

Quantum machine learning advantages beyond hardness of evaluation

Riccardo Molteni¹, Simon C. Marshall¹, and Vedran Dunjko¹

¹applied Quantum algorithms (aQa), Leiden University, The Netherlands

Link at the full paper: [ICLR](#), [arXiv:2504.15964](#)

1 Extended abstract

A central question in quantum machine learning (QML) is understanding which learning problems truly benefit from quantum computation. Several works have established formal quantum learning advantages [LAT21, SG04], particularly for problems involving cryptographic functions, which critically rely on their so-called “random generatability” property, introduced later. More recently, [GD23, MGD24] demonstrated exponential quantum–classical separations for target functions arising from genuinely quantum problems, leveraging the requirement in standard PAC-learning definitions that the learned model must *evaluate* new data points. An open question for years was whether the random generatability property is truly necessary to obtain hardness results in the more general setting where the learner is only required to *identify* the target labeling function from labeled data—that is, to output a description of the unknown target function without necessarily being able to evaluate it on unseen inputs. The identification task is especially relevant in physical applications such as Hamiltonian learning and discovering order parameters, where recovering the description of the process that generated the data is the ultimate goal. From a foundational viewpoint, studying the hardness of the identification helps isolate the true sources of learning advantage in quantum settings. Crucially, it remained unclear whether hardness results for the identification task could be established for quantum problems whose target functions lie in $\text{BQP} \setminus (\text{BPP} \cup \text{NP})$ (the functions that, under standard assumptions, are genuinely quantum and thus were conjectured not to have the generatability property), since Hamiltonian learning, for example, admits an efficient polynomial-time classical algorithm. In this work, we address this question and show that learning advantages for the identification task involving BQP functions can indeed be established in rather generally setting under certain complexity-theoretic assumptions that we discuss in detail in the paper. Concretely our contributions are (we [highlight](#) the new contributions present in this year’s updated version of the paper):

1. We show that quantum functions are not random generatable unless BQP is in the second level of the polynomial hierarchy. This result now formally rules out the possibility of directly extending the proof techniques used for cryptographic functions to establish hardness in the identification setting for BQP functions. Furthermore, it has consequences on quantum generative modeling as well, see Corollary 1.
2. We introduce the task of verifiable-identification, where the algorithm must reject invalid datasets, and solve identification correctly otherwise, and prove classical verifiable-identification for quantum target function is impossible, unless $\text{BQP} \not\subseteq \text{HeurBPP}^{\text{NP}}$ if the learning algorithm only needs to work for a fixed distribution. [In case the learner is required to succeed for every input distributions, we show that classical hardness holds under the more standard assumption \$\text{BQP} \not\subseteq \text{BPP}^{\text{NP}}\$.](#)
3. We identify sufficient conditions on learning algorithm and concept class such that “mere” classical identification implies approximate-verifiable identification (the algorithm must reject datasets unless most of the samples are labeled according to the same concept) within the polynomial hierarchy. This shows that identification is classically intractable, [assuming \$\text{BQP} \not\subseteq \text{BPP}^{\text{NP}^{\text{NP}}}\$ when the learner must succeed for every input distribution](#), or assuming $\text{BQP} \not\subseteq \text{HeurBPP}^{\text{NP}}$ for a fixed distribution.
4. We provide a physical examples of learning tasks which satisfies the above condition, for which then the quantum advantage for identification tasks holds, see Theorem 3.
5. [We provide a list of physical scenarios in which quantum functions arise naturally, together with a list of practical tasks that can be captured within our framework for the identification task in Appendix G4. We discuss the assumptions required of our learning algorithm in Appendix H and explain how they relate to the framework of proper PAC learning. In Appendix F2, we construct a concept class of functions satisfying the average-case smoothness condition required for our hardness result together with a simple circuit implementation. This helps identify physically relevant tasks where our results can be applied.](#)

1.1 Quantum functions

We report here the definition of quantum functions adopted in this work. For the definitions of the complexity classes involved in our results we refer to the full manuscript.

