



CHORUS

This is the accepted manuscript made available via CHORUS. The article has been published as:

Information-Theoretic Bounds on Quantum Advantage in Machine Learning

Hsin-Yuan Huang, Richard Kueng, and John Preskill

Phys. Rev. Lett. **126**, 190505 — Published 14 May 2021

DOI: [10.1103/PhysRevLett.126.190505](https://doi.org/10.1103/PhysRevLett.126.190505)

Information-theoretic bounds on quantum advantage in machine learning

Hsin-Yuan Huang,^{1,2} Richard Kueng,³ and John Preskill^{1,2,4,5}

¹*Institute for Quantum Information and Matter, Caltech, Pasadena, CA, USA*

²*Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA*

³*Institute for Integrated Circuits, Johannes Kepler University Linz, Austria*

⁴*Walter Burke Institute for Theoretical Physics, Caltech, Pasadena, CA, USA*

⁵*AWS Center for Quantum Computing, Pasadena, CA, USA*

(Dated: April 7, 2021)

We study the performance of classical and quantum machine learning (ML) models in predicting outcomes of physical experiments. The experiments depend on an input parameter x and involve execution of a (possibly unknown) quantum process \mathcal{E} . Our figure of merit is the number of runs of \mathcal{E} required to achieve a desired prediction performance. We consider classical ML models that perform a measurement and record the classical outcome after each run of \mathcal{E} , and quantum ML models that can access \mathcal{E} coherently to acquire quantum data; the classical or quantum data is then used to predict outcomes of future experiments. We prove that for any input distribution $\mathcal{D}(x)$, a classical ML model can provide accurate predictions *on average* by accessing \mathcal{E} a number of times comparable to the optimal quantum ML model. In contrast, for achieving accurate prediction on *all* inputs, we prove that exponential quantum advantage is possible. For example, to predict expectations of all Pauli observables in an n -qubit system ρ , classical ML models require $2^{\Omega(n)}$ copies of ρ , but we present a quantum ML model using only $\mathcal{O}(n)$ copies. Our results clarify where quantum advantage is possible and highlight the potential for classical ML models to address challenging quantum problems in physics and chemistry.

I. INTRODUCTION

The widespread applications of machine learning (ML) to problems of practical interest have fueled interest in machine learning using quantum platforms [20, 47, 79]. Though many potential applications of quantum ML have been proposed, so far the prospect for quantum advantage in solving purely classical problems remains unclear [10, 41, 85, 86]. On the other hand, it seems plausible that quantum ML can be fruitfully applied to problems faced by quantum scientists, such as characterizing the properties of quantum systems and predicting the outcomes of quantum experiments [6, 28, 30, 40, 67, 81, 88].

Here we focus on an important class of learning problems motivated by quantum mechanics. Namely, we are interested in predicting functions of the form

$$f(x) = \text{tr}(O \mathcal{E}(|x\rangle\langle x|)), \quad (1)$$

where x is a classical input, \mathcal{E} is an arbitrary (possibly unknown) completely positive and trace preserving (CPTP) map, and O is a known observable. Equation (1) encompasses *any* physical process that takes a classical input and produces a real number as output. The goal is to construct a function $h(x)$ that accurately approximates $f(x)$ after accessing the physical process \mathcal{E} as few times as possible.

A particularly important special case of setup (1) is training an ML model to predict what would happen in physical experiments [67]. Such experiments might explore, for instance, the outcome of a reaction in quantum chemistry [96], ground state properties of a novel molecule or material [17, 27, 40, 59, 73, 74, 92], or the behavior of neutral atoms in an analog quantum simulator [19, 25, 63]. In these cases, the input x subsumes parameters that characterize the process, e.g., chemicals involved in the reaction, a description of the molecule, or the intensity of lasers that control the neutral atoms. The map \mathcal{E} characterizes a quantum evolution happening in the lab. Depending

on the parameter x , it produces the quantum state $\mathcal{E}(|x\rangle\langle x|)$. Finally, the experimentalist measures a certain observable O at the end of the experiment. The goal is to predict the measurement outcome for new physical experiments, with new values of x that have not been encountered during the training process.

Motivated by these concrete applications, we want to understand the power of classical and quantum ML models in predicting functions of the form given in Equation (1). On the one hand, we consider classical ML models that can gather classical measurement data $\{(x_i, o_i)\}_{i=1}^{N_C}$, where o_i is the outcome when we perform a POVM measurement on the state $\mathcal{E}(|x_i\rangle\langle x_i|)$. We denote by N_C the number of such experiments performed during training in the classical ML setting. On the other hand, we consider quantum ML models in which multiple runs of the CPTP map \mathcal{E} can be composed coherently to collect quantum data, and predictions are produced by a quantum computer with access to the quantum data. We denote by N_Q the number of times \mathcal{E} is used during training in the quantum setting. The classical and quantum ML settings are illustrated in Figure 1.

We focus on the question of whether quantum ML can have a large advantage over classical ML: to achieve a small prediction error, can the optimal N_Q in the quantum ML setting be much less than the optimal N_C in the classical ML setting? For the purpose of this comparison, we disregard the runtime of the classical or quantum ML models that generate the predictions; we are only interested in how many times the process \mathcal{E} must run during the learning phase in the quantum and classical settings.

