Learning Low-Degree Quantum Objects

Srinivasan Arunachalam
IBM Quantum, Thomas J Watson Research Center, Yorktown Heights, NY, USA

Arkopal Dutt
IBM Quantum, IBM Research Cambridge, MA, USA

Francisco Escudero Gutiérrez
CWI & QuSoft, Amsterdam, The Netherlands

Carlos Palazuelos
Dpto. Análisis Matemático y Matemática Aplicada, Fac. Ciencias Matemáticas,
Universidad Complutense de Madrid, Spain
Instituto de Ciencias Matemáticas, Madrid, Spain

Abstract

We consider the problem of learning low-degree quantum objects up to ε-error in ℓ_2-distance. We show the following results: (i) unknown n-qubit degree-d (in the Pauli basis) quantum channels and unitaries can be learned using $O(1/\varepsilon^d)$ queries (which is independent of n), (ii) polynomials $p : \{-1,1\}^n \to [-1,1]$ arising from d-query quantum algorithms can be learned from $O((1/\varepsilon)^d \cdot \log n)$ many random examples $(x, p(x))$ (which implies learnability even for $d = O(\log n)$), and (iii) degree-d polynomials $p : \{-1,1\}^n \to [-1,1]$ can be learned through $O(1/\varepsilon^d)$ queries to a quantum unitary U_p that block-encodes p. Our main technical contributions are new Bohnenblust-Hille inequalities for quantum channels and completely bounded polynomials.

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases Tomography

Digital Object Identifier 10.4230/LIPIcs.ICALP.2024.13

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: https://arxiv.org/abs/2405.10933 [6]

Funding This research was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 945045, and by the NWO Gravitation project NETWORKS under grant no. 024.002.003.

Acknowledgements We thank the referees of ICALP'24 and TQC'24 for useful comments. We thank Jop Briët and Antonio Pérez Hernández for useful discussions. We thank Alexandros Eskenazis, Paata Ivanisvili, Alexander Volberg and Haonan Zhang for useful comments. A.D. and S.A. thank the Institute for Pure and Applied Mathematics (IPAM) for its hospitality throughout the long program "Mathematical and Computational Challenges in Quantum Computing" in Fall 2023 during which part of this work was initiated.

1 Introduction

Computational learning theory refers to the mathematical framework for understanding machine learning models and quantifying their complexity. The seminal result of Leslie Valiant [54] (who introduced the Probably Approximately Correct (PAC) model) gives a complexity-theoretic definition of what it means for a class of functions $f : \{0,1\}^n \to \{0,1\}$ to be learnable information-theoretically and computationally.

One of the foundational results in computational learning theory is the one of Linial, Mansour and Nisan [39] who showed that AC_0, or constant-depth n-bit classical circuits consisting of AND, OR and NOT gates, can be learned in quasi-polynomial time. A crucial aspect of their proof is the following structural theorem: if a Boolean function f is computable by AC_0 then f can be approximated by a *low-degree polynomial*. Using this structural property, the learning algorithm approximates the coefficients of all the low-degree monomials of f in the PAC model and hence approximately learns the unknown AC_0 function. Since their work, the notion of Boolean functions being low-degree or being well-approximated by low-degree polynomials has been a central technique [16] in obtaining new learning algorithms. Furthermore, low-degree approximations have played a significant role in theoretical computer science topics such as quantum computing, circuit complexity, learning theory and cryptography.

In the last few years, there have been several works in quantum learning theory where the goal has been to learn an unknown object on a quantum computer under various access models. Motivated by classical learning theory, in this work our primary focus will be on learning objects that have the additional structure of being *low-degree*. Since we consider different objects, when presenting our results we will make the definition of being low-degree clear, but the main motivation of this work can be summarized by the following question:

Can we learn low-degree n-qubit quantum objects information-theoretically with complexity that scales only polynomial (or better polylogarithmic) in n?

We give a positive answer to this question for low-degree channels, unitaries, quantum query algorithms, polynomials, and states. The organization of the paper is as follows. In Section 2 we present our main technical contribution, then in Section 3 we describe our applications to learning, in Section 4 we prove the result for channels, and in Section 5 we prove the result for quantum query algorithms. Due to page restrictions, we defer the preliminaries of our paper as well as the remaining proofs to the extended version [6].

2 Technical contribution: New Bohnenblust-Hille inequalities

In 1931, Bohnenblust and Hille [13] (generalizing the classic theorem of Littlewood [40]) gave a solution to the famous *Bohr strip problem* of Dirichlet series [14]. To do that, they showed the following: let $T : ([-1,1]^n)^d \to [-1,1]$ be a d-tensor specifed by the coefficients $T = (\widehat{T}_{i_1,\ldots,i_d})_{i_1,\ldots,i_d\in[n]}$, then

$$\Big(\sum_{i_1,\ldots,i_d=1}^{n} |\widehat{T}_{i_1,\ldots,i_d}|^{2d/(d+1)}\Big)^{(d+1)/2d} \leq C(d), \tag{1}$$

where $C(d)$ is a universal constant independent of n.[1] Their work marked the birth of the Bohnenblust-Hille (BH) inequality, which has became a fundamental tool in functional analysis. Despite being studied over a century, the best known upper bound on $C(d)$ scales polynomially with d, while the best lower bound is a constant. Closing this gap has been an active area of research in mathematics. In 2011 Defant et. al [22] refined the BH inequality and found a striking application of the BH inequality: they determined the precise asymptotic behavior of the n-dimensional Bohr radius using the BH inequality. Since then,

[1] We remark that the inequality above is a simplified version of the original Bohnenblust-Hille inequality and we discuss this in more detail in the preliminaries.

there has been renewed interest in the BH inequality and has found several applications in theoretical computer science such as Fourier-Entropy influence conjecture [4], classical learning theory [26], non-local games [42], and quantum computing [33, 57].

Of particular relevance to our work, the BH inequality recently captured the attention of the computer science community when Eskenazis and Ivanisvili [26] used a version of it to prove a major improvement in the problem of classical learning bounded low-degree multilinear polynomials (which we discuss in detail below). The key insight of Eskenazis and Ivanisvili is that BH inequalities imply that for bounded operators the small coefficients have a low total contribution, so one does not have to learn them. Multiple extensions of the BH have been proved since then and applied to learning quantum objects [33, 50, 51, 57, 36]. In this work, we extend the BH inequality in two ways:

1. We consider a variant of the BH inequality, that can be regarded as a hybrid between the BH inequality and the celebrated Grothendieck inequality [30]. We show that degree-d completely bounded tensors $\widehat{T}$ (which are known to be the the output of d-query quantum algorithms [3]) satisfy

$$\Big(\sum_{i_1,\ldots,i_d=1}^{n} |\widehat{T}_{i_1,\ldots,i_d}|^{2d/d+1}\Big)^{(d+1)/2d} \leq 1.$$

   In other words, we improve the BH constant for completely bounded tensors from $\mathrm{poly}(d)$ to 1. See Theorem 15 for a precise statement.
2. More recently, the works of [33, 57] considered non-commutative variants of the BH inequality. They showed that the Pauli coefficients of n-qubit degree-d (i.e., d-local) observables that are bounded in operator norm, can be bounded in a similar fashion to Eq. (1), but with $\exp(d)$ instead of $\mathrm{poly}(d)$. Here, we prove another non-commutative version of the BH inequality for quantum *channels* (we in fact prove a stronger BH inequality for *maps* that are bounded in $S_1 \to S_1$ norm, and refer the reader to Theorem 11). In particular, if Φ is a quantum channel defined as $\Phi(\rho) := \sum_{x,y\in\{0,1,2,3\}^n} \widehat{\Phi}(x,y)\sigma_x\rho\sigma_y$ (where $\sigma_x = \otimes_i \sigma_{x_i}$ and $\sigma_0 = I, \sigma_1 = X, \sigma_2 = Y, \sigma_3 = Z$ are the usual single-qubit Pauli operators) where $\widehat{\Phi}(x,y) = 0$ if $|x|, |y| > d$, then the Pauli coefficients $\widehat{\Phi}(x,y)$ satisfy

$$\Big(\sum_{x,y} |\widehat{\Phi}(x,y)|^{2d/(d+1/2)}\Big)^{(d+1/2)/2d} \leq \exp(d). \tag{2}$$

   While $\exp(d)$ in Equation (2) might seem much higher than the factor $\mathrm{poly}(d)$ of Equation (1), this is not a fair comparison. Equation (1) corresponds to tensors, which are a very structured class of polynomials, while Equation (2) is a non-commutative analogue of the BH inequality for general polynomials, for which the best known upper bounds [23] are superpolynomial in d.

These BH inequalities might be of independent interest both for mathematicians and quantum computing; in this work we crucially use them for our learning algorithms.