Definition 1 (BQP function f). We refer to a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ as a BQP function if there exists a language $\mathcal{L} \in \text{BQP}$ such that f is its characteristic function. In particular:

$$f(x) = \begin{cases} f(x) = 0 & \text{if } x \notin \mathcal{L} \\ f(x) = 1 & \text{if } x \in \mathcal{L} \end{cases} . \quad (1)$$

1.2 Hardness of random generatability

Given a uniform family of functions $f = \{f_n\}_{n \in \mathbb{N}}$, with $f_n : \{0, 1\}^n \rightarrow \{0, 1\}$, we say that f is “random-generable” under a distribution $\mathcal{D}$ if there exists a uniform (randomized) algorithm capable of producing samples $(x, f_n(x))$ with x sampled from $\mathcal{D}$ efficiently for any n . Previous classical identification-hardness proofs for learning tasks involving cryptographic functions often hinge on random generatability, since it implies that labeled data can be sampled efficiently and therefore offers no help in evaluating the target function. Here, we prove that quantum functions violate this, necessitating a completely new proof strategy. Specifically, we consider two cases: exact random generatability, where the algorithm is not allowed to make any errors, and approximate random generatability, where the algorithm outputs samples from a distribution only close to the true target distribution. We present here our hardness results only for the more general case of approximate random generatability.

Definition 2 (Approximate random generatability). Let $f = \{f_n\}_{n \in \mathbb{N}}$ be a uniform family of functions, with $f_n : \{0, 1\}^n \rightarrow \{0, 1\}$. f is approximately random generatable under the distribution $\mathcal{D}$ if there exists an efficient uniform (randomized) algorithm $\mathcal{A}_{\mathcal{D}}$ such that given as input a description of f_n for any n , a random string r and an error value ϵ outputs:

$$\mathcal{A}_{\mathcal{D}}(f_n, 0^{1/\epsilon}, r) = (x_r, f_n(x_r)) \quad (2)$$

with $(x_r, f_n(x_r)) \sim \pi_{\epsilon}^f(\mathcal{D})$ when r is sampled uniformly at random from the distribution of all the random strings (with polynomial size). In particular, $\pi_{\epsilon}^f(\mathcal{D})$ is a probability distribution over $\{0, 1\}^n \times \{0, 1\}$ which satisfies $\|\pi_{\epsilon}^f(\mathcal{D}) - \pi^f(\mathcal{D})\|_{TV} \leq \epsilon$, where $\pi^f(\mathcal{D})$ samples x from the target distribution $\mathcal{D}$ over $x \in \{0, 1\}^n$ and assigns the corresponding label $f_n(x_r) \in \{0, 1\}$.

We then show hardness of approximate random generatability of quantum functions based on the assumptions that BQP cannot be heuristically decided within HeurBPP by an efficient classical machine with a NP oracle.

Theorem 1 (Approximate Random generatability implies $(\text{BQP}, \text{Unif}) \in \text{HeurBPP}^{\text{NP}}$). *Let $f = \{f_n\}_{n \in \mathbb{N}}$ be a family of BQP functions which is the characteristic function of a language $\mathcal{L} \in \text{BQP}$ as in Def. 1, and let Unif be the uniform distribution over $\{0, 1\}^n$. If there exists a polynomial time algorithm $\mathcal{A}$ which satisfies Def. 2 for the uniform input distribution Unif, then $(\mathcal{L}, \text{Unif}) \in \text{HeurBPP}^{\text{NP}}$.*

Proof sketch. The full proof can be found in the full paper, here we outline the main idea. Given the algorithm $\mathcal{A}$, we present an algorithm $\mathcal{A}' \in \text{HeurBPP}^{\text{NP}}$ which can heuristically decide the language $\mathcal{L}$ with respect to the uniform distribution over the inputs. Let $f = \{f_n\}_n$ be the family of functions associated to the BQP language as in Def. 1. For a fixed function f_n and error parameter ϵ , the algorithm $\mathcal{A}(f_n, 0^{1/\epsilon}, \cdot)$ in Def. 2 maps random strings $r \in \{0, 1\}^{\text{poly}(n)}$ to tuples $(x_r, f_n(x_r)) \in \{0, 1\}^n \times \{0, 1\}$. The key idea is that, on an input $x_{\tilde{r}}$, $\mathcal{A}'$ inverts the algorithm $\mathcal{A}$ in order to obtain a random string $\tilde{r}$ such that $\mathcal{A}(f_n, 0^{1/\epsilon}, \tilde{r}) = (x_{\tilde{r}}, f_n(x_{\tilde{r}}))$. Because $\mathcal{A}$ is efficient, this inversion can be done in the first layer of the polynomial hierarchy yielding the desired conclusion; see the full version for complete details. $\square$

The result has an interesting Corollary, which proves that a certain class of quantum generative models called expectation value samplers (introduced in [RAG21], and proven universal in [BGV⁺24] and generalized in [SKS⁺24]) is classically intractable under certain plausible complexity theoretic assumptions (see full paper).

