this setup, each amplitude of x is simply a polynomial of degree 1 in these complex variables.

As we fix the variables corresponding to x and y in the shattering assumption that appears in our proof of Theorem 2, their corresponding degrees are not relevant to

our argument; only the degree in the entries of the unitaries enters our analysis. Therefore, both product measurements or entangled measurements lead to the same pseudo-dimension bound. This is due to the fact that allowing for entangled measurements changes the set of allowed inputs but not the function class itself.

Now that we have established Lemma 1, we can prove Theorem 2 with reasoning analogous to that in Goldberg and Jerrum (1995).

Proof (Theorem 2) Let $\{(x_i, y_i)\}_{i=1}^m \subseteq X \times \mathbb{R}$ be such that for every $C \subseteq \{1, \ldots, m\}$ there exists $f_C \in \mathscr{F}_{\mathscr{N}}$ such that $f_C(x_i) - y_i \geq 0$ if and only if $i \in C$.

By Lemma 1, there exists a polynomial $p_{\mathscr{N}}$ in $2\gamma d^4 + 2dn$ real variables of degree $\leq 2(\gamma + n)$ such that for every $C \subseteq \{1, \ldots, m\}$ there exists an assignment Ξ_C to the first $2\gamma d^4$ variables of $p_{\mathscr{N}}$ such that $p_{\mathscr{N}}(\Xi_C, x_i) - y_i \geq 0$ if and only if $i \in C$.

In particular, this implies (using the "moreover" part of Lemma 1) that the set $\mathscr{P} = \{p_{\mathscr{N}}(\cdot, x_i) - y_i\}_{i=1}^m$ is a set of m polynomials of degree $\leq 2\gamma$ in $2\gamma d^4$ real variables that has at least 2^m different consistent sign assignments.

We now claim that $m \leq 8d^4 \cdot \gamma \cdot \log(16e \cdot \gamma)$. If $m < 2\gamma d^4$, this holds trivially. Hence, w.l.o.g. $m \geq 2\gamma d^4$. So by Corollary 1, we have

$$2^m \leq \left(\frac{8e \cdot 2\gamma \cdot m}{2\gamma d^4} \right)^{2\gamma d^4}.$$

Taking logarithms now gives

$$m \leq 2\gamma d^4 \left(\log(16e \cdot \gamma) + \log\left(\frac{m}{2\gamma d^4} \right) \right).$$

Now we distinguish cases. If $16e \cdot \gamma \geq \frac{m}{2\gamma d^4}$, then the above immediately implies $m \leq 4\gamma d^4 \cdot \log(16e \cdot \gamma)$. If $16e \cdot \gamma \leq \frac{m}{2\gamma d^4}$, then we obtain $m \leq 4\gamma d^4 \cdot \log\left(\frac{m}{2\gamma d^4} \right)$, which in turn implies $m \leq 8\gamma d^4$. In both cases we have $m \leq 8d^4 \cdot \gamma \cdot \log(16e\gamma)$. By definition of the pseudo-dimension, we conclude $\mathrm{Pdim}(\mathscr{F}_{\mathscr{N}}) \leq 8d^4 \cdot \gamma \cdot \log(16e\gamma)$, as claimed. $\qquad\square$

The attentive reader may notice that we do not explicitly refer to the unitarity assumption in our reasoning; our argument mainly uses linearity. This already hints at a generalization to quantum circuits not of unitaries but of operations, which we will describe in Section 3.4. In that subsection, we will also see how the unitarity assumption implicit in this proof produces a better upper bound than in the general setting of quantum operations.

Remark 2 We formulate our bounds in terms of the pseudo-dimension, not its scale-sensitive version called fat-shattering dimension, even though the latter is more commonly used in classical learning. In our scenario, however, the pseudo-dimension and the fat-shattering dimension effectively coincide. This is because we could apply our reasoning for general matrices instead of only unitaries in the setting of Theorem 2 as well and achieve the same bounds. In that case, however, the resulting real-valued function class is closed under scalar multiplication with non-negative scalars and it follows from the definition that for such classes, the fat-shattering dimension equals the pseudo-dimension.

3.2 Variable circuit structure

Whereas in the previous subsection we fixed a quantum circuit architecture and only varied the entries of the two-qudit unitaries plugged into this structure, we now additionally vary the structure of the quantum circuit architecture itself and consider the complexity of the class of all quantum circuits of a given depth and size. Once again, we consider 2-local quantum circuits, i.e., circuits with one- and two-qudit gates acting on arbitrary pairs of qudits.

The class of states which is of relevance in this analysis is

$$\mathscr{S}_{\delta,\gamma}\left((\mathbb{C}^d)^{\otimes n} \right) := \{|\psi\rangle \mid \exists \text{ quantum circuit } \mathscr{N} \text{ of depth } \delta$$
$$\text{and size } \gamma \text{ such that } |\psi\rangle \in \mathscr{S}_{\mathscr{N}}\left((\mathbb{C}^d)^{\otimes n} \right)\}.$$

Again, this set of states gives rise to a function class via

$$\mathscr{F}_{\delta,\gamma} := \{f : X \to [0, 1] \mid \exists |\psi\rangle \in \mathscr{S}_{\delta,\gamma}\left((\mathbb{C}^d)^{\otimes n} \right) :$$
$$f(x) = |\langle x|\psi\rangle|^2\},$$

where X is as above given by $X = S_d \times \ldots \times S_d$. As before, we want to bound the pseudo-dimension of this function class.