Our first main result addresses small *average* prediction error, i.e. the prediction error $|h(x) - f(x)|^2$ averaged over some specified input distribution $\mathcal{D}(x)$. We rigorously show that, for any \mathcal{E} , O , and \mathcal{D} , and for any quantum ML model, one can always design a classical ML model achieving a similar average prediction error such that N_C is larger than N_Q by at worst a small polynomial factor. Hence, there is no

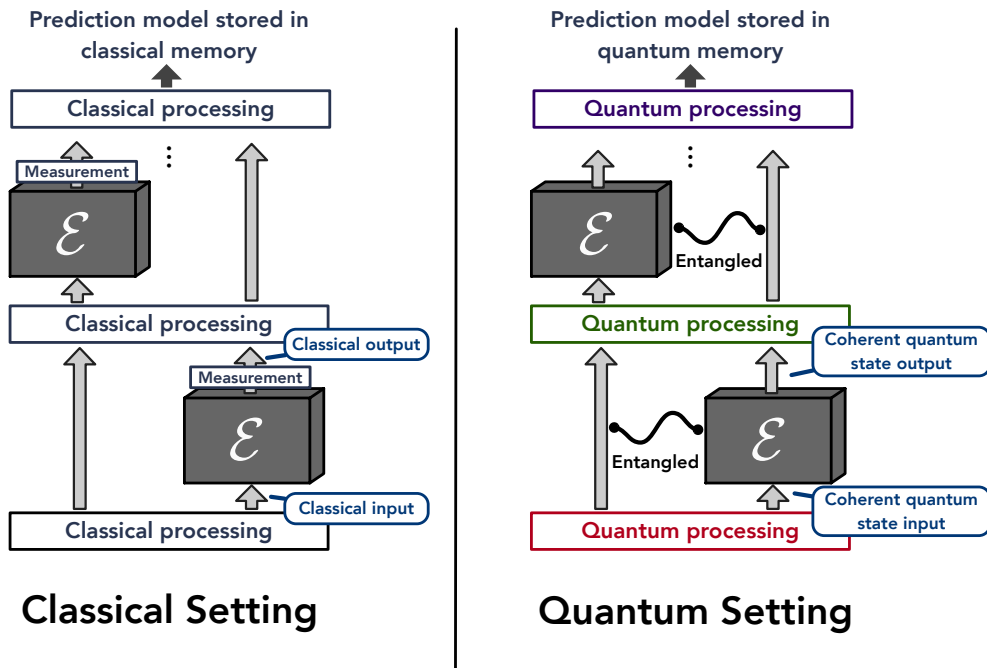


Figure 1: *Illustration of classical and quantum machine learning settings:* The goal is to learn about an unknown CPTP map \mathcal{E} by performing physical experiments. (Left) In the learning phase of the classical ML setting, a measurement is performed after each query to \mathcal{E} ; the classical measurement outcomes collected during the learning phase are consulted during the prediction phase. (Right) In the learning phase of the quantum ML setting, multiple queries to \mathcal{E} may be included in a single coherent quantum circuit, yielding an output state stored in a quantum memory; this stored quantum state is consulted during the prediction phase.

exponential advantage of quantum ML over classical ML if the goal is to achieve a small average prediction error, and if the efficiency is quantified by the number of times \mathcal{E} is used in the learning process. This statement holds for existing quantum ML models running on near-term devices [47, 53, 79] and future quantum ML models yet to be conceived. We note, though, that while there is no large advantage in query complexity, a substantial quantum advantage in computational complexity is possible [80].

However, the situation changes if the goal is to achieve a small *worst-case* prediction error rather than a small average prediction error — an exponential separation between N_C and N_Q becomes possible if we insist on predicting $f(x) = \text{tr}(O \mathcal{E}(|x\rangle\langle x|))$ accurately for *every* input x . We illustrate this point with an example: accurately predicting expectation values of Pauli observables in an unknown n -qubit quantum state ρ . This is a crucial subroutine in many quantum computing applications; see e.g. [34, 52, 54, 56–58, 61, 74]. We present a quantum ML model that uses $N_Q = \mathcal{O}(n)$ copies of ρ to predict expectation values of all n -qubit Pauli observables. In contrast we prove that *any* classical ML model requires $N_C = 2^{\Omega(n)}$ copies of ρ to achieve the same task even if the ML model can perform arbitrary adaptive single-copy POVM measurements.

II. MACHINE LEARNING SETTINGS

We assume that the observable O (with $\|O\| \leq 1$) is known and the physical experiment \mathcal{E} is an unknown CPTP map that belongs to a set of CPTP

maps \mathcal{F} . Apart from $\mathcal{E} \in \mathcal{F}$, the process can be arbitrary — a common assumption in statistical learning theory [12, 15, 21, 87, 89]. For the sake of concreteness, we assume that \mathcal{E} is a CPTP map from a Hilbert space of n qubits to a Hilbert space of m qubits. Regarding inputs, we consider bit-strings of size n : $x \in \{0, 1\}^n$. This is not a severe restriction, since floating-point representations of continuous parameters can always be truncated to a finite number of digits. We now give precise definitions for classical and quantum ML settings; see Fig. 1 for an illustration.

a. Classical (C) ML: The ML model consists of two phases: learning and prediction. During the learning phase, a randomized algorithm selects classical inputs x_i and we perform a (quantum) experiment that results in an outcome o_i from performing a POVM measurement on $\mathcal{E}(|x_i\rangle\langle x_i|)$. A total of N_C experiments give rise to the classical training data $\{(x_i, o_i)\}_{i=1}^{N_C}$. After obtaining this training data, the ML model executes a randomized algorithm \mathcal{A} to learn a prediction model

$$s_C = \mathcal{A}(\{(x_1, o_1), \dots, (x_{N_C}, o_{N_C})\}), \quad (2)$$

where s_C is stored in the classical memory. In the prediction phase, a sequence of new inputs $\tilde{x}_1, \tilde{x}_2, \dots \in \{0, 1\}^n$ is provided. The ML model will use s_C to evaluate predictions $h_C(\tilde{x}_1), h_C(\tilde{x}_2), \dots$ that approximate $f(\tilde{x}_1), f(\tilde{x}_2), \dots$ up to small errors.

