


Classical Verification of Quantum Learning

Matthias C. Caro ✉ 🏠 

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany
Institute for Quantum Information and Matter, Caltech, Pasadena, CA, USA

Marcel Hinsche ✉ 

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany

Marios Ioannou ✉

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany

Alexander Nietner ✉ 

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany

Ryan Sweke ✉ 

IBM Quantum, Almaden Research Center, San Jose, CA, USA

Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany

Abstract

Quantum data access and quantum processing can make certain classically intractable learning tasks feasible. However, quantum capabilities will only be available to a select few in the near future. Thus, reliable schemes that allow classical clients to delegate learning to untrusted quantum servers are required to facilitate widespread access to quantum learning advantages. Building on a recently introduced framework of interactive proof systems for classical machine learning, we develop a framework for classical verification of quantum learning. We exhibit learning problems that a classical learner cannot efficiently solve on their own, but that they can efficiently and reliably solve when interacting with an untrusted quantum prover. Concretely, we consider the problems of agnostic learning parities and Fourier-sparse functions with respect to distributions with uniform input marginal. We propose a new quantum data access model that we call “mixture-of-superpositions” quantum examples, based on which we give efficient quantum learning algorithms for these tasks. Moreover, we prove that agnostic quantum parity and Fourier-sparse learning can be efficiently verified by a classical verifier with only random example or statistical query access. Finally, we showcase two general scenarios in learning and verification in which quantum mixture-of-superpositions examples do not lead to sample complexity improvements over classical data. Our results demonstrate that the potential power of quantum data for learning tasks, while not unlimited, can be utilized by classical agents through interaction with untrusted quantum entities.

2012 ACM Subject Classification Theory of computation → Machine learning theory; Theory of computation → Quantum computation theory; Theory of computation → Interactive proof systems

Keywords and phrases computational learning theory, quantum learning theory, interactive proofs, quantum oracles, agnostic learning

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.24

Related Version *Full Version:* <https://arxiv.org/abs/2306.04843>

Funding The authors acknowledge support from the BMWK (PlanQK, EniQmA), the BMBF (Hybrid), and the Munich Quantum Valley (K-8). This work has also been funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany’s Excellence Strategy, The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689) as well as CRC 183 (B1).

Matthias C. Caro: This research was supported by a DAAD PRIME fellowship. The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center.

Acknowledgements We thank Jens Eisert for his valuable input to discussions on this project and for suggestions on improving the draft. We thank Srinivasan Arunachalam, Jack O’Connor, Yihui Quek, Jonathan Shafer, Thomas Vidick, and the ITCS reviewers for insightful comments and discussions.



© Matthias C. Caro, Marcel Hinsche, Marios Ioannou, Alexander Nietner, and Ryan Sweke; licensed under Creative Commons License CC-BY 4.0

15th Innovations in Theoretical Computer Science Conference (ITCS 2024).

Editor: Venkatesan Guruswami; Article No. 24; pp. 24:1–24:23

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

For many learning problems, the data to which we have access determine our ability to obtain a good hypothesis. Unfortunately, in practical settings there is often a cost associated with collecting high quality data, and this cost prohibits us from solving a learning problem of interest. In light of this, it would be desirable to delegate learning problems to *untrusted* servers with access to more or higher-quality data than ourselves. Ideally we would like such “data-rich” servers to efficiently solve the learning problem, and we would like to efficiently verify, using both the limited data available to us and interaction with the server, that the server has indeed successfully solved the learning problem. Recently, a formal framework – *interactive proofs for the verification of machine learning* – has been introduced to explore when, and to which extent, such delegation of learning tasks is possible [46].

In this work, we are interested in verifying learning with untrusted *quantum* servers, with access to quantum data. Indeed, there is a rich history of work on quantum learning theory [6], aimed at understanding quantum learning algorithms with access to different types of quantum data oracles. Notably, there exist classically intractable learning problems that can be efficiently solved by quantum learners. However, the most realistic future scenario is that quantum devices will be accessed remotely, and that only certain parties have access to hard-to-prepare and hard-to-store quantum data. Therefore, to realize the advantages of quantum learning, it becomes crucial that classical clients (verifiers) can delegate learning to untrusted quantum servers (provers) and efficiently verify the provided hypotheses, using only interaction with the server and the classical data that is readily available.

To explore the setting just described, it is necessary to fix a formal learning problem. In supervised learning, [46] showed that for standard Probably Approximately Correct (PAC) learning there exist trivial techniques for the verification of hypotheses and as such the verification problem is only non-trivial for *agnostic* PAC learning. Agnostic learning also captures an important feature of modern machine learning in practice: Often, one has few or no promises on the structure of the data, and one attempts to do the best possible by optimizing over a chosen model class. Given the necessity of working within the framework of agnostic learning, the question of whether or not it is possible for classical clients to delegate learning problems to untrusted quantum servers is only interesting if there exist agnostic learning problems in which the amount of resources required for classical learning exceeds that sufficient for quantum learners with access to quantum data. Unfortunately, however, little is known about the power of quantum learning algorithms for agnostic learning.

In light of the above, the main contributions in this work are two-fold: Firstly, we identify and motivate a novel quantum oracle for agnostic learning and, with respect to this oracle, provide the first efficient fully agnostic quantum learning algorithms for parities and Fourier-sparse functions. To the best of our knowledge, these are the first agnostic quantum learning algorithms for any model class for *distributional* agnostic learning. Secondly, we leverage these results to give a concrete example of a classically intractable agnostic learning problem that can be efficiently and reliably delegated to an untrusted quantum server. Namely, we provide an explicit interactive verification protocol which, despite the classical intractability of the learning task, allows the classical client to efficiently verify the hypothesis provided by a potentially dishonest quantum server. This demonstrates that classical clients can reap the benefits of quantum advantages in learning, in the realistic setting where learning needs to be delegated to untrusted servers. Our hope is that these results provide new tools and insights for agnostic quantum learning, as well as motivation for the development of further techniques for the secure delegation of learning problems to quantum servers.