We summarize the result of this subsection in the following:

Theorem 3 *With the notation and assumptions from above, it holds that $Pdim(\mathscr{F}_{\delta,\gamma}) \leq \mathcal{O}(\delta \cdot d^4 \cdot \gamma^2 \log \gamma)$.*

As with Theorem 2, the main step towards this result consists of relating the functions appearing in $\mathscr{F}_{\delta,\gamma}$ to polynomials. The difference here is that we must upper bound the number of polynomials, as below.

Lemma 2 *With the notation and assumptions from above, there exists a set $\mathscr{P}_{\delta,\gamma}$ of polynomials with real coefficients, in $2\gamma d^4 + 2dn$ real variables of degree $\leq 2(\gamma + n)$ such that for every $f \in \mathscr{F}_{\delta,\gamma}$ there exists a polynomial $p \in \mathscr{P}_{\delta,\gamma}$*

such that f can be obtained from p by fixing values for the first $2\gamma d^4$ variables, and such that

$$|\mathscr{P}_{\delta,\gamma}| \leq \frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!}(n!)^\delta.$$

Moreover, in each term of $p \in \mathscr{P}_{\delta,\gamma}$ the degree in the first $2\gamma d^4$ real variables is $\leq 2\gamma$ and the degree in the last $2dn$ real variables is $\leq 2n$.

Proof There are at most $\frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!}$ ways to assign them among the δ layers. The term $\frac{\gamma!}{(\gamma-\delta)!}$ counts assigning a single gate to each layer, to ensure that there are no trivial (empty) layers. Having assigned each layer one gate, the remaining $\gamma - \delta$ gates may be distributed to any of the δ layers.

Next, we bound the number of ways of assigning qudits to the circuit layers, so that the qudits are inputs to the fixed-position unitaries. For our purposes, it suffices to crudely upper bound this by $n!$ for each single layer and thus by $(n!)^\delta$ overall. Hence, there are at most

$$\frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!}(n!)^\delta$$

different quantum circuit architectures. The proof is completed by applying Lemma 1 to every such quantum circuit architecture. □

Now that we have established Lemma 2, we can prove Theorem 3 by reasoning analogous to that in Goldberg and Jerrum (1995) (see the Appendix for the proof of Theorem 3).

3.3 Extension to circuits with variable inputs

We now modify the results of Sections 3.1 and 3.2 to allow not only for the fixed input $|0\rangle^{\otimes n}$, but also for variable input. This is of use, for instance, in Section 4.2, in which we consider the PAC-learnability of quantum circuits (of unitary gates or more general quantum channels). In that context, allowing variable input amounts to learning the entire quantum circuit, rather than just its action on $|0\rangle^{\otimes n}$. This is necessary in order to meaningfully compare the learning problem in Section 4.2 to exact circuit tomography.

To consider variable input states, we define the following function classes, analogously to those in Sections 3.1 and 3.2:

$$\mathscr{F}'_{\mathcal{N}} := \{ \qquad f : X \times Y \to [0, 1] \mid \exists U_{\mathcal{N}\mid\{U^{(i,j)}\}},$$
$$U^{(i,j)} \in \mathscr{U}\left((\mathbb{C}^d)^{\otimes 2}\right) : f(x, y) = |\langle x|U_{\mathcal{N}}|y\rangle|^2\},$$

where Y can be taken as the computational basis states $\{0, 1, ..., d-1\}^n$, or more generally as $Y = X = S_d \times ... \times S_d$.

Lemma 3 *With the notation and assumptions from above the following holds: There exists a polynomial $p'_{\mathcal{N}}$ in $2\gamma d^4 + 4dn$ real variables of degree $\leq 2\gamma + 4n$ such that every $f \in \mathscr{F}'_{\mathcal{N}}$ can be obtained from $p'_{\mathcal{N}}$ by fixing values for the first $2\gamma d^4$ variables. Moreover, in each term of $p'_{\mathcal{N}}$ the degree in the first $2\gamma d^4$ real variables is $\leq 2\gamma$, the degree in the $2dn$ real variables corresponding to $x \in X$ is $\leq 2n$, and the degree in the $2dn$ real variables corresponding to $y \in Y$ is $\leq 2n$.*

Proof Consider the product state input $|y\rangle = \sum_z y_z |z\rangle$. As we consider product states, each y_z is a product of n complex parameters. Following the same reasoning as before, for a fixed $z \in \{0, \ldots, d-1\}$, $\langle z|U_{\mathcal{N}}|x\rangle$ is a multilinear polynomial $q^z_{\mathcal{N}}$. Then, the amplitude $\langle y|U_{\mathcal{N}}|x\rangle$ is

$$q'_{\mathcal{N}}(x, y) = \langle y|U_{\mathcal{N}}|x\rangle = \sum_{z\in\{0,1,...,d-1\}^n} \overline{y_z} \, \langle z|U_{\mathcal{N}}|x\rangle$$
$$= \sum_{z\in\{0,1,...,d-1\}^n} \overline{y_z} \, q^z_{\mathcal{N}}(x).$$

In the above equation, $q'_{\mathcal{N}}(x, y)$ has degree at most n in y, and so upon squaring the amplitude $q'_{\mathcal{N}}(x, y)$ to obtain $p'_{\mathcal{N}}(x, y)$ as in Lemma 1, we have a degree at most $2n$ in the $2dn$ real variables corresponding to y. The rest follows from Lemma 1. □

The bound from Theorem 2 still holds for the case of variable circuit input, with the proof proceeding almost identically upon replacing Lemma 1 by Lemma 3. The $2d \cdot n$ additional variables that arise from the polynomial y-dependence do not alter the bound because we fix the values of these variables in the pseudo-shattering assumption.