3 Applications and main results

3.1 Result 1: Learning channels

Learning a quantum process is a fundamental task in quantum computing and this can be modelled as learning an unknown *quantum channel*, also referred to as *quantum process tomography.* On an experimental level, the dynamics of closed quantum systems can be modeled as a unitary transformation from the initial state to the final state. However, in practice, quantum systems interact with the environment and must be treated as an open quantum system. To learn the behavior of these open quantum systems, it is convenient to model this map as a quantum channel [45]. Learning an n-qubit quantum channel is however challenging and is known to require $\Theta(4^n)$ queries to the channel [31]. This exponential sample complexity can be drastically improved when prior information on the structure of the channel is available. For example, a recent work of Bao and Yao [10] considered k-junta quantum channels, i.e., n-qubit channels that act non-trivially only on at most k of the n (unknown) qubits leaving rest of qubits unchanged. These channels were shown to be learnable using $\widetilde{\Theta}(4^k)$ queries to the channel [10]. In [21], it was shown that quantum channels that can be efficiently generated, can be learned efficiently, albeit in the PAC learning model. In this work, we consider learning n-qubit quantum channels which have a Pauli decomposition only involving low-degree Pauli operators.

General quantum channels. To describe our result, we first describe Pauli analysis for quantum channels. An n-qubit to n qubit quantum channel Φ can be expressed as

$$\Phi(\rho) = \sum_{x,y\in\{0,1,2,3\}^n} \widehat{\Phi}(x,y)\cdot \sigma_x\rho\sigma_y, \tag{3}$$

where $\sigma_x = \otimes_{i\in[n]}\sigma_{x_i}$ and σ_i for $i \in \{0,1,2,3\}$ are the Pauli matrices

$$\sigma_0 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\ \sigma_1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\ \sigma_2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\ \sigma_3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix};$$

and $\widehat{\Phi}(x,y)$ are the Pauli coefficients of the channel. Given $x \in \{0,1,2,3\}^n$, $|x|$ is the number of non-zero entries of x. The degree of a channel Φ is the minimum integer d such that $\widehat{\Phi}(x,y) = 0$ if $|x| > d$ or $|y| > d$. Our first result is an efficient learning algorithm for low-degree channels. The learning model we consider is the same as the recent work of Bao and Yao [10]. Given a channel Φ, a learning algorithm is allowed to make *queries to* Φ as follows: it can apply Φ to an arbitrary ρ (or subsystem of ρ) of its choice and measure the resulting state in in any basis. From the measurement outcomes, the learner should output a classical description of a superoperator $\widetilde{\Phi}$ that is close to Φ in the ℓ_2-distance defined by the usual inner product for superoperators, i.e., $\langle \Phi, \widetilde{\Phi}\rangle = \mathrm{Tr}[J(\Phi), J(\widetilde{\Phi})]/4^n$, where $J(\Phi)$ is the Choi-Jaminkowski (CJ) representation of Φ.[2]

▶ **Theorem 1.** *Let Φ be a n-qubit degree-d quantum channel. There is an algorithm that (ε,δ)-learns Φ (in ℓ_2-distance) using* $\exp\left(\widetilde{O}(d^2 + d\log(1/\varepsilon))\right)\cdot\log(1/\delta)$ *queries to Φ.*

We remark that the sample complexity of learning general quantum channels requires $\Omega(4^n)$ queries, but if we are promised the channel is *low-degree*, then our algorithm is much faster than the general algorithm. Additionally, observe that the sample complexity of our learning algorithm is independent of n, in contrast to the results [33, 57, 50, 51, 36] on quantum learning observables (which also are based on the BH inequality) that have a logarithmic dependence on n.

To prove the theorem, we first note the fact that the matrix $\widehat{\Phi}$ (whose entries are given by $\widehat{\Phi}_{x,y} = \widehat{\Phi}(x,y)$) is the density matrix of a state that is unitarily equivalent to the CJ state of the channel Φ [10, Lemma 8]. Hence, $\widehat{\Phi}$ can be prepared by applying Φ to the first n-qubits of n EPR pairs defined over $2n$ qubits, and $\{\widehat{\Phi}(x,x)\}_x$ is a probability distribution. The high-level idea behind the learning algorithm is the following.

---

[2] In this paper we say an algorithm (ε,δ)-learns a quantum object if it succeeds with probability $\geq 1-\delta$ and outputs an ε-approximation to the unknown quantum object in a metric that will be made clear.

1. Prepare T few copies of $\widehat{\Phi}$ and measure them in the computational basis, allowing the learner to sample from the distribution $\{\widehat{\Phi}(x,x)\}_x$.
2. Using a well-known result in distribution learning theory, we observe that $O(1/\alpha^2)$ many samples from $\{\widehat{\Phi}(x,x)\}_x$ suffices to obtain the xs such that $\Phi(x,x)$ is α-large.
3. Approximate all the large Pauli coefficients in the step above using a SWAP test for mixed states.
4. Output $\widetilde{\Phi}$ with coefficients that were estimated above and the remaining coefficients set to 0.

At this point, we use the BH inequality for quantum channels and show that, as long as $T \sim \exp(d^2/\varepsilon^d)$, then the output $\widetilde{\Phi}$ is ε-close to the target Φ in ℓ_2-distance.

Remark on learning Pauli channels. One can also consider a special case of quantum channels called *Pauli channels* Φ, motivated by the fact that Pauli channels are often the dominant noise on quantum devices and a practical noise model for analyzing fault-tolerance [52]. For these channels, the only non-zero terms in the Fourier expansion (3) are $\widehat{\Phi}(x,x)$ (i.e., $\widehat{\Phi}(x,y) = 0$ when $x \neq y$) and these coefficients are often called *error rates*. Since learning Pauli channels is an important task on near-term quantum devices for error mitigation [55] and analyzing error correction, it is desirable to avoid using entangled copies of a state and access to ancillary qubits as part of the learning algorithm. With this requirement, it was shown that, in order to ε-learn (in diamond norm) an unknown n-qubit Pauli channels using unentangled measurements, one needs to use the channel $\Omega(4^n/\varepsilon^2)$ many times [28]. In this work, we show that the subclass of low-degree Pauli channels are efficiently learnable.

Noise on current large-scale quantum devices is modeled as Pauli channels containing a sparse set of local Paulis [55], which is a subclass of low-degree Pauli channels. Such models are available from device physics and experiments used to characterize noise on quantum hardware. When non-local interactions but only over few qubits are included in the Pauli channel [53], the corresponding Paulis are still low-degree which fits into the class of Pauli channels considered. Our learning result is as follows.

▶ **Fact 2.** *Let Φ be an n-qubit degree-d Pauli channel. There is an algorithm that (ε,δ)-learns Φ (under the diamond norm) using $O\left(n^{2d}/\varepsilon^2 \cdot \log(n/\delta)\right)$ queries to Φ. The learning algorithm only requires preparation of product states and measurements in the Pauli basis.*

We remark that the dependence on n in our sample complexity matches the algorithm in [29] for learning low-degree Pauli channels under the diamond norm but our analysis differs from theirs; our result is obtained using a Fourier-analytic approach in contrast to the the result of [29] which uses ideas from population recovery.

3.2 Result 2: Learning unitaries

Apart from learning channels, in this paper we also consider the task of learning unknown n-qubit unitaries. Similar to the case for channels, it is well-known that learning an unstructured n-qubit unitary requires $\widetilde{\Theta}(4^n)$ applications of U [31]. This complexity can be significantly improved if structural information is available. For example, it has been show efficient learning is possible if the unitary corresponds to Clifford circuits [41], corresponds to Clifford circuits with few non-Clifford gates [38], or are the diagonal unitaries of the Clifford hierarchy [2].

In a recent work, Chen et al. [20] considered the task of learning unitaries U that are k-juntas and showed that this class can be learned querying the unitary $\widetilde{\Theta}(4^k)$ many times. In this work, we consider the scenario where the unitary is degree-d and show such an exponential saving (in comparison to naive tomography) is possible. This structured class of low-degree unitaries occur in many instances. For example, in nature, the dynamics of many physical systems are governed by local Hamiltonians, whose unitary time evolution operators for short time evolution are close to being low-degree unitaries. Moreover, through the application of Lieb-Robinson bounds to structured many-body Hamiltonians, the corresponding unitary evolution operator can be seen to only have support only on low-degree Paulis [19]. In addition, the quantum unitaries corresponding to quantum circuits producing degree-d phase states containing all commuting diagonal quantum gates [2], are low-degree. The learning algorithm that we will present is thus applicable for learning and verifying such circuits.

Consider the Pauli decomposition of an n-qubit unitary as follows:

$$U = \sum_{x\in\{0,1,2,3\}^n} \widehat{U}(x)\sigma_x,$$

where $\widehat{U}(x)$ are its Pauli coefficients. The degree of U is the minimum integer d such that $|x|> d$ implies $\widehat{U}(x) = 0$. Our learning model is the one by Chen et al. [20]. Given an unitary U, a learning algorithm is allowed to make *queries to* U (and to control-U) as follows: it can choose a state ρ, apply U to the state to obtain $U\rho U^*$, and measure $U\rho U^*$ in a chosen basis. From the measurement outcomes, the learner should output a classical description of an operator $\widetilde{U}$ that is close to U in the ℓ_2-distance determined by the usual inner product for operators defined as $\langle U, V\rangle = \mathrm{Tr}[U^*V]/2^n$.