b. Restricted classical ML: We will also consider a restricted version of the classical setting. Rather than performing arbitrary POVM measurements, we restrict the ML model to measure the target observable O on the output state $\mathcal{E}|x_i\rangle\langle x_i|$ to obtain the

measurement outcome o_i . In this case, we always have $o_i \in \mathbb{R}$ and $\mathbb{E}[o_i] = \text{tr}(O\mathcal{E}(|x_i\rangle\langle x_i|))$.

c. Quantum (Q) ML: During the learning phase, the model starts with an initial state ρ_0 in a Hilbert space of arbitrarily high dimension. Subsequently, the quantum ML model accesses the unknown CPTP map \mathcal{E} a total of N_Q times. These queries are interleaved with quantum data processing steps:

$$\rho_{\mathcal{E}} = \mathcal{C}_{N_Q}(\mathcal{E} \otimes \mathcal{I}) \mathcal{C}_{N_Q-1} \dots \mathcal{C}_1(\mathcal{E} \otimes \mathcal{I})(\rho_0), \quad (3)$$

where each \mathcal{C}_i is an arbitrary but known CPTP map, and we write $\mathcal{E} \otimes \mathcal{I}$ to emphasize that \mathcal{E} acts on an n -qubit subsystem of a larger quantum system. The final state $\rho_{\mathcal{E}}$, encoding the prediction model learned from the queries to the unknown CPTP map \mathcal{E} , is stored in a quantum memory. In the prediction phase, a sequence of new inputs $\tilde{x}_1, \tilde{x}_2, \dots \in \{0, 1\}^n$ is provided. A quantum computer with access to the stored quantum state $\rho_{\mathcal{E}}$ executes a computation to produce prediction values $h_Q(\tilde{x}_1), h_Q(\tilde{x}_2), \dots$ that approximate $f(\tilde{x}_1), f(\tilde{x}_2), \dots$ up to small errors¹.

The quantum ML setting is strictly more powerful than the classical ML setting. During the prediction phase, classical ML models are restricted to processing classical data, albeit data obtained by measuring a quantum system during the learning phase. In contrast, quantum ML models can work directly with the quantum data and perform quantum data processing. A quantum ML model can have an exponential advantage relative to classical ML models for some tasks, as we demonstrate in Sec. IV.

III. AVERAGE-CASE PREDICTION ERROR

For a prediction model $h(x)$, we consider the average-case prediction error

$$\sum_{x \in \{0,1\}^n} \mathcal{D}(x) |h(x) - \text{tr}(O\mathcal{E}(|x\rangle\langle x|))|^2, \quad (4)$$

with respect to a fixed distribution \mathcal{D} over inputs. This could, for instance, be the uniform distribution.

Although learning from quantum data is strictly more powerful than learning from classical data, there are fundamental limitations. The following rigorous statement limits the potential for quantum advantage.

Theorem 1. *Fix an n -bit probability distribution \mathcal{D} , an m -qubit observable O ($\|O\| \leq 1$) and a set \mathcal{F} of CPTP maps with n input qubits and m output qubits. Suppose there is a quantum ML model which accesses the map $\mathcal{E} \in \mathcal{F}$ N_Q times, producing with high probability a function $h_Q(x)$ that achieves*

$$\sum_{x \in \{0,1\}^n} \mathcal{D}(x) |h_Q(x) - \text{tr}(O\mathcal{E}(|x\rangle\langle x|))|^2 \leq \epsilon. \quad (5)$$

¹ Due to non-commutativity of quantum measurements, the ordering of new inputs matters. For instance, the two lists \tilde{x}_1, \tilde{x}_2 and \tilde{x}_2, \tilde{x}_1 can lead to different outcome predictions $h_Q(\tilde{x}_i)$. Our main results do not depend on this subtlety — they are valid, irrespective of prediction input ordering.

Then there is an ML model in the restricted classical setting which accesses \mathcal{E} $N_C = \mathcal{O}(mN_Q/\epsilon)$ times and produces with high probability a function h_C that achieves

$$\sum_{x \in \{0,1\}^n} \mathcal{D}(x) |h_C(x) - \text{tr}(O\mathcal{E}(|x\rangle\langle x|))|^2 = \mathcal{O}(\epsilon). \quad (6)$$

Proof sketch. The proof consists of two parts. First, we cover the entire set of CPTP maps \mathcal{F} with a maximal packing net, i.e. the largest subset $\mathcal{S} = \{\mathcal{E}_s\}_{s=1}^{|\mathcal{S}|} \subset \mathcal{F}$ such that the functions $f_{\mathcal{E}_s}(x) = \text{tr}(O\mathcal{E}_s(|x\rangle\langle x|))$ obey $\sum_{x \in \{0,1\}^n} \mathcal{D}(x) |f_{\mathcal{E}_s}(x) - f_{\mathcal{E}_{s'}}(x)|^2 > 4\epsilon$ whenever $s \neq s'$. We then set up a communication protocol as follows. Alice chooses an element s of the packing net uniformly at random, records her choice s , and then applies \mathcal{E}_s N_Q times to prepare a quantum state $\rho_{\mathcal{E}_s}$ as in Eq. (3). Alice's random ensemble of quantum states is thus given by