3.4 Extension to circuits of quantum operations

We finish this section by describing an extension of Theorems 2 and 3 to the case of circuits of quantum operations, instead of only unitaries. This generalization is relatively straightforward because the decisive property of unitaries used in our previous proofs was not the preservation of inner products, but rather linearity. This setting is useful to, e.g., describe circuits with imperfect gates. Rather than consider a logical gate that implements a unitary exactly, each gate can instead be considered a quantum operation that executes the desired unitary with some probability, and, e.g., depolarizes input qudits with some probability. (Other noise models are of course possible.) Note that although quantum operations can, by Stinespring's dilation theorem, be viewed as subsystem dynamics of a larger, unitarily evolving system, if we only

have access to measurement data for the subsystem then we cannot directly apply our result for the unitary case.

We use analogous notation to that introduced at the beginning of Section 3.1, writing $T_{\mathcal{N}|\{T^{(i,j)}\}}$ for the overall quantum operation implemented by $\mathcal{N}$ when plugging the two-qudit quantum operations $\{T^{(i,j)}\}_{1 \le i \le \delta, 1 \le j \le \gamma_i}$ into the respective positions of the quantum circuit.

The quantum circuit $\mathcal{N}$ (of operations) now gives rise to the set of output states

$$\mathcal{D}_{\mathcal{N}}\left((\mathbb{C}^d)^{\otimes n}\right) := \{T_{\mathcal{N}|\{T^{(i,j)}\}}(|0^n\rangle\langle 0^n|) \mid$$
$$T^{(i,j)} \in \mathcal{T}\left((\mathbb{C}^d)^{\otimes 2}\right)\},$$

where we write $|0^n\rangle = |0\rangle^{\otimes n}$, so $|0^n\rangle\langle 0^n| = (|0\rangle\langle 0|)^{\otimes n}$.

By taking into account all possible quantum circuits of size γ and depth δ, we obtain

$$\mathcal{D}_{\delta,\gamma}\left((\mathbb{C}^d)^{\otimes n}\right) := \{\rho \mid \exists \text{ circuit } \mathcal{N} \text{ of two-qudit operations}$$
$$\text{of size } \gamma \text{ and depth } \delta \text{ such that } \rho \in \mathcal{D}_{\mathcal{N}}\left((\mathbb{C}^d)^{\otimes n}\right)\}.$$

These states now yield again a p-concept class

$$\mathcal{G}_{\delta,\gamma} := \{f : X \to [0,1] \mid \exists \rho \in \mathcal{D}_{\delta,\gamma}\left((\mathbb{C}^d)^{\otimes n}\right) : f(x) = \langle x|\rho|x\rangle\}.$$

In this scenario, we show:

Theorem 4 *With the notation and assumptions from above, it holds that $Pdim(\mathcal{G}_{\delta,\gamma}) \le \mathcal{O}(\delta \cdot d^8 \cdot \gamma^2 \log \gamma)$.*

Proof We only sketch the reasoning, as it is similar to that in the proof of Theorem 3. We first need to establish an analogue of Lemma 2. To this end, observe that a quantum operation acting on two-qudit states can be interpreted as a $d^4 \times d^4$ matrix with complex entries. Moreover, we may write

$$\langle x|T_{\mathcal{N}}(|0^n\rangle\langle 0^n|)|x\rangle = \text{tr}[T_{\mathcal{N}}(|0^n\rangle\langle 0^n|)|x\rangle\langle x|]$$
$$= \text{tr}[|0^n\rangle\langle 0^n|T_{\mathcal{N}}^*(|x\rangle\langle x|)]$$
$$= \langle 0^n|T_{\mathcal{N}}^*(|x\rangle\langle x|)|0^n\rangle,$$

where $T_{\mathcal{N}}^*$ denotes the adjoint operation of $T_{\mathcal{N}}$ with regard to the Hilbert-Schmidt inner product.

As before, we can do a layer-wise analysis of the transformation of $|x\rangle\langle x|$ and observe that the entries of the (sub-normalized) density matrix after a layer can be written as linear combinations of the entries of the (sub-normalized) density matrix before the layer. Moreover, the coefficients can be written as multilinear polynomials with the degree determined by the number of two-qudit operations in the layer. Hence, we obtain the result of Lemma 1 with d^8 instead of d^4. The bound on the number of different quantum circuit architectures can be derived in exactly the same way as before, so the analogue of Lemma 2 holds, completing the proof of the theorem. $\qquad\square$

Theorem 4 and its proof sketch also help to elucidate the relevance of the unitarity assumption in Theorems 2 and 3. Unitarity justifies our restriction to pure states, but in other respects Theorems 2 and 3 do not exploit unitary. The difference between Theorems 3 and 4 amounts to the size of the matrices that represent the unitaries or quantum operations.