▶ **Theorem 3.** *Let U be a n-qubit degree-d unitary. There is an algorithm that (ε, δ)-learns U (in ℓ_2-distance) using the unitary U $\exp\big(\widetilde{O}(d^2 + d\log(1/\varepsilon))\big)\cdot\log(1/\delta)$ many times.*

The proof of Theorem 3 follows the same structure as that of Theorem 1, but now we learn the Pauli coefficients via an extension of the algorithm of Montanaro and Osborne [43], and the control on the contribution of the small coefficients relies on the non-commutative BH inequality of Volberg and Zhang [57].

The BH inequality of Volberg and Zhang [57] works for matrices with bounded operator norm, of which unitaries are a very special case. As argued by Montanaro and Osborne [43], matrices bounded on the operator norm are the quantum analogue of bounded functions $f : \{-1, 1\}^n \to [-1, 1]$, while unitaries are the analogue of Boolean functions[3] $f : \{-1, 1\}^n \to \{-1, 1\}$. These two families of classical functions differ a lot with respect to the BH inequalities: the best upper bound [23] for the BH constant for bounded functions is $\exp(d^{1/2})$, while for Boolean functions of degree d one can even prove that the *usually much bigger* quantity $\sum_s |\widehat{f}(s)|$ is at most 2^{d-1} [46, Exercise 1.11]. An open question is if this fact can be generalized to the quantum setting.

▶ **Question 4.** *Is there a constant $C(d)$ such that $\sum_{x\in\{0,1,2,3\}^n} |\widehat{U}(x)|\leq C(d)$ for every n-qubit degree-d unitary U?*

If Question 4 was answered positively, then one could improve the ε dependence of Theorem 3 to $(1/\varepsilon)^2$. Some evidence in favor of an affirmative answer to Question 4 is that if a conjecture of Montanaro and Osborne was true [43, Conjecture 4], then every degree-d hermitian unitary would be a 2^d-junta, which would imply an affirmative answer to Question 4 for the hermitian case.

Remark on learning low-degree quantum states. Learning quantum states has been an active line of research given its fundamental importance and applications to quantum system characterization, assessing quality of quantum gates, verification of quantum circuits and

[3] To be precise, they argue that unitary Hermitian matrices are the analogue of Boolean functions.

validating performance of quantum algorithms. Breakthrough results of Haah et al. [32] and O'Donnell and Wright [47] showed that the sample complexity of learning an unknown n-qubit state, up to trace distance ε is $\Theta(4^n/\varepsilon^2)$. A natural consideration is the task of learning *low-degree* quantum states. To describe this, we first write down the Pauli expansion of an n-qubit state ρ as

$$\rho = \sum_{x\in\{0,1,2,3\}^n} \widehat{\rho}(x)\sigma_x.$$

Then, we say that ρ has degree at most d if $\widehat{\rho}(x) = 0$ for all $|x| > d$. It is not too hard to see that one can use the formalism of classical shadows [34] to obtain a learning (in trace norm) algorithm that has a sample complexity of $\widetilde{O}(n^d/\varepsilon^2 \cdot \log(n/\delta))$. A similar result (with a different norm and with a bit more structure than just being low-degree) was noted in a recent work of Nadimpalli et al. [44], where they used the result to give efficient algorithms to learn QAC^0 circuits.

3.3 Result 3: Learning quantum query algorithms

Eskenazis and Ivanisvili [26] established a surprising connection between the BH inequality and learning theory. They considered the following question: suppose $f : \{-1,1\}^n \to [-1,1]$ is a bounded degree-d function, and a learner is given uniformly random x and $f(x)$, then how many $(x, f(x))$ suffices to learn f up to error ε in ℓ_2^2 error? The seminal low-degree algorithm of Linial, Mansour and Nisan uses[4] $O_{d,\varepsilon}(n^d)$ many such samples [39]. This was not improved until recently, when Iyer et al. [35] reduced this complexity to $O_{d,\varepsilon}(n^{d-1})$. In a surprising work, [26] showed that one can learn f in sample complexity $O_{d,\varepsilon}(\log n)$. For the particular case of bounded d-linear tensors $T : (\{-1,1\}^n)^d \to [-1,1]$ they showed that it suffices to use

$$(1/\varepsilon)^d \cdot \Big(\sum_{i_1,\dots,i_d=1}^{n} |\widehat{T}_{i_1,\dots,i_d}|^{2d/d+1} \Big)^{(d+1)/2d} \cdot \log n \tag{4}$$

samples $(x, T(x))$, where x is uniform from $(\{-1,1\}^n)^d$ and $\widehat{T}$ is the tensor of coefficients of T, i.e., $T(x) = \sum_{i_1,\dots,i_d} \widehat{T}_{i_1,\dots,i_d} x_1(i_1)\dots x_d(i_d)$. Combining that with the upper bound of the BH constant for multilinear tensors [11], it yields

$$(d/\varepsilon)^{O(d)} \cdot O(\log n) \tag{5}$$

uniformly random samples are enough to learn T. Although this result is surprising since the complexity only scales polylogarithmic with n, observe that if $d = \omega(\log n)$, then the sample complexity is superpolynomial in n, motivating the natural question, are there classes of polynomials that can be learned using $\mathrm{poly}(n)$ samples for any $d = \omega(\log n)$? Below we show that the class of polynomials that arise from quantum query algorithms answers this question in the positive.

Quantum polynomials. The result of Equation (5) can be applied to learn the amplitudes of quantum algorithms that query different blocks of variables every time (see Figure 1), as they are multilinear tensors bounded on the supremum norm [12]. To be precise, we consider quantum query algorithms such that they prepare a state

$$|\psi_x\rangle = U_d(O_{x_d} \otimes \mathrm{Id}_m)U_{d-1}\dots U_1(O_{x_1} \otimes \mathrm{Id}_m)U_0|u\rangle,$$

[4] Here and below, we use $O_{d,\varepsilon}$ to hide the factors that depend on $d, 1/\varepsilon$ and independent of n.

where m is an integer, x stands for $(x_1,\dots,x_d)$, O_y is the n-dimensional unitary that maps $|i\rangle$ to $y_i|i\rangle$, $U_0,\dots,U_d$ are $(n+m)$-dimensional unitaries and $|u\rangle$ is a $n+m$-dimensional unit vector. The algorithm succeeds according to a projective measurement that measures the projection of the final state onto some fixed $n+m$ dimensional unit vector $|v\rangle$. Hence, the amplitude of $|v\rangle$ is $T(x)=\langle v|\Psi_x\rangle$, so $|T(x)|^2$ is the acceptance probability of the algorithm. These quantum algorithms have been considered in the quantum computing literature. For

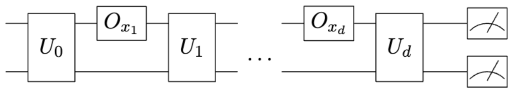

Figure 1 Quantum query algorithms considered in Theorem 5.

example, k-forrelation, that witnesses the biggest possible quantum-classical separation, has this structure [1, 8]. Also, for these algorithms the Aaronson and Ambainis conjecture is known to be true, so they can be classically and efficiently simulated almost everywhere [9, 25]. In addition, Arunachalam et al. showed that those amplitudes are not only bounded, but also completely bounded [3]. Our main contribution regarding these algorithms is showing that for d-linear tensors T that are *completely bounded*, we can improve the BH inequality to

$$\left(\sum_{i_1,\dots,i_d=1}^{n}|\widehat{T}_{i_1,\dots,i_d}|^{2d/d+1}\right)^{(d+1)/2d}\leq 1.$$

Using this upper bound, we show the following.

▶ **Theorem 5.** *For a quantum algorithm that makes d-queries as in Figure 1, its amplitudes can be learned up to error ε in ℓ_2^2 accuracy using $O\left((1/\varepsilon)^d\cdot\log n\right)$ uniformly random samples.*

This exponentially improves the complexity of [26], as stated in Eq. (5) for the natural class of polynomials arising from quantum query algorithms. In particular, for $d=\omega(\log n)$ and contant ε, one can learn this class of polynomials with sample complexity that is polynomial in n.

3.4 Result 4: Quantum learning of classical polynomials

3.4.1 Boolean functions

Quantum learning not only concerns quantum objects, but also classical ones. For instance, Boolean functions $f:\{-1,1\}^n\to\{-1,1\}$ can be accessed using a *quantum example*, given by

$$|\psi_f\rangle=\frac{1}{\sqrt{2^n}}\sum_{x\in\{-1,1\}^n}|x,f(x)\rangle.$$

This data access model has been vastly studied in the literature, where several notable quantum speedups have been proven [15, 7, 5]. Many of these speedups are analyzed trough the Fourier transform, that allows to identify every function via $f:\{-1,1\}^n\to\mathbb{R}$ with a multilinear polynomial $f=\sum_{s\in\{0,1\}^n}\widehat{f}(s)\chi_s$, where $\widehat{f}(s)\in\mathbb{R}$ are the Fourier coefficients and χ_s are the character functions $\chi_s(x)=\prod_{i\in[n]}x_i^{s_i}$. The degree for these functions is the minimum integer d such that if $|s|>d$ then $\widehat{f}(s)=0$. It is a well-known fact that Boolean functions $f:\{-1,1\}^n\to\{-1,1\}$ of degree d are 2^{1-d}-granular, meaning that their Fourier coefficients lie in $2^{1-d}\mathbb{Z}$. This has immediate consequences for both learning theory and BH inequalities.