Corollary 1. *The expectation value sampling (EVS) generative model, as defined in [BGV⁺24] is classically intractable.*

1.3 Hardness of Identification task

Imagine we are given a machine learning task defined by a concept class $\mathcal{F}$ composed of 2^m target functions, each one specified by a vector $\alpha \in \{0, 1\}^m$ with m scaling polynomially with input size n . Given a training set of inputs labeled by one of these concepts, the goal of the *identification learning algorithm* is to “recognize” the target concept which labeled the data and output the corresponding α . We prove hardness results for two closely related but subtly different versions of identification tasks. In the first one, we assume the learning algorithm can decide if a dataset is *invalid*, i.e. if it is not consistent with any of the concepts in the concept class. Nevertheless, in our main hardness result, we will show how to relax the verifiability condition on the learning algorithm and provide hardness result for the existence of a proper PAC learner algorithm which does not reject any input dataset but satisfies the following additional assumptions in Def. 3. We call such learning algorithm an *approximate-correct* identification algorithm.

Definition 3 (Identification task - non verifiable case). Let $\mathcal{F} = \{f^{\alpha} : \{0, 1\}^n \rightarrow \{0, 1\} \mid \alpha \in \{0, 1\}^m\}$ be a concept class. An *approximate-correct* identification algorithm is a (randomized) algorithm $\mathcal{A}_B$ such that when given as input a set $T = \{(x_{\ell}, y_{\ell})\}_{\ell=1}^B$ of at least B pairs $(x, y) \in \{0, 1\}^n \times \{0, 1\}$, an error parameter $\epsilon > 0$ and a random string $r \in R$ satisfies the definition of a proper PAC learner (see for example [KV94]) along with the following additional properties:

1. If for any α all the $(x_\ell, y_\ell) \in T$ are such that $y_\ell \neq f^\alpha(x_\ell)$ then there exists an ϵ_1 such that for all $\epsilon \leq \epsilon_1$ and all $r \in R$:

$$\mathcal{A}_B(T, \epsilon_1, r) \neq \alpha. \quad (3)$$

In other words, the algorithm will never output a totally incorrect α , i.e. an α inconsistent with all the inputs in the dataset.

2. If $T = \{(x_\ell, y_\ell)\}_{\ell=1}^B$ is composed of different inputs x_ℓ and if there exists an α such that $y_\ell = f^\alpha(x_\ell)$ for all $(x_\ell, y_\ell) \in T$, then there exists a threshold ϵ_2 such that for any $\epsilon \leq \epsilon_2$ there exists at least one $r \in R$ for which:

$$\mathcal{A}_B(T, \epsilon_2, r) = \alpha_2 \quad (4)$$

With the condition: $\mathbb{E}_{x \sim \text{Unif}(\{0,1\}^n)} |f^\alpha(x) - f^{\alpha_2}(x)| \leq \frac{1}{3}$.

Therefore, if the dataset is fully consistent with one of the concept α , then there is at least one random string for which the identification algorithm will output a $\tilde{\alpha}$ closer than $\frac{1}{3}$ in PAC condition to the true labelling α .

We say that $\mathcal{A}_B$ solves the identification task for a concept class $\mathcal{F}$ under the input distribution $\mathcal{D}$ if the algorithm works for any value of $\epsilon, \delta \geq 0$. The required minimum size B of the input set T is assumed to scale as $\text{poly}(n, 1/\epsilon, 1/\delta)$, while the running time of the algorithm scales as $\text{poly}(B, n)$. Moreover, the ϵ_1 and ϵ_2 required values scale at most inverse polynomially with n .

We then show that identification is hard for concept classes satisfying either of two conditions: *i*) they are c -distinct, meaning any pair of functions disagree on at least a c fraction of inputs, or *ii*) they are average-case smooth, meaning the label space has a metric under which PAC-close concepts are also close.