4 Applications

In this section, we explore two different applications of our pseudo-dimension upper bounds. First, we employ the pseudo-dimension to exhibit a large but finite discrete set of quantum states, out of which at least one is hard to implement in the sense that preparing it requires exponentially many 2-qubit unitaries. Second, we combine the pseudo-dimension bound with results from the theory of p-concept learning to derive the PAC-learnability of quantum circuits.

4.1 Lower bounds on the gate complexity of quantum state preparation

It is well known that almost all n-qubit unitaries require an exponential (in n) number of 2-qubit unitaries to be implemented. Similarly, almost all pure n-qubit states require an application of exponentially (in n) many 2-qubit unitaries to be generated from the $|0\rangle^{\otimes n}$ state (see, e.g., Nielsen and Chuang 2010). However, in neither case are there explicit examples of unitaries or states saturating this exponentiality bound (see Aaronson 2016 for more information on the gate complexity of unitary implementation and state preparation). We will use the pseudo-dimension as a tool to exhibit a discrete set of pure qubit states such that at least one of them requires exponentially many 2-qubit unitaries to be generated from $|0\rangle^{\otimes n}$.

The drawback of our result is that the size of this set is 2^{2^n} and thus unsatisfyingly large. By relatively simple deliberations this size can be reduced by an order of 2^n elements, though this is negligible compared to the overall size.

We now describe the construction of the candidate set of states. For a subset $C \subseteq \{|x0\rangle\}_{x \in \{0,1\}^n}$, namely a subset of the set of all computational basis states of $n + 1$ qubits that end on 0, with $C \ne \emptyset$, define

$$|\psi_C\rangle = \frac{1}{\sqrt{|C|}} \sum_{x0 \in C} |x0\rangle.$$

For $C = \emptyset$ we take

$$|\psi_\emptyset\rangle = |0\rangle^{\otimes n} \otimes |1\rangle.$$

(Note that the $(n + 1)^{st}$ qubit only really matters for $|\psi_\emptyset\rangle$.)
Our set of interest will be

$$\mathscr{S} := \{|\psi_C\rangle \mid C \subseteq \{|x0\rangle\}_{x\in\{0,1\}^n}\}.$$

This discrete set of 2^{2^n} multi-qubit quantum states now
gives rise to a class of p-concepts

$$\mathscr{F}_{\mathscr{S}} = \{f_C : X \to [0, 1] \mid \exists C \subseteq \{|x0\rangle\}_{x\in\{0,1\}^n} :$$
$$f_C(x) = |\langle x|\psi_C\rangle|^2\}.$$

This class has large pseudo-dimension, as described in
the following lemma.

Lemma 4 *With the notation introduced above, it holds that*
$Pdim(\mathscr{F}_{\mathscr{S}}) \geq 2^n$.

Proof Consider the subset of computational basis states
$\{|x0\rangle\}_{x\in\{0,1\}^n}$ and the corresponding threshold values
$y_{x0} = \frac{1}{2^n} = \min\limits_{C\subseteq\{|x0\rangle\}_{x\in\{0,1\}^n}} \frac{1}{|C|}$ independently of $x0$. By
construction of $\mathscr{S}$ and thus $\mathscr{F}_{\mathscr{S}}$ the following holds:

For any $C \subseteq \{|x0\rangle\}_{x\in\{0,1\}^n}$

$$f_C(x0) = |\langle x0|\psi_C\rangle|^2 = \begin{cases} \frac{1}{|C|} & \text{if } |x0\rangle \in C \\ 0 & \text{else} \end{cases}.$$

In particular, we have

$$f_C(x0) \geq y_{x0} \iff |x0\rangle \in C.$$

Hence, $\mathrm{Pdim}(\mathscr{F}_{\mathscr{S}}) \geq 2^n$, because we have found an
example of a set of size 2^n that is pseudo-shattered. $\square$

We now combine this simple observation with Theo-
rem 3, which gives us the following:

Theorem 5 *With the notation introduced above, if γ and
δ are such that each state in $\mathscr{S}$ can be generated from the
state $|0\rangle^{\otimes(n+1)}$ by some circuit of size γ and depth δ, then*

$$2^n \leq \mathscr{O}\left(\delta \cdot 2^4 \cdot \gamma^2 \log \gamma\right)$$

Proof Under the assumption of the Theorem we can
conclude $\mathscr{F}_{\mathscr{S}} \subseteq \mathscr{F}_{\delta,\gamma}$. Now combine the lower bound of
Lemma 4 with the upper bound from Theorem 3. $\square$

Corollary 2 *There exists a $C \subseteq \{|x0\rangle\}_{x\in\{0,1\}^n}$ such that
$|\psi_C\rangle = \frac{1}{\sqrt{|C|}} \sum\limits_{|x0\rangle\in C} |x0\rangle$ cannot be implemented by a
quantum circuit of 2-qubit unitaries with subexponential (in
n) size or depth.*

Note that any set of functions which pseudo-shatters a set
of size 2^n has to have at least 2^{2^n} elements. Hence, the large
size of the set C is an automatic consequence of our line of
reasoning.