▶ **Fact 6.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a degree-d function. There is a quantum algorithm that $(0,\delta)$ learns f using $O\left(4^d d \log(1/\delta)\right)$ quantum examples. However, a classical algorithm can learn it using $O\left(4^d d \log(n/\delta)\right)$ uniform examples and requires $\Omega(2^d + \log n)$ examples.*

Despite the simplicity of the proof of Fact 6, we state it for completeness and because it seems not to be well-known (see for instance [44, Corollary 34], which proposes a quantum algorithm for the same problem that requires $O(n^d)$ samples, or [27, Corollary 4] that proposes a classical algorthm that requires $O(2^{d^2} \log n)$ samples). Furthermore, we observe a BH-type inequality for Boolean functions.

▶ **Fact 7.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ of degree at most d. Then,*

$$\left(\sum_{s \in \{0,1\}^n} |\widehat{f}(s)|^{\frac{2d}{d+1}} \right)^{\frac{d+1}{2d}} \leq 2^{\frac{d-1}{d}}.$$

The equality it is witnessed by the address function.

Fact 7 might be of interest in functional analysis for two reasons: (i) it is conjectured that the value of the BH constant for d-linear tensors is $2^{\frac{d-1}{d}}$ [48], so this fact proves the conjecture for particular case of d-linear Boolean tensors, (ii) the address function[5], that saturates the inequality, is a d-linear form that gives a lower bound for the BH constant for multilinear tensors of $2^{\frac{d}{d-1}}$, which matches the best lower bound known so far [24].

3.4.2 Real-valued polynomials

For bounded functions $f : \{-1,1\}^n \to [-1,1]$ the definition of quantum uniform examples $|\psi_f\rangle$ is unclear. Given that bounded polynomials have received attention in a few works [39, 35, 26], we propose a way learning them quantumly by accessing these polynomials through a block encoding [18], i.e., a learning algorithm has access to a block encoding of the 2^n-dimensional diagonal matrix whose diagonal entries all equal f. To this end, we prove the following theorem.

▶ **Proposition 8.** *Let $f : \{-1,1\}^n \to [-1,1]$ be a degree-d polynomial. There is an algorithm that (ε,δ)-learns f (in ℓ_2-distance) using $\exp(\widetilde{O}(\sqrt{d^3} + d\log(1/\varepsilon))\log(1/\delta))$ copies of a block-encoding of f.*

The proof of Proposition 8 is a combination of the ideas of Eskenazis and Ivanisvili [26] with the Fourier sampling and block-encoding quantum primitives. We remind the reader that the merit of Eskenazis and Ivanisvili [26] was to bring down the classical complexity of the problem from $O_{d,\varepsilon}(n^d)$ to $O_{d,\varepsilon}(\log n)$. Proposition 8 shows that the quantum complexity this could be even reduced to $O_{d,\varepsilon}(1)$. Proposition 8 also implies a quantum speedup (with respect to n), as the lower bound of $\Omega(2^d + \log n)$ also holds for membership queries[6], which are the classical analogue of accessing a unitary block-encoding of f.

[5] For $d \in \mathbb{N}$, the address function $f : (\{-1,1\}^2)^{d-1} \times \{-1,1\}^{2^{d-1}} \to \{-1,1\}$ is defined as $f(x,y) = \sum_{a \in \{-1,1\}^{d-1}} g_a(x)y(a)$, where we identify $\{-1,1\}^{d-1}$ with $[2^{d-1}]$ and $g_a(x)$ is 0 unless $x_i(1) = a_i x_i(2)$ for every $i \in [n]$, in which case it takes the value $\prod_{i\in[n]} x_i(1)$.

[6] Making a membership query to f consists on accessing one pair $(x, f(x))$ where x is chosen by the learner, not necessarily uniformly at random. If f is a Boolean function and U_f is the unitary defined by $U_f|x\rangle = f(x)|x\rangle$, then a membership query $(x, f(x))$ can be simulated by applying $(H \otimes \mathrm{Id}_n)CU_f(H \otimes \mathrm{Id}_n)$ to $|0\rangle|x\rangle$ and measuring the first qubit in the computational basis. Note that U_f can be regarded as a block-encoding of f.

Remarkably, we highlight that although our results are about sample complexity, the time complexity of our quantum algorithm of Proposition 8 scales as $\mathrm{poly}(n, \exp(d^{1.5}))$, while the current state-of-the-art classical algorithm [26] approximates the $\binom{n}{d}$ Fourier coefficients, so it requires $\mathrm{poly}(n^d)$ time. In particular, for $d = (\log n)^{2/3}$, constant ε, δ, our quantum algorithm has a time complexity of $\mathrm{poly}(n)$ time, while the state-of-the-art classical algorithm runs in time $n^{\mathrm{polylog} n}$.

4 Learning algorithm for low-degree quantum channels

In this section we give our learning algorithm for quantum channels, i.e., prove Theorem 1. To do that, we first need to prove a new BH inequality for quantum channels.

4.1 Bohnenblust-Hille inequality for quantum channels

In this section, we prove a Bohnenblust-Hille inequality for n-qubit quantum channels. In fact, it is a result for superoperators which are bounded in the S_1 to S_1 norm (defined below), of which quantum channels are a particular example. Hence, we will treat Φ as a linear map from $\mathcal{M}_N$ to $\mathcal{M}_N$, the space of N-dimensional matrices with $N = 2^n$. In particular, we will evaluate Φ on matrices that are not states. The S_1 to S_1 norm of superoperator is defined by

$$\|\Phi\|_{S_1 \to S_1} = \sup_{M \neq 0} \frac{\|\Phi(M)\|_{S_1}}{\|M\|_{S_1}},$$

where $\|M\|_{S_1}$ is the Schatten 1-norm of M, i.e., the sum of the singular values of M.

To prove our theorem will reduce to the classical case of functions $f : \{-1,1\}^n \to \mathbb{R}$.

▶ **Theorem 9** ([23]). *Let $p : \{-1,1\}^n \to \mathbb{R}$ of degree at most d. Then,*

$$\|\widehat{p}\|_{\frac{2d}{d+1}} \leq C^{\sqrt{d \log d}} \|p\|_\infty ,$$

where $C > 0$ is a constant.

To achieve this reduction, for every superoperator $\Phi : \mathcal{M}_M \to \mathcal{M}_N$, we assign it a function $f_\Phi : \{-1,1\}^{3n} \times \{-1,1\}^{3n} \to \mathbb{C}$ defined as follows. For $a = (a^1, a^2, a^3)$, $b = (b^1, b^2, b^3) \in \{-1,1\}^n \times \{-1,1\}^n \times \{-1,1\}^n$ and $s, t \in \{1,2,3\}^n$, define the following matrices (which are not necessarily states)

$$|a^s\rangle\langle b^t| = \bigotimes_{i \in [n]} |\chi^{s(i)}_{a^s(i)}\rangle\langle \chi^{t(i)}_{b^t(i)}|,$$

where $|\chi^s_{\pm 1}\rangle$ are the ± 1 eigenstates of the single-qubit Pauli operators σ_s. The function $f_\Phi : \{-1,1\}^{3n} \times \{-1,1\}^{3n} \to \mathbb{C}$ is then given by

$$f_\Phi(a,b) = \frac{1}{9^n} \sum_{s,t \in \{1,2,3\}^n} \mathrm{Tr}[\Phi\left(|a^s\rangle\langle b^t|\right) |b^t\rangle\langle a^s|],$$

where f_Φ has the following properties, allowing us to reduce to the classical BH inequality.