$$\rho_{\mathcal{E}_s} \text{ with probability } p_s = \frac{1}{|\mathcal{S}|} \quad (7)$$

for $s = 1, \dots, |\mathcal{S}|$. Alice then sends the randomly sampled quantum state $\rho_{\mathcal{E}_s}$ to Bob, hoping that Bob can decode the state $\rho_{\mathcal{E}_s}$ to recover her chosen message s . Using the quantum ML model, Bob can produce the function $h_{Q,s}(x)$. Because by assumption the function $h_{Q,s}(x)$ achieves a small average-case prediction error with high probability, and because the packing net has been constructed so that the functions $\{f_{\mathcal{E}_s}\}$ are sufficiently distinguishable, Bob can determine s successfully with high probability. Because Alice chose from among $|\mathcal{S}|$ possible messages, the mutual information of the chosen message s and Bob's measurement outcome must be at least of order $\log |\mathcal{S}|$ bits. According to Holevo's theorem, the Holevo χ quantity of Alice's ensemble Eq. (7) upper bounds this mutual information, and therefore must also be $\chi = \Omega(\log |\mathcal{S}|)$. Furthermore, we can analyze how χ depends on N_Q , finding that each additional application of \mathcal{E}_s can increase χ by at most $\mathcal{O}(m)$. We conclude that $\chi = \mathcal{O}(mN_Q)$, yielding the lower bound $N_Q = \Omega(\log(|\mathcal{S}|)/m)$. The lower bound applies to any quantum ML model, where the size $|\mathcal{S}|$ of the packing net depends on the average-case prediction error ϵ . This completes the first part of the proof.

In the second part, we explicitly construct an ML model in the restricted classical setting that achieves a small average-case prediction error using a modest number of experiments. In this ML model, an input x_i is selected by sampling from the probability distribution \mathcal{D} , and an experiment is performed in which the observable O is measured in the output quantum state $\mathcal{E}(|x_i\rangle\langle x_i|)$, obtaining measurement outcome o_i which has expectation value $\text{tr}(O\mathcal{E}(|x_i\rangle\langle x_i|))$. A total of N_C such experiments are conducted. Then, the ML model minimizes the least-squares error to find the best fit within the aforementioned maximal packing net \mathcal{S} :

$$h_C = \arg \min_{f \in \mathcal{S}} \frac{1}{N_C} \sum_{i=1}^{N_C} |f(x_i) - o_i|^2. \quad (8)$$

Because the measurement outcome o_i fluctuates about the expectation value of O , it may be impossi-

ble to achieve zero training error. Yet it is still possible for h_C to achieve a small average-case prediction error, potentially even smaller than the training error. We use properties of maximal packing nets and of quantum fluctuations of measurement outcomes to perform a tight statistical analysis of the average-case prediction error, finding that with high probability $\sum_{x \in \{0,1\}^n} \mathcal{D}(x) |h_C(x) - \text{tr}(O\mathcal{E}(|x\rangle\langle x|))|^2 = \mathcal{O}(\epsilon)$, provided that N_C is of order $\log(|\mathcal{S}|)/\epsilon$.

Finally, we combine the two parts to conclude $N_C = \mathcal{O}(mN_Q/\epsilon)$. The full proof is in Appendix ??.

Theorem 1 shows that all problems that are approximately learnable by a quantum ML model are also approximately learnable by some restricted classical ML model which executes the quantum process \mathcal{E} a comparable number of times. This applies in particular, to predicting outputs of quantum-mechanical processes. The relation $N_C = \mathcal{O}(mN_Q/\epsilon)$ is tight. We give an example in Appendix ?? with $N_C = \Omega(mN_Q/\epsilon)$.

For the task of learning classical Boolean circuits, fundamental limits on quantum advantage have been established in previous work [11–13, 31, 80, 94]. Theorem 1 generalizes these existing results to the task of learning outcomes of quantum processes.

IV. WORST-CASE PREDICTION ERROR

Rather than achieving a small average prediction error, one may be interested in obtaining a prediction model that is accurate for all inputs $x \in \{0,1\}^n$. For a prediction model $h(x)$, we consider the worst-case prediction error to be

$$\max_{x \in \{0,1\}^n} |h(x) - \text{tr}(O\mathcal{E}(|x\rangle\langle x|))|^2. \quad (9)$$

Under such a stricter performance requirement, exponential quantum advantage becomes possible.

We highlight this potential by means of an illustrative and practically relevant example: predicting expectation values of Pauli operators in an unknown n -qubit quantum state ρ . This is a central task for many quantum computing applications [34, 52, 54, 56–58, 61, 74]. To formulate this problem in our framework, suppose the $2n$ -bit input x specifies one of the 4^n n -qubit Pauli operators $P_x \in \{I, X, Y, Z\}^{\otimes n}$, and suppose that $\mathcal{E}_\rho(|x\rangle\langle x|)$ prepares the unknown state ρ and maps P_x to the fixed observable O , which is then measured; hence

$$f(x) = \text{tr}(O\mathcal{E}_\rho(|x\rangle\langle x|)) = \text{tr}(P_x\rho). \quad (10)$$

In this setting, according to Theorem 1, there is no large quantum advantage if our goal is to estimate the Pauli operator expectation values with a small *average* prediction error. However, an exponential quantum advantage is possible if we insist on accurately predicting *every one* of the 4^n Pauli observables.