Remark 3 We note that a set of n-qubit states with
cardinality doubly exponential in n s.t. at least one of them
needs an exponential number of gates (up to logarithmic
factors) to be implemented can also be obtained with more
standard reasoning. Namely, it is well known that there
are n-qubit states the approximation of which up to trace-
distance ε requires $\Omega\left(\frac{2^n \log\left(\frac{1}{\varepsilon}\right)}{\log(n)}\right)$ unitary gates (see Nielsen
and Chuang [2010], chap. 4.5.4). So if we pick a $\frac{1}{2}$-net of size
$\mathscr{O}\left(2^{2^n}\right)$ for the set of pure n-qubit quantum states, this will
have the desired properties.

We sketch another way of using our pseudo-dimension
bound to study the gate complexity of state preparation
and which might lead to a smaller set of candidates.
Given n-qubit pure states $|\psi_1\rangle, \ldots, |\psi_m\rangle$ and efficiently
implementable (i.e., with polynomially many 2-qubit
unitary gates arranged in polynomially many layers)
unitaries $U_1, \ldots, U_k$, one can study the set of states
$\{U_i|\psi_j\rangle\}_{1\leq i\leq k, 1\leq j\leq m}$.

If an exponential (in n) pseudo-dimension lower bound
can be established for

$$\{f : X \to [0, 1] \mid \exists 1 \leq i \leq k, 1 \leq j \leq m : f(x) = |\langle x|U_i|\psi_j\rangle|^2\},$$

then, since every U_i is efficiently implementable, one can
conclude that at least one among the states $|\psi_j\rangle$ is not
efficiently implementable.

The advantage of such a pseudo-dimension-based rea-
soning would be that m need not be doubly exponential
in n, since we can compensate for this in k. This realiza-
tion can already be used to reduce the size of the set of
candidate states given in Corollary 2. However, we have
not yet been able to identify sufficiently many efficiently
implementable unitaries to reduce the size below doubly
exponential. Nevertheless, there is likely room for improve-
ment in applying our method to the gate complexity of
quantum state preparation.

4.2 Learnability of quantum circuits

We now use our pseudo-dimension bounds to study
learnability. Specifically, we use the pseudo-dimension
bound for the case of variable inputs (Section 3.3) combined
with the generalization to quantum operations (Section 3.4).
We proceed quite similarly to Aaronson ([2007]).

The learning problem which we want to study is the fol-
lowing: Let μ be a probability measure on $(X \times Y) \times [0, 1]$,
unknown to the learner. Let $S = \{((x^{(i)}, y^{(i)}), p^{(i)})\}_{i=1}^m$ be
corresponding training data drawn i.i.d. according to μ. A
learner must, upon input of training data S, size $\Gamma \in \mathbb{N}$,
depth $\Delta \in \mathbb{N}$, confidence $\delta \in [0, 1]$, accuracy, $\varepsilon \in [0, 1]$
and error margin $\beta \in (0, 1)$, output a hypothesis quantum
circuit $\mathscr{N}$ of size Γ and depth Δ consisting of two-qudit

operations such that, with probability $\geq 1 - \delta$ with regard to the choice of training data,

$$\mathbb{P}_{((x,y),p)\sim\mu}\left[|f_{\mathcal{N}}(x,y) - p| > \beta\right] \leq \varepsilon$$
$$+ \inf_{\mathcal{M}} \mathbb{P}_{(x,p)\sim\mu}\left[|f_{\mathcal{M}}(x,y) - p| > \beta\right],$$

where the infimum runs over all quantum circuits $\mathcal{M}$ of size Γ and depth Δ. Here, $f_{\mathcal{N}}$ denotes the function $f_{\mathcal{N}}(x,y) = \langle x|T_{\mathcal{N}}(|y\rangle\langle y|)|x\rangle$ and $f_{\mathcal{M}}$ is defined analogously, similarly to Section 3.3.

We use our pseudo-dimension bound in order to upper bound the size of the training data sufficient for solving this task. More precisely, we make use of sample complexity upper bounds from the fat-shattering dimension as proved in Anthony and Bartlett (2000) and Bartlett and Long (1998), together with the fact that the fat-shattering dimension is upper-bounded by the pseudo-dimension.

First we restrict our scope to the "realizable" scenario, i.e., we will assume the probability measure to be of the form

$$\mu((x,y),p) = \begin{cases} \mu_1(x,y) & \text{if } p = f_{\mathcal{N}_*}(x,y) \\ 0 & \text{else} \end{cases}$$

for some quantum circuit $\mathcal{N}_*$ of size Γ and depth Δ. This will in particular imply that for quantum circuits $\mathcal{M}$ of size Γ and depth Δ

$$\inf_{\mathcal{M}} \mathbb{P}_{((x,y),p)\sim\mu}\left[|f_{\mathcal{M}}(x,y) - p| > \beta\right] = 0.$$

Colloquially, realizability means that there exists a set of "correct" parameters Γ and Δ and these are known to the learner, i.e., training samples are promised to be drawn from circuits of size Γ and depth Δ.

We will focus on a proper learning scenario, i.e., we will assume the unknown target circuit to be in some (known) class, namely the class of circuits whose size and depth satisfy certain polynomial bounds, and require the learner to output an element of that same class as hypothesis.