▶ **Lemma 10.** *Let Φ be a degree-d superoperator. Then, $|f_\Phi(a,b)| \leq \|\Phi\|_{S_1 \to S_1}$ for all a, b and $\left\|\widehat{\Phi}\right\|_p \leq 9^d \left\|\widehat{f_\Phi}\right\|_p$. The degree of f_Φ as a multilinear polynomials is $2d$.*

Proof. We first show the bound on $|f_\Phi|$. Given that $(|a^s\rangle\langle b^t|)(|a^s\rangle\langle b^t|)^* = |a^s\rangle\langle a^s|$, we have that

$$\left\||a^s\rangle\langle b^t|\right\|_{S_1} = \left\||a^s\rangle\langle b^t|\right\|_{S_\infty} = 1. \tag{6}$$

Thus, we have that f_Φ is bounded:

$$\begin{aligned}|f_\Phi(a,b)| &\leq \frac{1}{9^n}\sum_{s,t\in\{1,2,3\}^n} |\mathrm{Tr}[\Phi\left(|a^s\rangle\langle b^t|\right)|b^t\rangle\langle a^s|]| \\ &\leq \frac{1}{9^n}\sum_{s,t\in\{1,2,3\}^n} \left\|\Phi\left(|a^s\rangle\langle b^t|\right)\right\|_{S_1}\left\||b^t\rangle\langle a^s|\right\|_{S_\infty} \\ &\leq \frac{1}{9^n}\sum_{s,t\in\{1,2,3\}^n} \|\Phi\|_{S_1\to S_1}\left\||a^s\rangle\langle b^t|\right\|_{S_1}\left\||b^t\rangle\langle a^s|\right\|_{S_\infty} \\ &\leq \frac{1}{9^n}\sum_{s,t\in\{1,2,3\}^n} \|\Phi\|_{S_1\to S_1} = \|\Phi\|_{S_1\to S_1},\end{aligned}$$

where in the first inequality we have used the triangle inequality, in the second inequality Riesz theorem, in the third the definition of $S_1 \to S_1$ norm and in the fourth Equation (6). We now prove that $\left\|\widehat{\Phi}\right\|_p \leq 9^{-d}\left\|\widehat{f_\Phi}\right\|_p$ and that the degree of f_Φ is $2d$. It suffices to show that

$$f_\Phi(a,b) = \sum_{x,y\in\{0,1,2,3\}^n} \frac{\widehat{\Phi}(x,y)}{3^{|x|+|y|}}\prod_{i\in\mathrm{supp}(x)}\prod_{j\in\mathrm{supp}(y)} a_i^{x(i)}b_j^{y(j)}, \tag{7}$$

where $\mathrm{supp}(x) = \{i\in[n] : x_i \neq 0\}$ and $|x|$ is the size of $\mathrm{supp}(x)$. To prove Equation (7) the key is observing that for every $s,t\in\{1,2,3\}$, $x,y\in\{0,1,2,3\}$ and $a,b\in\{-1,1\}$ we have that

$$\mathrm{Tr}[\sigma_x|\chi_a^s\rangle\langle\chi_b^t|\sigma_y|\chi_b^t\rangle\langle\chi_a^s|] = \begin{cases} 0 & \text{if } (s\neq x \text{ and } x\neq 0) \text{ or } (t\neq y \text{ and } y\neq 0),\\ 1 & \text{if } x=0 \text{ and } y=0,\\ a & \text{if } s=x \text{ and } y=0,\\ b & \text{if } x=0 \text{ and } t=y,\\ ab & \text{if } s=x \text{ and } y=t.\end{cases}$$

After taking tensor products, we observe that for every $s,t\in\{1,2,3\}^n$, $x,y\in\{0,1,2,3\}^n$ and $a=(a^1,a^2,a^3),\ b=(b^1,b^2,b^3)\in\{-1,1\}^n\times\{-1,1\}^n\times\{-1,1\}^n$, it holds that

$$\mathrm{Tr}[\sigma_x|a^s\rangle\langle b^t|\sigma_y|b^t\rangle\langle a^s|] = \prod_{i\in\mathrm{supp}\,x}\prod_{j\in\mathrm{supp}\,y} a_i^{x(i)}b_j^{y(j)}\delta_{x(i),s(i)}\delta_{y(j),t(j)},$$

where $\delta_{x,y}$ is the delta function taking value of 1 when $x=y$, and 0 otherwise. In particular, from this follows that

$$f_{\Phi_{x,y}}(a,b) = \frac{1}{9^n}\sum_{s,t\in\{1,2,3\}^n}\mathrm{Tr}[\sigma_x|a^s\rangle\langle b^t|\sigma_y|b^t\rangle\langle a^s|] = \frac{1}{9^n}\prod_{i\in\mathrm{supp}\,x}\prod_{j\in\mathrm{supp}\,y} a_i^{x(i)}b_j^{y(j)}\sum_{s\in\mathcal{X},t\in\mathcal{Y}} 1,$$

where $\mathcal{X} = \{s\in\{1,2,3\}^n : s(i) = x(i)\ \forall\ i\in\mathrm{supp}(x)\}$. Hence, as $|\mathcal{X}| = 3^{n-|x|}$, Equation (7) follows for the case of $\Phi_{x,y}$. By linearity, Equation (7) follows for every superoperator. ◀

▶ **Theorem 11** (Bohnenblust-Hille inequality for $S_1 \to S_1$ maps)**.** *Let Φ be a super-operator of degree at most d. Then there exists a constant C such that*

$$\left\|\widehat{\Phi}\right\|_{\frac{2d}{d+1/2}} \leq C^d\|\Phi\|_{S_1\to S_1}.$$

In particular, if Φ is a quantum channel, then there exists a constant C such that

$$\left\|\widehat{\Phi}\right\|_{\frac{2d}{d+1/2}} \leq C^d.$$

Proof. Let $\Re f_\Phi : \{-1,1\}^n \to \mathbb{R}$ be defined as taking the real part of f_Φ i.e., $(\Re f_\phi)(x) = \Re(f_\Phi(x))$ and $\Im f_\Phi : \{-1,1\}^n \to \mathbb{R}$ as taking the imaginary part of f_Φ i.e., $(\Im f_\phi)(x) = \Im(f_\Phi(x))$. Note that we have that $\widehat{f_\phi} = \widehat{\Re f}_\Phi + i\widehat{\Im f}_\Phi$. By Lemma 10, $|(\Re f_\phi)(x)|, |(\Im f_\phi)(x)| \leq |f_\Phi(x)| \leq \|\Phi\|_{S_1 \to S_1}$ and that the degree of both the real and imaginary part is at most $2d$. Hence, by the triangle inequality and Theorem 9 we have

$$\left\|\widehat{f}_\Phi\right\|_{\frac{4d}{2d+1}} \leq \left\|\widehat{\Re f}_\Phi\right\|_{\frac{4d}{2d+1}} + \left\|\widehat{\Im f}_\Phi\right\|_{\frac{4d}{2d+1}} \leq C^{\sqrt{d \log d}} \|\Phi\|_{S_1 \to S_1}.$$

Thus, using that $\left\|\widehat{\Phi}\right\|_{2d/(d+1)} \leq 9^d \left\|\widehat{f}_\Phi\right\|_{2d/(d+1)}$ it follows that $\left\|\widehat{\Phi}\right\|_{2d/(d+1)} \leq C^d \|\Phi\|_{S_1 \to S_1}$. This proves the first part of the statement.

For the second we just have to show that if Φ is a quantum channel, then $\|\Phi\|_{S_1 \to S_1}$ is bounded by a constant. Indeed, if M is self-adjoint, we can write it as $M = M^+ - M^-$, with M^+ and M^- being positive semidefinite, so

$$\frac{\|\Phi(M)\|_{S_1}}{\|M\|_{S_1}} \leq \frac{\mathrm{Tr}[\Phi(M^+)] + \mathrm{Tr}[\Phi(M^-)]}{\mathrm{Tr}[M^+] + \mathrm{Tr}[M^-]} = 1, \tag{8}$$

where in the first equality we have used that Φ is positive and in the second that is trace perserving. Finally, any matrix M can be written as $M = \Re M + i\Im M$, where $\Re M = (M + M^*)/2$ and $\Im M = (M - M^*)/2$ are self-adjoint. Hence,

$$\frac{\|\Phi(M)\|_{S_1}}{\|M\|_{S_1}} \leq \frac{\|\Phi(\Re M)\|_{S_1} + \|\Phi(\Im M)\|_{S_1}}{\|M\|_{S_1}} \leq \frac{\|\Re M\|_{S_1} + \|\Im M\|_{S_1}}{\|M\|_{S_1}} \leq \frac{\|M\|_{S_1} + \|M\|_{S_1}}{\|M\|_{S_1}} = 2,$$

where in the first inequality we have used the triangle inequality, in the second inequality we have used Equation (8) and in the third inequality that $\|\Re M\|_{S_1}, \|\Im M\|_{S_1} \leq \|M\|_{S_1}$. ◀

4.2 Learning low-degree quantum channels

Before we prove the main theorem of the section, we show that for given $x, y \in \{0,1,2,3\}^n$, the corresponding Fourier coefficient $\widehat{\Phi}(x,y)$ can be efficiently learned. This is accomplished through the combination of a few SWAP tests.

▶ **Fact 12** (SWAP test for mixed states [37]). *Let ρ, ρ' be two states. Then, one can estimate $\mathrm{Tr}[\rho\rho']$ up to error ε with probability $1-\delta$ using $O((1/\varepsilon)^2 \log(1/\delta))$ copies of ρ and ρ'.*

▶ **Lemma 13** (Pauli coefficient estimation for channels). *Let $x, y \in \{0,1,2,3\}^n$. Then, $\widehat{\Phi}(x,y)$ can be estimated with error ε and probability $1-\delta$ using $O((1/\varepsilon)^2 \log(1/\delta))$ queries to Φ.*

Proof. If $x = y$, we just have to prepare $\widehat{\Phi}$ (which can be done by preparing the Choi state $J(\Phi)$ following by a unitary transformation) and apply Fact 12 to $\widehat{\Phi}$ and the state $\rho = |x\rangle\langle x|$. If $x \neq y$, one first learns $\widehat{\Phi}(x,x)$ and $\widehat{\Phi}(y,y)$ with error ε as before. One the one hand, one can learn $\widehat{\Phi}(x,x) + \widehat{\Phi}(x,x) + 2\Re\widehat{\Phi}(x,y)$, with error ε by applying Fact 12 to $\widehat{\Phi}$ and $1/2\sum_{z,t\in\{x,y\}}|z\rangle\langle t|$. Hence, one learns $\Re\widehat{\Phi}(x,y)$ with error $3\varepsilon/2$. On the other hand, one can learn $\widehat{\Phi}(x,x) + \widehat{\Phi}(y,y) + 2\Im\widehat{\Phi}(x,y)$, with error ε by applying Fact 12 to $\widehat{\Phi}$ and $1/2(|x\rangle\langle x| + i|x\rangle\langle y| - i|y\rangle\langle x| + |y\rangle\langle y|)$, and thus one can learn $\Im\widehat{\Phi}(x,y)$ with error $3\varepsilon/2$. ◀

We will also need the following well-known result on learning discrete probability distributions. See [17, Theorem 9] for a proof.