First we show there is an efficient quantum ML model that achieves a small prediction error. Details are in Appendix ??; here we just sketch the main ideas. The procedure for predicting $\text{tr}(P_x\rho)$

has two stages. The goal of the first stage is to predict the absolute value $|\text{tr}(P_x\rho)|$ for each x , and the goal of the second stage is to determine the *sign* of $\text{tr}(P_x\rho)$. The key idea used in the first stage is that, although two different Pauli operators P_x and P_y may either commute or anticommute, the tensor products $P_x \otimes P_x$ and $P_y \otimes P_y$ are mutually commuting for all x and y . Therefore, although it is not possible to measure anticommuting Pauli operators simultaneously using a single copy of the state ρ , it is possible to measure $P_x \otimes P_x$ simultaneously for all x using two copies of ρ . Indeed, all 4^n expectation values $\text{tr}((P_x \otimes P_x)(\rho \otimes \rho)) = \text{tr}(P_x\rho)^2$ can be determined by measuring pairs of qubits in the Bell basis, which is highly efficient. This completes the first stage.

If $|\text{tr}(P_x\rho)|$ is found to be small in the first stage, we may predict $h(x) = 0$ and be assured that the prediction error is small. Therefore, in the second stage, we need only determine the sign if $|\text{tr}(P_x\rho)|$ was found to be reasonably large in the first stage. In that case we can perform a coherent measurement across several copies of ρ which performs a majority vote and yields the correct value of the sign with high success probability. Because the measurement is strongly biased in favor of one of the two possible outcomes, it introduces only a very “gentle” disturbance of the pre-measurement state. Therefore, by performing many such measurements in succession on the same quantum memory register, we can determine the sign of $\text{tr}(P_x\rho)$ for many different values of x . The second stage can also be more amenable to near-term implementation using a heuristic that groups commuting observables [56, 57]; see Appendix ?? for further discussion. Each of the two stages requires only a small number of copies of ρ ; a careful analysis yields the following theorem.

Theorem 2. *The quantum ML model only needs $N_Q = \mathcal{O}(\log(M/\delta)/\epsilon^4)$ copies of ρ to predict expectation values of any M Pauli observables to error ϵ with probability at least $1 - \delta$.*

More details regarding the quantum ML model, as well as a rigorous proof, are provided in Appendix ??. The sample complexity stated in Theorem 2 improves upon previously known shadow tomography protocols [4, 5, 26, 54] for the special case of predicting Pauli observables; see Appendix ??. Because each access to \mathcal{E}_ρ allows us to obtain one copy of ρ , we only need $N_Q = \mathcal{O}(n)$ to predict expectation values of all 4^n Pauli observables up to a constant error.

For classical ML models, we prove the following fundamental lower bound; see Appendix ??.

Theorem 3. *Any classical ML must use $N_C \geq 2^{\Omega(n)}$ copies of ρ to predict expectation values of all Pauli observables up to a small error with a constant success probability.*

This theorem holds even when the POVM measurements performed by the classical ML model could depend on the previous POVM measurement outcomes adaptively. When combined with Theorem 2, Theorem 3 establishes an exponential gap separating classical ML models from fully quantum ML models. Table 2 provides a summary of the upper and lower

Model	Upp. bd.	Low. bd.
Quantum ML	$\mathcal{O}(n)$	$\Omega(n)$
Classical ML	$\mathcal{O}(n^{2^n})$	$2^{\Omega(n)}$
Restricted CML	$\mathcal{O}(2^{2^n})$	$\Omega(2^{2^n})$

Figure (Table) 2: Sample complexity for predicting expectations of *all* 4^n Pauli observables (worst-case prediction error) in an n -qubit quantum state. *Upp. bd.* is the achievable sample complexity of a specific algorithm. *Low. bd.* is the lower bound for any algorithm. The classical ML upper bound can be achieved using classical shadows based on random Clifford measurements [54]. The rest of the bounds are obtained in Appendix ??.

bounds on the sample complexity for predicting expectation values of Pauli observables.

V. NUMERICAL EXPERIMENTS

We support our theoretical findings with numerical experiments, focusing on the task of predicting the expectation values of all 4^n Pauli observables in an unknown n -qubit quantum state ρ , with a small worst-case prediction error. In this case, the function is $f(x) = \text{tr}(O \mathcal{E}_\rho(|x\rangle\langle x|)) = \text{tr}(P_x \rho)$, where $x \in \{I, X, Y, Z\}^n$ indexes the Pauli observables, and \mathcal{E}_ρ prepares the unknown state ρ then maps P_x to the fixed observable O . This is the task we considered in Section IV. Note that average-case prediction of Pauli observables is a much easier task, because most of the 4^n expectation values are exponentially small in n .

We consider two classes of underlying states ρ : (i) *Mixed states*: $\rho = (I + P)/2^n$, where P is a tensor product of n Pauli operators. States in this class have rank 2^{n-1} . (ii) *Product states*: $\rho = \bigotimes_{i=1}^n |s_i\rangle\langle s_i|$, where each $|s_i\rangle$ is one of the six possible single-qubit stabilizer states. We consider stabilizer states to ensure that classical simulation of the quantum ML model is tractable for reasonably large system size.

The numerical experiment in Figure 3 implements the best-known ML procedures. We can clearly see that there is an exponential separation between the number of copies of the state ρ required for classical and quantum ML to predict expectation values when ρ is in the class of mixed states. However, for the class of product states, the separation is much less pronounced. Restricted classical ML can only obtain outcomes $o_i \in \{\pm 1\}$ with $\mathbb{E}[o_i] = \text{tr}(P_{x_i} \rho)$. Hence each copy of ρ provides at most one bit of information, and therefore $\mathcal{O}(n)$ copies are needed to predict expectation values of all 4^n Pauli observables. In contrast, standard classical ML can perform arbitrary POVM measurements on the state ρ , so each copy can provide up to n bits of information. The separation between classical ML and quantum ML is marginal for product states.