We will make use of the following classical result:

Theorem 6 (Anthony and Bartlett 2000, Corollary 3.3) *Let X be an input space, let $\mathcal{F} \subseteq [0,1]^X$. Let D be a probability measure on X, let $f_* \in \mathcal{F}$. Let $\delta, \varepsilon, \alpha, \beta \in (0,1)$ with $\beta > \alpha$. Let $\mathcal{S} = \{x_1, \ldots, x_m\}$ be a set of m samples drawn i.i.d. according to D. Let $h \in \mathcal{F}$ be such that $|h(x_i) - f_*(x_i)| \leq \alpha$ for all $1 \leq i \leq m$.*

Then, a sample size

$$m = \mathcal{O}\left(\frac{1}{\varepsilon}\left(\text{fat}_{\mathcal{F}}\left(\frac{\beta-\alpha}{8}\right)\log^2\left(\frac{\text{fat}_{\mathcal{F}}\left(\frac{\beta-\alpha}{8}\right)}{(\beta-\alpha)\varepsilon}\right) + \log\frac{1}{\delta}\right)\right)$$

suffices to guarantee that, with probability $\geq 1 - \delta$ with regard to the choice of training data $\mathcal{S}$,

$$\mathbb{P}_{x\sim D}[|h(x) - f_*(x)| > \beta] \leq \varepsilon.$$

In our setting, this result implies:

Corollary 3 *Let $\mathcal{N}_*$ be a quantum circuit of quantum operations with size Γ and depth Δ. Let μ be probability measure on $X \times Y$ unknown to the learner. Let*

$$S = \{((x^{(i)}, y^{(i)}), f_{\mathcal{N}_*}(x^{(i)}, y^{(i)}))\}_{i=1}^m$$

be corresponding training data drawn i.i.d. according to μ. Let $\delta, \varepsilon, \alpha, \beta \in (0,1)$. Then, training data of size $m = \mathcal{O}\left(\frac{1}{\varepsilon}\left(\Delta d^8 \Gamma^2 \log(\Gamma)\log^2\left(\frac{\Delta d^8 \Gamma^2 \log(\Gamma)}{(\beta-\alpha)\varepsilon}\right) + \log\frac{1}{\delta}\right)\right)$ suffice to guarantee that, with probability $\geq 1 - \delta$ with regard to choice of the training data, any quantum circuit $\mathcal{N}$ of size Γ and depth Δ that satisfies

$$|f_{\mathcal{N}}(x_i, y_i) - f_{\mathcal{N}_*}(x_i, y_i)| \leq \alpha \quad \forall 1 \leq i \leq m$$

also satisfies

$$\mathbb{P}_{(x,y)\sim\mu}[|f_{\mathcal{N}}(x,y) - f_{\mathcal{N}_*}(x,y)| > \beta] \leq \varepsilon.$$

Proof Combine Theorem 6 with Theorem 3 (more precisely, with its version for variable input states, which can be proved for operations analogously to the reasoning in Section 3.3) and use that the fat-shattering dimension is always upper-bounded by the pseudo-dimension. $\qquad\square$

Note that in particular, this implies that for the class of circuits of quantum operations with polynomial size and depth in the number of qudits, a hypothesis that performs well on training data will also perform well in a probably approximately correct sense.

Next, we want to discuss briefly how our result compares to the work (Aaronson 2007) on the learnability of quantum states. There, it is shown that quantum states can be PAC-learned with a sample complexity that depends linearly on the number of qubits and (among other dependencies) polynomially on $\frac{1}{\varepsilon}$, where ε denotes the desired accuracy. However, this result does not imply learnability of quantum channels with a sample complexity that depends polynomially on the number of qubits. This observation is already stated in Aaronson (2007), and we provide an alternate, intuitive explanation for why the result on states does not directly apply to operations.

One can straightforwardly apply the result of Aaronson (2007) to learn the Choi-Jamiolkowski state of a quantum channel. One can then compute measurement probabilities of output states of a channel T acting on n-qubit states, using its Choi-Jamiolkowski state τ. For this we must make use of the formula

$$\text{tr}[ET(\rho)] = 2^n \text{tr}[\tau(E \otimes \rho^T)].$$

Here, we see that any error on the side of the Choi-Jamiolkowski state will be multiplied by a factor exponential in n, and thus in this case the overall n-dependence of the sample complexity bound from Aaronson (2007) becomes exponential via the accuracy-dependence.

This motivates our study of learnability of a restricted class of quantum operations. Finding such operations for which process tomography is possible was left as an open problem in Aaronson (2007). Our answer to this question is that a PAC-version of quantum process tomography is possible when we restrict our scope to operations that can be implemented by quantum circuits of depth and size polynomial in the number of qudits. However, note that this is subject to a realizability assumption: the learner must known in advance a polynomial bound on the size and depth of the circuit. We show that imposing the operations be efficiently implementable automatically reduces the information-theoretic complexity of learning, requiring only a modest number of training examples. We do not make any statement about the computational complexity of this learning task; this remains an open problem.

How can this probably approximately correct version of quantum process tomography be put to use? Given polynomially many uses of a black box implementing an unknown quantum operation of polynomial size and depth, one can exhibit a circuit of two-qudit quantum operations that approximates the unknown channel. In other words, we obtain a classical description of an approximate copy of the channel.

5 Open problems

Finally, in this section we discuss future directions and possible generalizations of our results.