Theorem 2 (Hardness of the identification task - non verifiable case). *Let $\mathcal{F} = \{f^\alpha : \{0,1\}^n \rightarrow \{0,1\} \mid \alpha \in \{0,1\}^m\}$ be a concept class containing at least a BQP function f as in Def. 1 associated to a language $\mathcal{L} \in \text{BQP}$. Assume further that $\mathcal{F}$ is either a c -distinct concept class with $c \geq 1/3$ or an average-case-smooth concept class. If the non-verifiable identification task for $\mathcal{F}$, as defined in Def. 3, is solvable by a classically efficient approximate-correct identification algorithm $\mathcal{A}_B$, then it follows that $(\mathcal{L}, \text{Unif}) \in \text{HeurBPP}^{\text{NP}^{\text{NP}}}$.*

Proof sketch. The proof can be found in the full paper. The intuition is the following. We first show that for c -distinct or average-case-smooth concept classes the existence of an approximate-correct identification algorithm in Def. 3 allows for the construction of an identification algorithm in the first level of the PH which is able to reject any dataset which contains a fraction of inputs greater than $\frac{1}{\beta}$ not labeled by f^α , for any $\beta = \text{poly}(n)$. We then show that if such an algorithm exists for a given β , then for a fixed α , we can invert it by ascending two additional levels in the polynomial hierarchy, allowing us to construct a dataset of inputs that is mostly consistent with f^α . Because each of these datasets will contain only a fraction of $\frac{1}{\beta}$ inputs incorrectly labeled, we can guarantee that the fraction of misclassified inputs can be made polynomially small. Thus we can evaluate f^α in $\text{HeurBPP}^{\text{NP}^{\text{NP}}}$.

We notice that Theorem 2 remains valid assuming only $\text{BQP} \not\subseteq \text{PH}$ when the identification algorithm is required to output a consistent hypothesis for every input distribution. This then leads to the main result of our paper, which establishes the existence of a physically motivated learning problem—learning local observables—that already exhibits a quantum advantage at the identification stage.

Main result of the paper

Theorem 3 (Informal, see Theorem 15 in the full paper for a formal version). *There exists a natural learning problem involving expectation values of quantum functions for which there exist a quantum learning advantage for the identification task (i.e. the target concepts are efficiently identifiable on a quantum computer but not on a classical computer) unless BQP is in the polynomial hierarchy.*

Proof sketch. The proof can, again, be found in the full paper. The key idea of the proof is to consider as target functions expectation values of a families of local observables O on a particular family of quantum states ρ_U such that the corresponding concept class contains c -distinct concepts with $c \geq 1/3$. We achieve this by leveraging the Reed-Solomon code to construct functions that differ from each other on at least $1/3$ of all the input x . As this is still an instance of the same category of concept classes as studied in [MGD24], the quantum learning algorithm from this work will solve the identification task, and also the corresponding evaluation task. $\square$

Finally, we consider two physically relevant tasks—Hamiltonian learning and learning the order parameter—that fit naturally as identification problems, and outline the extra components needed for our techniques to apply, clarifying why classically efficient Hamiltonian learning does not contradict our results.

References

- [BGV⁺24] Alice Barthe, Michele Grossi, Sofia Vallecora, Jordi Tura, and Vedran Dunjko. Parameterized quantum circuits as universal generative models for continuous multivariate distributions, 2024.
- [GD23] Casper Gyurik and Vedran Dunjko. Exponential separations between classical and quantum learners. arXiv preprint arXiv:2306.16028, 2023.
- [KV94] Michael J Kearns and Umesh Vazirani. An introduction to computational learning theory. MIT press, 1994.
- [LAT21] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. A rigorous and robust quantum speed-up in supervised machine learning. Nature Physics, 17(9):1013–1017, 2021.
- [MGD24] Riccardo Molteni, Casper Gyurik, and Vedran Dunjko. Exponential quantum advantages in learning quantum observables from classical data. arXiv preprint arXiv:2405.02027, 2024.
- [RAG21] Jonathan Romero and Alán Aspuru-Guzik. Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions. Advanced Quantum Technologies, 4(1):2000003, 2021.
- [SG04] Rocco A Servedio and Steven J Gortler. Equivalences and separations between quantum and classical learnability. SIAM Journal on Computing, 33(5):1067–1092, 2004.
- [SKS⁺24] Kevin Shen, Andrii Kurkin, Adrián Pérez Salinas, Elvira Shishenina, Vedran Dunjko, and Hao Wang. Shadow-frugal expectation-value-sampling variational quantum generative model, 2024.