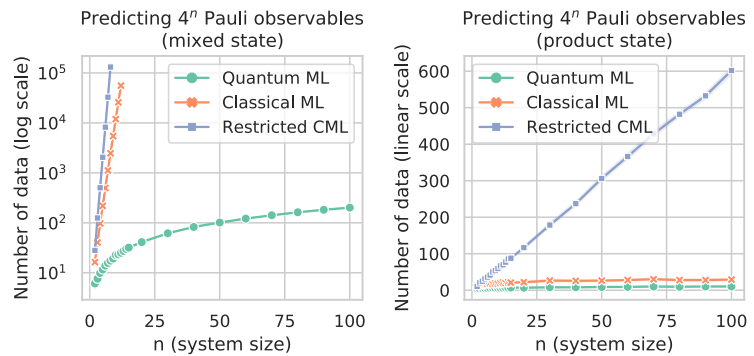


Figure 3: Numerical experiments – number of copies of an unknown n -qubit state needed for predicting expectation values of all 4^n Pauli observables, with constant worst-case prediction error. *Mixed states*: Quantum states of the form $(I + P)/2^n$, where P is an n -qubit Pauli observable. *Product states*: Tensor products of single-qubit stabilizer states.

VI. CONCLUSION AND OUTLOOK

We have studied the task of learning functions of the form Equation (1), using as a figure of merit the number of runs of \mathcal{E} . Our main result Theorem 1 shows that, when the objective is achieving a specified *average* prediction error, a classical ML model can perform as well as a quantum ML model, using a comparable number of runs of \mathcal{E} . This result establishes a fundamental limit on quantum advantage in machine learning that holds for any quantum ML model [47, 53, 79].

From a different perspective, Theorem 1 means that the classical ML setting, in which a measurement is performed after each query to \mathcal{E} , can be surprisingly effective. The quantum ML setting, in which multiple queries to \mathcal{E} can be included in a single coherent quantum circuit, is far more challenging and may be infeasible until far in the future. Therefore finding that classical and quantum ML have comparable power (for average-case prediction) boosts our hopes that the combination of classical ML and near-term quantum algorithms [47, 52, 53, 76, 79] may fruitfully address challenging quantum problems in physics, chemistry, and materials science.

On the other hand, Theorem 2 and 3 rigorously establish that quantum ML can have an exponential advantage over classical ML for certain problems where the objective is achieving a specified *worst-case* prediction error. This exponential advantage of quantum ML over classical ML may be viewed as an exponential separation between coherent measurements (in which a measurement apparatus interacts coherently multiple times with a measured system, storing quantum data which is then processed by a quantum computer) and incoherent measurements (in which a POVM measurement is performed and the outcome recorded after each interaction between system and apparatus, and the classical measurement outcomes are then processed by a classical computer). Such a separation has been challenging to establish because incoherent measurements are difficult to analyze in the adaptive setting, where each measurement performed may depend on the outcomes of all previous measurements. Our proof technique overcomes this

challenge, enabling us to identify tasks which allow substantial quantum advantage. An important future direction will be identifying further learning problems which allow substantial quantum advantage, pointing toward potential practical applications of quantum technology.

Acknowledgments:

The authors thank Victor Albert, Sitan Chen, Jerry Li, Seth Lloyd, Jarrod McClean, Spiros Michalakis, Yuan Su, and Thomas Vidick for valuable input and

inspiring discussions. We would also like to thank anonymous reviewers for in-depth comments and suggestions. HH is supported by the J. Yang & Family Foundation. JP acknowledges funding from the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, (DE-NA0003525, DE-SC0020290), and the National Science Foundation (PHY-1733907). The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center.