Two natural parameters of a circuit, depth and size, appear polynomially in the pseudo-dimension upper bounds. Notably, these bounds are independent of the number of qudits in the circuit. Are our upper bounds tight in their dependence on size and depth? Can similar techniques produce pseudo-dimension lower bounds? For example, by considering a single 2-qudit unitary it is relatively straightforward to see that the pseudo-dimension of a circuit is $\geq \Omega(d)$. Can we close the gap in dimension-dependence between this linear lower bound and our quartic upper bound?

Our application of pseudo-dimension for lower bounds on the gate complexity of state preparation complements known methods (described, e.g., in Nielsen and Chuang 2010), based on counting dimensions or covering arguments. We exhibit a class of states of size 2^{2^n}, for which at least one has exponential gate complexity of state prepara-

tion. Can we exploit this new technique to exhibit a smaller set of states? Perhaps the most exciting application of pseudo-dimension bounds could be provable lower bounds on the gate complexity of state preparation, if the reasoning in Section 4.1 is sharpened or the tools are developed further.

If circuit depth and size are known in advance, one can information-efficiently learn the circuit. If the learner receives training data generated by an approximation of the circuit, does the result still hold? Can the realizability assumption be relaxed?

Does "pretty-good circuit tomography" have applications? On the theory side, this might involve exploiting the learning process as an approximate copy-machine for quantum circuits. Of interest for both theory and experiment is whether circuits can be learned with a reasonable amount of computation. One can imagine progress on this question for process tomography similar to that for state tomography; demonstrating a class of states for which learning is computationally efficient in Rocchetto (2017) made it possible to learn physically interesting states in a laboratory in Rocchetto et al. (2019). An efficiency improvement in the process tomography case might also have experimental ramifications.

Appendix

Here, we prove Theorem 3, namely that $\mathrm{Pdim}(\mathscr{F}_{\delta,\gamma}) \leq \mathscr{O}(\delta \cdot d^4 \cdot \gamma^2 \log \gamma)$.

Proof (Theorem 3)

We rely upon Lemma 2. Let $\{(x_i, y_i)\}_{i=1}^m \subseteq X \times \mathbb{R}$ be such that for every $C \subseteq \{1, \ldots, m\}$, there exists $f_C \in \mathscr{F}_{\delta,\gamma}$ such that $f_C(x_i) - y_i \geq 0$ if and only if $i \in C$.

By Lemma 2, there exists a set of polynomials $\mathscr{P}_{\delta,\gamma}$ in $2\gamma d^4 + 2d^n$ real variables such that $|\mathscr{P}_{\delta,\gamma}| \leq \frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!} (n!)^\delta$ and such that for every $C \subseteq \{1, \ldots, m\}$, there exists a $p_C \in \mathscr{P}_{\delta,\gamma}$ and an assignment Ξ_C to the first $2\gamma d^4$ variables of p_C such that $p_C(\Xi_C, x_i) - y_i \geq 0$ if and only if $i \in C$.

In particular, this implies (using the "moreover"-part of Lemma 2) that the set $\mathscr{P} = \{p(\cdot, x_i) - y_i\}_{i=1}^m \mid p \in \mathscr{P}_{\delta,\gamma}\}$ is a set of $m \cdot |\mathscr{P}_{\delta,\gamma}| \leq m \frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!} (n!)^\delta$ polynomials of degree $\leq 2\gamma$ in $2\gamma d^4$ real variables that has at least 2^m different consistent sign assignments. So by Corollary 1, we have

$$2^m \leq \left(\frac{8e \cdot 2\gamma \cdot m}{2\gamma d^4} \cdot \frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!} (n!)^\delta \right)^{2\gamma d^4}.$$

Taking logarithms yields

$$m \leq 2\gamma d^4 \left(\log(16e \cdot \gamma) + \log \left(\frac{m}{2\gamma d^4} \cdot \frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!} (n!)^\delta \right) \right).$$

Repeating the argument in the proof of Theorem 2, we distinguish cases and observe that in both cases,

$$m \leq 8d^4 \cdot \gamma \cdot \log \left(16e\gamma \cdot \frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!} (n!)^\delta \right).$$

Expanding the logarithm and using Stirling's formula up to two terms, we have

$$
\log \left(16e\gamma \cdot \frac{\gamma! \, \delta^{\gamma-\delta}}{(\gamma-\delta)!} (n!)^\delta \right)
$$
$$
= \frac{1}{\ln 2} \left(4\ln 2 + 1 + \ln \gamma + [n \cdot \delta \ln n - n \cdot \delta + \mathscr{O}(\ln n) \right.
$$
$$
+ \gamma \ln \gamma - \gamma + \mathscr{O}(\ln \gamma) - (\gamma - \delta) \ln(\gamma - \delta)
$$
$$
\left. + (\gamma - \delta) + \mathscr{O}(\ln(\gamma - \delta)) + (\gamma - \delta) \ln \delta] \right)
$$
$$
\leq \frac{1}{\ln 2} \left(4\ln 2 + 1 + \ln \gamma + [2\gamma \cdot \delta(\ln(2\gamma) - 1) \right.
$$
$$
\left. + \gamma \ln \gamma - (\gamma - \delta) \ln(\gamma - \delta) - \delta + (\gamma - \delta) \ln \delta] \right)
$$
$$
= \mathscr{O}(\gamma \cdot \delta \log \gamma).
$$