▶ **Lemma 14.** *Let $p = \{p(x)\}_x$ be a probability distribution over some set $\mathcal{X}$. Let $p' = (p'(x))_x$ the empirical probability distribution obtained after sampling T times from p. Then, for $T = O((1/\varepsilon)^2 \log(1/\delta))$ with probability $1-\delta$ we have that $|p(x) - p'(x)| \leq \varepsilon$ for every $x \in \mathcal{X}$.*

Now, we are ready to prove Theorem 1, which we restate for the convenience of the reader.

▶ **Theorem 1.** *Let Φ be a n-qubit degree-d quantum channel. There is an algorithm that (ε, δ)-learns Φ (in ℓ_2-distance) using* $\exp(\widetilde{O}(d^2 + d\log(1/\varepsilon))) \cdot \log(1/\delta)$ *queries to Φ.*

Proof. We first state the algorithm.

Algorithm 1 Learning low-degree channels via BH inequality.

Input: A quantum channel Φ of degree at most d, and error ε and a failure probability δ

1: Let $c = \varepsilon^{2d+1}C^{-d^2}$

2: Prepare $T_1 = O((1/c)^2 \log(1/\delta))$ copies of $\widehat{\Phi}$ to sample from $(\widehat{\Phi}(x,x))_x$. Let $(\widehat{\Phi}'(x,x))_x$ be the associated empirical distribution

3: **for** $x, y \in \mathcal{X}_c = \{x : |\widehat{\Phi}'(x,x)| \geq c\}$ **do**

4: Prepare $O((1/c)^2(1/\varepsilon)^2 \log((1/c)^2(1/\delta)))$ copies of $\widehat{\Phi}$ and use them to approximate $\widehat{\Phi}(x,y)$ with $\widehat{\Phi}''(x,y)$ using Lemma 13.

5: **end for**

Output: $\sum_{x,y\in\mathcal{X}_c} \widehat{\Phi}''(x,y)\Phi_{x,y}$

Let $c > 0$ to be determined later. In the first part of the algorithm we prepare $\widehat{\Phi}$ and measure, i.e., we sample from $(\widehat{\Phi}(x,x))_{x\in\{0,1,2,3\}^n}$. Let $(\widehat{\Phi}'(x,x))_{x\in\{0,1,2,3\}^n}$ be the empirical distribution one obtains after T_1 samples. $\mathcal{E} = \{|\widehat{\Phi}(x,x) - \widehat{\Phi}'(x,x)| \leq c \ \forall x \in \{0,1,2,3\}^n\}$. By Lemma 14, taking T_1 to be $O((1/c)^2 \log(1/\delta))$ ensures that

$$\Pr[\mathcal{E}] \geq 1 - \delta.$$

Let $\mathcal{X}_c = \{x : \ |\widehat{\Phi}'(x,x)| \geq c\}$. Note that, as $\sum_{x\in\mathcal{X}_c} \Phi(x,x) \leq 1$,

$$|\mathcal{X}_c| \leq c^{-1}, \tag{9}$$

and in the event of $\mathcal{E}$ we have that

$$x \notin \mathcal{X}_c \implies |\widehat{\Phi}(x,x)| \leq |\widehat{\Phi}'(x,x)| + ||\widehat{\Phi}(x,x)| - |\widehat{\Phi}'(x,x)|| \leq 2c. \tag{10}$$

In particular, it follows that

$$x \notin \mathcal{X}_c \implies |\widehat{\Phi}(x,y)| \leq \sqrt{|\widehat{\Phi}(x,x)||\widehat{\Phi}(y,y)|} \leq \sqrt{2c} \quad \forall y \in \{0,1,2,3\}^n, \tag{11}$$

where in the first inequality we have used that $\widehat{\Phi}$ is positive semidefinite and in the second inequality Equation (10) and that $\widehat{\Phi}(y,y) \leq 1$. We assume that the first part of the algorithm succeeds, meaning that $\mathcal{E}$ happens. In the second part of the algorithm we approximate all the Pauli coefficients of $\mathcal{X}_c \times \mathcal{X}_c$ with error $c\varepsilon$ and probability $1-\delta$ querying Φ just

$$T_2 = O((1/c)^4(1/\varepsilon)^2 \log((1/c)^2(1/\delta)))$$

times from Lemma 13. Note that $T_2 > T_1$, so this complexity dominates the one of the first part of the algorithm. Let $\widehat{\Phi}''(x,y)$ be these approximations and let $\Phi_c = \sum_{x,y\in\mathcal{X}_c} \widehat{\Phi}''(x,y)\sigma_x \cdot \sigma_y$. Now, we have that

$$\begin{aligned}
\|\Phi - \Phi_c\|_2^2 &= \sum_{x,y\in\mathcal{X}_c} |\widehat{\Phi}(x,y) - \widehat{\Phi}''(x,y)|^2 + \sum_{x\vee y\notin\mathcal{X}_c} |\widehat{\Phi}(x,y)|^2 \\
&\le \varepsilon^2 + \sum_{x\vee y\notin\mathcal{X}_c} |\widehat{\Phi}(x,y)|^{\frac{1}{d+1/2}} |\widehat{\Phi}(x,y)|^{\frac{2d}{d+1/2}} \\
&\le \varepsilon^2 + (2c)^{\frac{1/2}{d+1/2}} \left\|\widehat{\Phi}\right\|_{\frac{2d}{d+1/2}}^{\frac{2d}{d+1/2}} \\
&\le \varepsilon^2 + c^{\frac{1/2}{d+1/2}} C^d,
\end{aligned}$$

where in the equality we have used Parseval's identity; in the first inequality we used Equation (9), the learning guarantees of the second part of the algorithm and that $2 = 1/(d+1/2) + 2d/(d+1/2)$; in the second inequality we have used Equation (11); and in the third inequality we used the Bohnenblust-Hille inequality for channels (Theorem 11). Hence, by choosing

$$c = \varepsilon^{4d+2} C^{-d^2}$$

we obtain the desired result. ◀

5 Learning quantum query algorithms

In this section, our goal will be to prove Theorem 5, restated below for the reader's convenience.

▶ **Theorem 5.** *For a quantum algorithm that makes d-queries as in Figure 1, its amplitudes can be learned up to error ε in ℓ_2^2 accuracy using $O\left((1/\varepsilon)^d \cdot \log n\right)$ uniformly random samples.*

Proof. Eskenazis and Ivanishvili [26] showed that a function $f : \{-1,1\}^n \to [-1,1]$ with degree at most d and $\left\|\widehat{f}\right\|_{\frac{2d}{d+1}} \le C$, can be learned with success probablity $1-\delta$ and error ε in ℓ_2^2 accuracy using $O(\varepsilon^{-(d+1)} C^{2d} \log(n/\delta))$ samples $(x, f(x))$, where x is drawn uniformly at random from $x \in \{-1,1\}^n$. Arunachalam et al. [3] showed that the amplitudes of quantum algorithms that d queries as in Figure 1 are completely bounded d-tensors. Hence, using Theorem 15 (which we prove below) we can let $C = 1$, and obtain the desired result. ◀

5.1 The constant of the completely bounded BH inequality is 1

In this section we determine that the exact value of the constant of the completely bounded BH inequality is 1. Before that, we first define the completely bounded norm. For d-linear tensors, the completely bounded norm is defined as

$$\|T\|_{\mathrm{cb}} = \sup \left\| \sum_{\mathbf{i}\in[n]^d} \widehat{T}_{\mathbf{i}} X_1(i_1) \dots X_t(i_t) \right\|_{\mathrm{op}}, \tag{12}$$

where $X_s(i_s)$ are matrices of size $m \times m$ that have operator norm at most 1 and $m \in \mathbb{N}$.

▶ **Theorem 15.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. Let $T : (\mathbb{K}^n)^d \to \mathbb{K}$ be a d-linear form. Then,*

$$\left\|\widehat{T}\right\|_{\frac{2d}{d+1}} \le \|T\|_{\mathrm{cb}},$$

and the inequality can be saturated.