-
- [1] S. Aaronson. Qma/qpoly/spl sube/pspace/poly: demerlinizing quantum protocols. In *CCC*, pages 13–pp, 2006.
- [2] S. Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007.
- [3] S. Aaronson. Read the fine print. *Nat. Phys.*, 11(4):291–293, 2015.
- [4] S. Aaronson. Shadow tomography of quantum states. In *STOC*, pages 325–338, 2018.
- [5] S. Aaronson and G. N. Rothblum. Gentle measurement of quantum states and differential privacy. In *STOC*, pages 322–333, 2019.
- [6] D. Aharonov, J. Cotler, and X.-L. Qi. Quantum algorithmic measurement. *arXiv preprint arXiv:2101.04634*, 2021.
- [7] N. Alon, S. Ben-David, N. Cesa-Bianchi, and D. Haussler. Scale-sensitive dimensions, uniform convergence, and learnability. *Journal of the ACM (JACM)*, 44(4):615–631, 1997.
- [8] A. Ambainis and J. Emerson. Quantum t-designs: t-wise independence in the quantum world. In *CCC*, pages 129–140, 2007.
- [9] H. Araki and E. H. Lieb. Entropy inequalities. In *Inequalities*, pages 47–57. Springer, 2002.
- [10] J. M. Arrazola, A. Delgado, B. R. Bardhan, and S. Lloyd. Quantum-inspired algorithms in practice. *arXiv preprint arXiv:1905.10415*, 2019.
- [11] S. Arunachalam and R. de Wolf. Optimal quantum sample complexity of learning algorithms. *arXiv preprint arXiv:1607.00932*, 2016.
- [12] S. Arunachalam and R. de Wolf. Guest column: A survey of quantum learning theory. *ACM SIGACT News*, 48(2):41–67, 2017.
- [13] S. Arunachalam, A. B. Grilo, and H. Yuen. Quantum statistical query learning. *arXiv preprint arXiv:2002.08240*, 2020.
- [14] P. L. Bartlett and P. M. Long. Prediction, learning, uniform convergence, and scale-sensitive dimensions. *Journal of Computer and System Sciences*, 56(2):174–190, 1998.
- [15] P. L. Bartlett and S. Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *J Mach Learn Res*, 3(Nov):463–482, 2002.
- [16] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. De Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.
- [17] A. D. Becke. A new mixing of hartree–fock and local density-functional theories. *J. Chem. Phys.*, 98(2):1372–1377, 1993.
- [18] I. Bengtsson and K. Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge university press, 2017.
- [19] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature*, 551(7682):579–584, 2017.
- [20] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
- [21] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *J. ACM*, 36(4):929–965, 1989.
- [22] X. Bonet-Monroig, R. Babbush, and T. E. O’Brien. Nearly optimal measurement scheduling for partial tomography of quantum states. *Phys. Rev. X*, 10(3):031064, 2020.
- [23] F. G. Brandão, A. Kalev, T. Li, C. Y.-Y. Lin, K. M. Svore, and X. Wu. Quantum SDP solvers: Large speed-ups, optimality, and applications to quantum learning. In *ICALP*, 2019.
- [24] S. Bubeck, S. Chen, and J. Li. Entanglement is necessary for optimal quantum property testing. *arXiv preprint arXiv:2004.07869*, 2020.
- [25] I. Buluta and F. Nori. Quantum simulators. *Science*, 326(5949):108–111, 2009.
- [26] C. Bădescu and R. O’Donnell. Improved quantum data analysis. *arXiv preprint arXiv:2011.10908*, 2020.
- [27] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55(22):2471, 1985.
- [28] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.
- [29] M. C. Caro. Binary classification with classical instances and quantum labels. *arXiv preprint arXiv:2006.06005*, 2020.
- [30] J. Carrasquilla and R. G. Melko. Machine learning phases of matter. *Nat. Phys.*, 13(5):431–434, 2017.
- [31] K.-M. Chung and H.-H. Lin. Sample efficient algorithms for learning quantum channels in pac model and the approximate state discrimination problem. *arXiv preprint arXiv:1810.10938*, 2018.
- [32] I. Cong, S. Choi, and M. D. Lukin. Quantum convolutional neural networks. *Nat. Phys.*, 15(12):1273–1278, 2019.
- [33] J. Cotler, S. Choi, A. Lukin, H. Gharibyan, T. Grover, M. E. Tai, M. Rispoli, R. Schittko, P. M. Preiss, A. M. Kaufman, et al. Quantum virtual cooling. *Phys. Rev. X*, 9(3):031013, 2019.
- [34] O. Crawford, B. van Straaten, D. Wang, T. Parks, E. Campbell, and S. Brierley. Efficient quantum measurement of pauli operators in the presence of finite sampling error. *arXiv preprint arXiv:1908.06942*,

- 2020.
- [35] V. Dunjko and H. J. Briegel. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Rep. Prog. Phys.*, 81(7):074001, 2018.
- [36] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch. Mixed-state entanglement from local randomized measurements. *Phys. Rev. Lett.*, 125:200501, Nov 2020.
- [37] T. J. Evans, R. Harper, and S. T. Flammia. Scalable bayesian hamiltonian learning. *arXiv preprint arXiv:1912.07636*, 2019.
- [38] E. Farhi and H. Neven. Classification with quantum neural networks on near term processors. *arXiv preprint arXiv:1802.06002*, 2018.
- [39] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New J. Phys.*, 14(9):095022, 2012.
- [40] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017.
- [41] A. Gilyén, S. Lloyd, and E. Tang. Quantum-inspired low-rank stochastic regression with logarithmic dependence on the dimension. *arXiv preprint arXiv:1811.04909*, 2018.
- [42] D. Gross, S. Nezami, and M. Walter. Schur-Weyl duality for the Clifford group with applications: Property testing, a robust Hudson theorem, and de Finetti representations. *arXiv preprint arXiv:1712.08628*, 2017.
- [43] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. *IEEE Trans. Inf. Theory*, 63(9):5628–5641, 2017.
- [44] I. Hamamura and T. Imamichi. Efficient evaluation of quantum observables using entangled measurements. *npj Quantum Inf.*, 6(1):1–8, 2020.
- [45] R. Harper, W. Yu, and S. T. Flammia. Fast estimation of sparse quantum noise. *arXiv preprint arXiv:2007.07901*, 2020.
- [46] A. W. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103(15):150502, 2009.
- [47] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.
- [48] C. W. Helstrom. Quantum detection and estimation theory. *J. Statist. Phys.*, 1:231–252, 1969.
- [49] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973.
- [50] A. S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3:337–394, 1973.
- [51] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81(2):865, 2009.
- [52] H.-Y. Huang, K. Bharti, and P. Rebentrost. Near-term quantum algorithms for linear systems of equations. *arXiv preprint arXiv:1909.07344*, 2019.
- [53] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean. Power of data in quantum machine learning. *arXiv preprint arXiv:2011.01938*, 2020.
- [54] H.-Y. Huang, R. Kueng, and J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.*, 16:1050–1057, 2020.
- [55] W. J. Huggins, S. McArdle, T. E. O’Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean. Virtual distillation for quantum error mitigation. *arXiv preprint arXiv:2011.07064*, 2020.
- [56] W. J. Huggins, J. McClean, N. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *arXiv preprint arXiv:1907.13117*, 2019.
- [57] A. F. Izmaylov, T.-C. Yen, R. A. Lang, and V. Verteletskyi. Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method. *J. Chem. Theory Comput.*, 16(1):190–195, 2019.
- [58] Z. Jiang, A. Kalev, W. Mroczkiewicz, and H. Neven. Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning. *Quantum*, 4:276, 2020.
- [59] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.
- [60] M. J. Kearns and R. E. Schapire. Efficient distribution-free learning of probabilistic concepts. *Journal of Computer and System Sciences*, 48(3):464–497, 1994.
- [61] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, et al. Self-verifying variational quantum simulation of lattice models. *Nature*, 569(7756):355–360, 2019.
- [62] R. Kueng. Quantum and classical information processing with tensors (lecture notes), Spring 2019. Caltech course notes: <https://iqim.caltech.edu/classes>.
- [63] H. Levine, A. Keesling, A. Omran, H. Bernien, S. Schwartz, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin. High-fidelity control and entanglement of rydberg-atom qubits. *Phys. Rev. Lett.*, 121(12):123603, 2018.
- [64] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. *arXiv preprint arXiv:1307.0411*, 2013.
- [65] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum principal component analysis. *Nat. Phys.*, 10(9):631–633, 2014.
- [66] W. Matthews, S. Wehner, and A. Winter. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Comm. Math. Phys.*, 291(3):813–843, 2009.
- [67] A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel. Active learning machine learns to create new quantum experiments. *Proc. Natl. Acad. Sci. U.S.A.*, 115(6):1221–1226, 2018.
- [68] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. The MIT Press, 2018.
- [69] A. Montanaro. Learning stabilizer states by bell sampling. *arXiv preprint arXiv:1707.04012*, 2017.
- [70] A. Montanaro and R. de Wolf. A survey of quantum property testing. *Theory Comput.*, pages 1–81, 2016.
- [71] M. Y. Niu, A. M. Dai, L. Li, A. Odena, Z. Zhao, V. Smelyanskyi, H. Neven, and S. Boixo. Learnability and complexity of quantum samples. *arXiv preprint arXiv:2010.11983*, 2020.
- [72] M. Painsi and A. Kalev. An approximate description of quantum states. *arXiv preprint arXiv:1910.10543*, 2019.