We use the fact that $n \leq 2\gamma$ (because we assume that each qudit is acted upon by at least one gate) in the second step, and note that because $\gamma \geq \delta$, the asymptotic behavior

of all of the above terms are subsumed by the first term in the bracket. We have also confirmed that the $\log(16e\gamma)$ term above may be neglected. Thus, by the definition of the pseudo-dimension we conclude $\mathrm{Pdim}(\mathscr{F}_\mathcal{N}) \leq \mathscr{O}(\delta \cdot d^4 \cdot \gamma^2 \log \gamma)$. $\qquad\square$

References

Aaronson S (2007) The learnability of quantum states. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 463(2088):3089–3114. https://doi.org/10.1098/rspa.2007.0113

Aaronson S (2016) The complexity of quantum states and transformations: From quantum money to black holes. Electronic Colloquium on Computational Complexity (ECCC) 23:109

Aaronson S (2018) Shadow tomography of quantum states. http://dl.acm.org/ft_gateway.cfm?id=3188802&type=pdf

Aaronson S, Chen X, Hazan E, Kale S (2018) Online learning of quantum states. http://dl.acm.org/ft_gateway.cfm?id=3327572&type=pdf

Alon N, Ben-David S, Cesa-Bianchi N, Haussler D (1997) Scale-sensitive dimensions, uniform convergence, and learnability. J ACM 44(4):615–631. https://doi.org/10.1145/263867.263927

Anthony M, Bartlett PL (2000) Function learning from interpolation. Comb Probab Comput 9(3):213–225. https://doi.org/10.1017/S0963548300004247

Arunachalam S, de Wolf R (2017) Guest column: A survey of quantum learning theory. SIGACT News 48 , https://pure.uva.nl/ws/files/25255496/p41_arunachalam.pdf

Bartlett PL, Long PM (1998) Prediction, learning, uniform convergence, and scale-sensitive dimensions. J Comput Sys Sci 56(2):174–190. https://doi.org/10.1006/jcss.1997.1557

Bartlett PL, Mendelson S (2002) Rademacher and gaussian complexities: Risk bounds and structural results. J Mach Learn Res 3(Nov):463–482. http://www.jmlr.org/papers/volume3/bartlett02a/bartlett02a.pdf

Blumer A, Ehrenfeucht A, Haussler D, Warmuth MK (1989) Learnability and the vapnik-chervonenkis dimension. J ACM 36(4):929–965. https://doi.org/10.1145/76359.76371

Cheng HC, Hsieh MH, Yeh PC (2016) The learnability of unknown quantum measurements. Quantum Information & Computation 16(7-8):615–656

Chung KM, Lin HH (2018) Sample efficient algorithms for learning quantum channels in pac model and the approximate state discrimination problem. arXiv:1810.10938

Goldberg PW, Jerrum MR (1995) Bounding the vapnik-chervonenkis dimension of concept classes parameterized by real numbers. Mach Learn 18(2-3):131–148. https://doi.org/10.1007/BF00993408

Hanneke S (2016) The optimal sample complexity of pac learning. J Mach Learn Res 17(1):1319–1333. http://dl.acm.org/ft_gateway.cfm?id=2946683&type=pdf

Heinosaari T, Ziman M (2013) The mathematical language of quantum theory: From uncertainty to entanglement. Cambridge University Press, Cambridge

Karpinski M, Macintyre A (1997) Polynomial bounds for vc dimension of sigmoidal and general pfaffian neural networks. J Comput Sys Sci 54(1):169–176. https://doi.org/10.1006/jcss.1997.1477

Kiani BT, Lloyd S, Maity R (2020) Learning unitaries by gradient descent. arXiv:2001.11897

Koiran P (1996) VC dimension in circuit complexity. In: Cai JY, Homer S (eds) Proceedings, Eleventh annual ieee conference on

computational complexity. IEEE Computer Society Press, Los Alamitos, pp 81–85. https://doi.org/10.1109/CCC.1996.507671

Nielsen MA, Chuang IL (2010) Quantum computation and quantum information. Cambridge University Press, Cambridge and New York

Pollard D (1984) Convergence of stochastic processes. Springer Series in Statistics. Springer, New York. https://doi.org/10.1007/978-1-4612-5254-2

Rocchetto A (2017) Stabiliser states are efficiently pac-learnable. Quantum Information and Computation, 18

Rocchetto A, Aaronson S, Severini S, Carvacho G, Poderini D, Agresti I, Bentivegna M, Sciarrino F (2019) Experimental learning of quantum states. Science Advances 5(3), https://doi.org/10.1126/sciadv.aau1946

Torlai G, Mazzola G, Carrasquilla J, Troyer M, Melko R, Carleo G (2018) Neural-network quantum state tomography. Nat Phy 14(5):447–450. https://doi.org/10.1038/s41567-018-0048-5, https://www.nature.com/articles/s41567-018-0048-5.pdf

Valiant LG (1984) A theory of the learnable. Commun ACM 27(11):1134–1142. https://doi.org/10.1145/1968.1972

Vapnik VN, Chervonenkis AY (1971) On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability & Its Applications 16(2):264–280. https://doi.org/10.1137/1116025

Warren HE (1968) Lower bounds for approximation by nonlinear manifolds. Trans Am Math Soc 133(1):167. https://doi.org/10.2307/1994937