Theorem 15 establishes that the best constant for the completely bounded BH inequality is exactly 1. This sharply contrasts with the current knowledge about the BH constants, where only poly(d) upper bounds are known. We thus *close* one of the edges of comparison of the three norms that appear in Grothendieck and Bohnenblust-Hille inequalities (see Figure 2).

$$\begin{tikzcd} & \|B\|_\infty \arrow[dl, phantom, "\leq \sqrt{2}"] \arrow[dr, phantom, "G_{\mathbb{R}} \geq"] & \\ \|B\|_{\frac{2d}{d+1}} \arrow[rr, phantom, "\leq 1"] & & \|B\|_{\mathrm{cb}} \end{tikzcd} \qquad \begin{tikzcd} & \|T\|_\infty \arrow[dl, phantom, "\leq \mathrm{poly}(d)"] \arrow[dr, phantom, "C(d) \ngeq"] & \\ \|T\|_{\frac{2d}{d+1}} \arrow[rr, phantom, "\leq 1"] & & \|T\|_{\mathrm{cb}} \end{tikzcd}$$

Figure 2 Triangles of norm comparisons. In the left triangle, we display the norm comparisons implied by the Littlewood [40] and Grothendieck inequalities [30] and our Theorem 15 for real bilinear maps. In the right triangle, we depict the best upper bound for the BH constant [11], the no extension of the Grothendieck inequality [49], and our Theorem 15 for d-linear tensors.

The main ingredient of the proof of Theorem 15 is a general lower bound for the completely bounded norm, Lemma 17. This technique is inspired by the idea of Varoupoulos to rule out a generalization of von Neumann's inequality [56] and was recently used by Escudero-Gutiérrez to show a particular case of the famous Aaronson and Ambainis conjecture of quantum query complexity [25]. Theorem 15 follows from combining Lemma 17 with Blei's inequality Lemma 16. See [11, Theorem 2.1] for a proof of Blei´s inequality.

▶ **Lemma 16** (Blei's inequality)**.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and let $\widehat{T} \in (\mathbb{K})^{n^d}$. Then,*

$$\Big(\prod_{s\in[d]} \sum_{i_s\in[n]} \sqrt{\sum_{i_1,\dots,i_{s-1},i_{s+1},\dots,i_d\in[n]} |\widehat{T}_{\mathbf{i}}|^2} \Big)^{\frac{1}{d}} \geq \left\|\widehat{T}\right\|_{\frac{2d}{d+1}}.$$

▶ **Lemma 17.** *Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, let $T : (\mathbb{K}^n)^d \to \mathbb{K}$ be a d-linear form, and let $s \in [d]$. Then,*

$$\|T\|_{\mathrm{cb}} \geq \sum_{i_s\in[n]} \sqrt{\sum_{i_1,\dots,i_{s-1},i_{s+1},\dots,i_d\in[n]} |\widehat{T}_{\mathbf{i}}|^2}.$$

Proof. The proof involves evaluating Equation (12) on an explicit set of contractions (matrices with operator norm at most 1). Let $m = \sum_{r=0}^{s-1} n^r + \sum_{r=0}^{d-r} n^r$ and let $\{e_{\mathbf{i}}, f_{\mathbf{j}} : \mathbf{i} \in [n]^r, r \in \{0\} \cup [d-s], \mathbf{j} \in [n]^t, t \in \{0\} \cup [s-1]\}$ be an orthonormal basis of $\ell_2^m(\mathbb{K})$, where we identify $[n]^0$ with $\emptyset$. We define $m \times m$ matrices $X(i)$ for $i \in [n]$ by

$$\begin{aligned} X(i)e_{\mathbf{j}} &= e_{(i,\mathbf{j})}, \text{ if } \mathbf{j} \in [n]^r, \ r \in \{0\} \cup [d-s-1], \\ X(i)e_{\mathbf{j}} &= \frac{\sum_{\mathbf{k}\in[n]^{s-1}} \widehat{T}^*_{\mathbf{k}i\mathbf{j}} f_{\mathbf{k}}}{\sqrt{\sum_{k_1,\dots,k_{s-1},k_{s+1},\dots,k_d\in[n]} |\widehat{T}|^2_{(k_1,\dots,k_{s-1},i,k_{s+1},\dots,k_d)}}}, \text{ if } \mathbf{j} \in [n]^{d-s}, \\ X(i)f_{\mathbf{j}} &= \delta_{i,j_s} f_{(i_1,\dots,i_{s-1})}, \text{ if } \mathbf{j} \in [n]^{d-s}, \ r \in \{0\} \cup [d-1]. \end{aligned}$$

For some intuition of the behaviour of these matrices, one may interpret the first $d-s$ applications of the matrices $X(i)$ as *creation* operators and the last $s-1$ as *destruction* operators. Assume for the moment that $X(i)$ are contractions. Given that,

$$\langle f_\emptyset, X(i_1)\dots X(i_d)e_\emptyset\rangle = \frac{\widehat{T}_{\mathbf{i}^*}}{\sqrt{\sum_{i_1,\dots,i_{s-1},i_{s+1},\dots,i_d\in[n]} |\widehat{T}_{\mathbf{i}}|^2}},$$

it would then follow that

$$\begin{aligned}
\|T\|_{\mathrm{cb}} &\geq \Bigg\| \sum_{\mathbf{i}\in[n]^d} \widehat{T}_{\mathbf{i}} X(i_1)\dots X(i_d) \Bigg\|_{\mathrm{op}} \\
&\geq \sum_{\mathbf{i}\in[n]^d} \widehat{T}_{\mathbf{i}} \frac{\widehat{T}^*_{\mathbf{i}}}{\sqrt{\sum_{k_1,\dots,k_{s-1},k_{s+1},\dots,k_d\in[n]} |\widehat{T}|^2_{(k_1,\dots,k_{s-1},i_s,k_{s+1},\dots,k_d)}}} \\
&= \sum_{i_s\in[n]} \sqrt{\sum_{i_1,\dots,i_{s-1},i_{s+1},\dots,i_d\in[n]} |\widehat{T}_{\mathbf{i}}|^2},
\end{aligned}$$

as desired. It then remains to prove that the matrices $X(i)$ are contractions. Given that $X(i)$ maps $\{e_{\mathbf{i}} : \mathbf{i}\in[n]^r,\ r\in\{0\}\cup[d-s-1]\}$, $\{e_{\mathbf{i}} : \mathbf{i}\in[n]^{d-s}\}$ and $\{f_{\mathbf{i}} : \mathbf{i}\in[n]^r,\ r\in\{0\}\cup[s]\}$ to orthogonal subspaces, it suffices to show that the $X(i)$ are contractions when restricted to those subspaces. For the first and third sets, this is true because $X(i)$ maps each basis vector of these sets to a different basis vector or to 0. For the second set, is also true because for every $\lambda\in\mathbb{K}^{n^{d-s}}$

$$\begin{aligned}
\Bigg\| X(i) \sum_{\mathbf{j}\in[n]^{d-s}} \lambda_{\mathbf{j}} e_{\mathbf{j}} \Bigg\|_2 &= \Bigg\| \frac{\sum_{\mathbf{k}\in[n]^{s-1}} (\sum_{\mathbf{j}\in[n]^{d-s}} \lambda_{\mathbf{j}} \widehat{T}^*_{\mathbf{k}i\mathbf{j}}) f_{\mathbf{k}}}{\sqrt{\sum_{k_1,\dots,k_{s-1},k_{s+1},\dots,k_d\in[n]} |\widehat{T}|^2_{(k_1,\dots,k_{s-1},i,k_{s+1},\dots,k_d)}}} \Bigg\|_2 \\
&= \frac{\sqrt{\sum_{\mathbf{k}\in[n]^{s-1}} |\sum_{\mathbf{j}\in[n]^{d-s}} \lambda_{\mathbf{j}} \widehat{T}^*_{\mathbf{k}i\mathbf{j}}|^2}}{\sqrt{\sum_{k_1,\dots,k_{s-1},k_{s+1},\dots,k_d\in[n]} |\widehat{T}|^2_{(k_1,\dots,k_{s-1},i,k_{s+1},\dots,k_d)}}} \\
&\leq \frac{\sqrt{\sum_{\mathbf{k}\in[n]^{s-1}} \sum_{\mathbf{j}\in[n]^{d-s}} |\widehat{T}^*_{\mathbf{k}i\mathbf{j}}|^2} \sqrt{\sum_{\mathbf{j}\in[n]^{d-s}} |\lambda_{\mathbf{j}}|^2}}{\sqrt{\sum_{k_1,\dots,k_{s-1},k_{s+1},\dots,k_d\in[n]} |\widehat{T}|^2_{(k_1,\dots,k_{s-1},i,k_{s+1},\dots,k_d)}}} \\
&= \|\lambda\|_2\,,
\end{aligned}$$

where in the inequality we have used Cauchy-Schwarz for the sum over $\mathbf{j}$. ◀

Proof of Theorem 15. The inequality $\|\widehat{T}\|_{\frac{2d}{d+1}} \leq \|T\|_{\mathrm{cb}}$ follows from Lemmas 16 and 17. The inequality is saturated by the form $T(x_1,\dots,x_d) = x_1(1)$. ◀

References

1. Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 307–316, New York, NY, USA, 2015. Association for Computing Machinery. `doi:10.1145/2746539.2746547`.
2. Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. Optimal Algorithms for Learning Quantum Phase States. In *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*, volume 266 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:24, 2023.
3. Srinivasan Arunachalam, Jop Briët, and Carlos Palazuelos. Quantum query algorithms are completely bounded forms. *SIAM J. Comput*, 48(3):903–925, 2019. Preliminary version in ITCS'18.
4. Srinivasan Arunachalam, Sourav Chakraborty, Michal Koucký, Nitin Saurabh, and Ronald De Wolf. Improved bounds on Fourier entropy and min-entropy. *ACM Transactions on Computation Theory (TOCT)*, 13(4):1–40, 2021.