- [73] R. G. Parr. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry*, pages 5–15. Springer, 1980.
- [74] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.*, 5:4213, 2014.
- [75] K. Poland, K. Beer, and T. J. Osborne. No free lunch for quantum machine learning. *arXiv preprint arXiv:2003.14103*, 2020.
- [76] J. Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.
- [77] P. Rebentrost, M. Mohseni, and S. Lloyd. Quantum support vector machine for big data classification. *Phys. Rev. Lett.*, 113(13):130503, 2014.
- [78] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, and F. Sciarrino. Experimental learning of quantum states. *Science advances*, 5(3):eaau1946, 2019.
- [79] M. Schuld and N. Killoran. Quantum machine learning in feature hilbert spaces. *Phys. Rev. Lett.*, 122(4):040504, 2019.
- [80] R. A. Servedio and S. J. Gortler. Equivalences and separations between quantum and classical learnability. *SIAM J. Comput.*, 33(5):1067–1092, 2004.
- [81] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. *Phys. Rev. Lett.*, 124(2):020503, 2020.
- [82] K. Sharma, M. Cerezo, Z. Holmes, L. Cincio, A. Sornborger, and P. J. Coles. Reformulation of the no-free-lunch theorem for entangled data sets. *arXiv preprint arXiv:2007.04900*, 2020.
- [83] G. Struchalin, Y. A. Zagorovskii, E. Kovalkov, S. Straupe, and S. Kulik. Experimental estimation of quantum state properties from classical shadows. *arXiv preprint arXiv:2008.05234*, 2020.
- [84] R. Sweke, J.-P. Seifert, D. Hangleiter, and J. Eisert. On the quantum versus classical learnability of discrete distributions. *arXiv preprint arXiv:2007.14451*, 2020.
- [85] E. Tang. Quantum-inspired classical algorithms for principal component analysis and supervised clustering. *arXiv preprint arXiv:1811.00414*, 2018.
- [86] E. Tang. A quantum-inspired classical algorithm for recommendation systems. In *STOC*, pages 217–228, 2019.
- [87] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.
- [88] E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber. Learning phase transitions by confusion. *Nat. Phys.*, 13(5):435–439, 2017.
- [89] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.
- [90] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- [91] V. Verteletskyi, T.-C. Yen, and A. F. Izmaylov. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *J. Chem. Phys.*, 152(12):124114, 2020.
- [92] S. R. White. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B*, 48(14):10345, 1993.
- [93] N. Wiebe, D. Braun, and S. Lloyd. Quantum algorithm for data fitting. *Phys. Rev. Lett.*, 109(5):050505, 2012.
- [94] C. Zhang. An improved lower bound on query complexity for quantum pac learning. *Inf. Process. Lett.*, 111(1):40–45, 2010.
- [95] L. Zhao, C. A. Pérez-Delgado, and J. F. Fitzsimons. Fast graph operations in quantum computation. *Phys. Rev. A*, 93(3):032314, 2016.
- [96] Z. Zhou, X. Li, and R. N. Zare. Optimizing chemical reactions with deep reinforcement learning. *ACS Cent. Sci.*, 3(12):1337–1344, 2017.