5 Srinivasan Arunachalam, Sourav Chakraborty, Troy Lee, Manaswi Paraashar, and Ronald De Wolf. Two new results about quantum exact learning. *Quantum*, 5:587, 2021.

6 Srinivasan Arunachalam, Arkopal Dutt, Francisco Escudero Gutierrez, and Carlos Palazuelos. Learling low-degree quantum objects. *arXiv preprint*, 2024.

7 Alp Atıcı and Rocco A Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2007.

8 Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 1303–1316, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451040.

9 Nikhil Bansal, Makrand Sinha, and Ronald de Wolf. Influence in Completely Bounded Block-Multilinear Forms and Classical Simulation of Quantum Algorithms. In *37th Computational Complexity Conference (CCC 2022)*, volume 234, pages 28:1–28:21, 2022. doi:10.4230/LIPIcs.CCC.2022.28.

10 Zongbo Bao and Penghui Yao. On Testing and Learning Quantum Junta Channels. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 1064–1094. PMLR, 12–15 July 2023. URL: https://proceedings.mlr.press/v195/bao23b.html.

11 Frédéric Bayart, Daniel Pellegrino, and Juan B Seoane-Sepúlveda. The Bohr radius of the n-dimensional polydisk is equivalent to $(\log n)/n$. *Advances in Mathematics*, 264:726–746, 2014.

12 Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

13 Henri Frédéric Bohnenblust and Einar Hille. On the absolute convergence of Dirichlet series. *Annals of Mathematics*, pages 600–622, 1931.

14 Harald Bohr. Ueber die Bedeutung der Potenzreihen unendlich vieler Variablen in der Theorie der Dirichletschen Reihen. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1913:441–488, 1913.

15 Nader H. Bshouty and Jeffrey C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory*, COLT '95, pages 118–127, New York, NY, USA, 1995. Association for Computing Machinery. doi:10.1145/225298.225312.

16 Mark Bun and Justin Thaler. Guest column: Approximate degree in classical and quantum computing. *ACM SIGACT News*, 51(4):48–72, 2021.

17 Clément L Canonne. A short note on learning discrete distributions. *arXiv preprint*, 2020. arXiv:2002.11457.

18 Shantanav Chakraborty, András Gilyén, and Stacey Jeffery. The Power of Block-Encoded Matrix Powers: Improved Regression Techniques via Faster Hamiltonian Simulation. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 33:1–33:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ICALP.2019.33.

19 Chi-Fang Chen, Andrew Lucas, and Chao Yin. Speed limits and locality in many-body quantum dynamics. *Reports on Progress in Physics*, 2023.

20 Thomas Chen, Shivam Nadimpalli, and Henry Yuen. Testing and Learning Quantum Juntas Nearly Optimally. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1163–1185. SIAM, 2023.

21 Kai-Min Chung and Han-Hsuan Lin. Sample efficient algorithms for learning quantum channels in pac model and the approximate state discrimination problem. *arXiv preprint*, 2018. arXiv:1810.10938.

22 Andreas Defant, Leonhard Frerick, Joaquim Ortega-Cerdà, Myriam Ounaïes, and Kristian Seip. The Bohnenblust–Hille inequality for homogeneous polynomials is hypercontractive. *Annals of mathematics*, pages 485–497, 2011.

23 Andreas Defant, Mieczysław Mastyło, and Antonio Pérez. On the Fourier spectrum of functions on boolean cubes. *Mathematische Annalen*, 374:653–680, 2019.

24 Diogo Diniz, Gustavo Muñoz-Fernández, Daniel Pellegrino, and J Seoane-Sepúlveda. Lower bounds for the constants in the Bohnenblust–Hille inequality: The case of real scalars. *Proceedings of the American Mathematical Society*, 142(2):575–580, 2014.

25 Francisco Escudero Gutiérrez. Influences of Fourier completely bounded polynomials and classical simulation of quantum algorithms. *arXiv:2304.06713*, 2023.

26 Alexandros Eskenazis and Paata Ivanisvili. Learning low-degree functions from a logarithmic number of random queries. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, pages 203–207, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3519935.3519981.

27 Alexandros Eskenazis, Paata Ivanisvili, and Lauritz Streck. Low-degree learning and the metric entropy of polynomials. *arXiv preprint*, 2022. arXiv:2203.09659.

28 Omar Fawzi, Aadil Oufkir, and Daniel Stilck França. Lower bounds on learning Pauli channels. *arXiv preprint*, 2023. arXiv:2301.09192.

29 Steven T Flammia and Ryan O'Donnell. Pauli error estimation via population recovery. *Quantum*, 5:549, 2021.

30 Alexandre Grothendieck. *Résumé de la théorie métrique des produits tensoriels topologiques*. Soc. de Matemática de São Paulo, 1953.

31 Gus Gutoski and Nathaniel Johnston. Process tomography for unitary quantum channels. *Journal of Mathematical Physics*, 55(3), 2014.

32 Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017.

33 Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes. *PRX Quantum*, 4:040337, December 2023. doi:10.1103/PRXQuantum.4.040337.

34 Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

35 Siddharth Iyer, Anup Rao, Victor Reis, Thomas Rothvoss, and Amir Yehudayoff. Tight bounds on the Fourier growth of bounded functions on the hypercube. *arXiv preprint*, 2021. arXiv:2107.06309.

36 Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. Quantum and Classical Low-Degree Learning via a Dimension-Free Remez Inequality. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 69:1–69:22, 2024.

37 Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. Quantum merlin-arthur proof systems: Are multiple merlins more helpful to arthur? In *Algorithms and Computation: 14th International Symposium, ISAAC 2003*, pages 189–198. Springer, 2003.

38 Ching-Yi Lai and Hao-Chung Cheng. Learning quantum circuits of T-depth one. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2213–2218, 2022. doi:10.1109/ISIT50566.2022.9834452.

39 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.

40 John E Littlewood. On bounded bilinear forms in an infinite number of variables. *The Quarterly Journal of Mathematics*, pages 164–174, 1930.

41 Richard A. Low. Learning and testing algorithms for the clifford group. *Phys. Rev. A*, 80:052314, November 2009.

42 Ashley Montanaro. Some applications of hypercontractive inequalities in quantum information theory. *Journal of Mathematical Physics*, 53(12), 2012.

43 Ashley Montanaro and Tobias J Osborne. Quantum Boolean functions. *arXiv preprint*, 2008. arXiv:0810.2435.

44 Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. On the Pauli spectrum of QAC0. *arXiv preprint*, 2023. arXiv:2311.09631.

45 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge, UK, 2010. doi:10.1017/CBO9780511976667.

46 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

47 Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 899–912, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2897518.2897544.

48 Daniel Pellegrino and Eduardo V Teixeira. Towards sharp Bohnenblust–Hille constants. *Communications in Contemporary Mathematics*, 20(03):1750029, 2018.

49 David Pérez-García, Michael M Wolf, Carlos Palazuelos, Ignacio Villanueva, and Marius Junge. Unbounded violation of tripartite Bell inequalities. *Communications in Mathematical Physics*, 279:455–486, 2008.

50 Joseph Slote, Alexander Volberg, and Haonan Zhang. Bohnenblust–Hille inequality for cyclic groups. *arXiv preprint*, 2023. arXiv:2305.10560.

51 Joseph Slote, Alexander Volberg, and Haonan Zhang. Noncommutative bohnenblust-hille inequality in the heisenberg-weyl and gell-mann bases with applications to fast learning. *arXiv preprint*, 2023. arXiv:2301.01438.

52 Barbara M. Terhal. Quantum error correction for quantum memories. *Rev. Mod. Phys.*, 87:307–346, April 2015.

53 Minh C Tran, Kunal Sharma, and Kristan Temme. Locality and error mitigation of quantum circuits. *arXiv preprint*, 2023. arXiv:2303.06496.

54 Leslie G. Valiant. A Theory of the Learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

55 Ewout Van Den Berg, Zlatko K Minev, Abhinav Kandala, and Kristan Temme. Probabilistic error cancellation with sparse Pauli–Lindblad models on noisy quantum processors. *Nature Physics*, pages 1–6, 2023.

56 N. Th. Varopoulos. On an inequality of von Neumann and an application of the metric theory of tensor products to operators theory. *J. Functional Analysis*, 16:83–100, 1974. doi:10.1016/0022-1236(74)90071-8.

57 Alexander Volberg and Haonan Zhang. Noncommutative Bohnenblust–Hille inequalities. *Mathematische Annalen*, pages 1–20, 2023.