1.1 Framework

Agnostic learning

When formalizing a learning task in which there may be a fundamental mismatch between the model used by the learner and the data-generating process, a so-called *agnostic* learning task [51, 63], there are two canonical choices:

- In *functional agnostic learning* w.r.t. uniformly random inputs, we assume that the data consists of labeled examples $(x_i, f(x_i))$, with the x_i drawn i.i.d. uniformly at random from $\mathcal{X}_n \equiv \{0, 1\}^n$ and with $f : \{0, 1\}^n \rightarrow \{0, 1\}$ an arbitrary unknown Boolean function. In this case, we denote the data-generating distribution as $\mathcal{D} = (\mathcal{U}_n, f)$.
- In *distributional agnostic learning* w.r.t. uniformly random inputs, we drop the assumption of a deterministic function that perfectly describes the data. That is, we assume labeled examples (x_i, y_i) drawn i.i.d. from some distribution \mathcal{D} over $\{0, 1\}^n \times \{0, 1\}$ with uniform marginal over $\{0, 1\}^n$. We denote this as $\mathcal{D} = (\mathcal{U}_n, \varphi)$ with conditional label expectation $\varphi : \{0, 1\}^n \rightarrow [0, 1]$, $\varphi(z) = \mathbb{E}_{(x,y) \sim \mathcal{D}}[y|x = z]$.

Whereas in functional agnostic learning there is a “correct” label for every input, this is no longer true in the distributional agnostic setting. In particular, in the latter case data could contain conflicting labels for the same input. Nevertheless, in both the functional and the distributional case, the goal is to learn an almost-optimal approximating function compared to a benchmark class \mathcal{B} : Given an accuracy parameter ε , a confidence parameter δ , and access to a training data set generated i.i.d. from \mathcal{D} , an α -agnostic learner has to output, with success probability $\geq 1 - \delta$, a hypothesis h such that

$$\mathbb{P}_{(x,y) \sim \mathcal{D}}[h(x) \neq y] \leq \alpha \cdot \inf_{b \in \mathcal{B}} \mathbb{P}_{(x,y) \sim \mathcal{D}}[b(x) \neq y] + \varepsilon. \quad (1)$$

Note that here we do not necessarily require that $h \in \mathcal{B}$. If we add this requirement, we speak of *proper* learning, otherwise the learner can be *improper*. Also, we recover the scenario of *realizable* PAC learning when assuming that $\inf_{b \in \mathcal{B}} \mathbb{P}_{(x,y) \sim \mathcal{D}}[b(x) \neq y] = 0$. For a formalization of this discussion, see [24, Definition 3].

Learning classical functions from quantum data

In quantum learning theory, a learner can have access to \mathcal{D} via a potentially more powerful resource than classical i.i.d. examples. Quantum training data for \mathcal{D} is canonically taken to consist of copies of the *quantum superposition example state* [17]

$$|\psi_{\mathcal{D}}\rangle = \sum_{(x,y) \in \{0,1\}^n \times \{0,1\}} \sqrt{\mathcal{D}(x,y)} |x, y\rangle. \quad (2)$$

Such quantum data is at least as powerful as its classical counterpart, since the former can simulate the latter via computational basis measurements. In fact, these quantum examples have proven to be useful for realizable learning and, to some degree, functional agnostic learning w.r.t. the uniform distribution. However, it is unknown how to use copies of $|\psi_{\mathcal{D}}\rangle$ to improve upon classical distributional agnostic learning.

Therefore, we propose a different quantum resource for distributional agnostic learning. Our starting point is that a distribution $\mathcal{D} = (\mathcal{U}_n, \varphi)$ induces a distribution $F_{\mathcal{D}}$ over the set of all functions mapping $\{0, 1\}^n$ to $\{0, 1\}$. Namely, $F_{\mathcal{D}}$ is defined by taking the probability that $f(x)$ equals 1 to be $\varphi(x)$ independently for each x , see Equation (3). We then consider quantum training data for $\mathcal{D} = (\mathcal{U}_n, \varphi)$ to consist of copies of the *mixture-of-superpositions example state* $\rho_{\mathcal{D}} = \mathbb{E}_{f \sim F_{\mathcal{D}}} [|\psi_{(\mathcal{U}_n, f)}\rangle \langle \psi_{(\mathcal{U}_n, f)}|]$, see Definition 5.

Interactive verification of agnostic learning

If quantum processing and quantum data are only available to a select few, enabling widespread use of quantum learning requires classical verification procedures. Extending the framework of [46], who formalized interactive verification of classical learning, we consider interactive classical verification of quantum learning. Here, an efficient classical verifier with classical data access, via random examples or statistical queries (SQs), interacts with an efficient quantum prover with mixture-of-superpositions quantum example or quantum SQ (QSQ) access. The goal of the verifier is twofold: On the one hand, when interacting with an honest quantum prover, the verifier should, with high probability, produce a hypothesis that satisfies the agnostic learning requirement. On the other hand, even when interacting with an arbitrarily powerful dishonest prover, the verifier should only accept the interaction and output a faulty hypothesis with small probability. If these two requirements are satisfied, the classical verifier can reliably profit from potential quantum advantages in learning. We work with the following small modification of [46, Definition 4] (see also [24, Definition 7]):

► **Definition 1** (Interactive verification of α -agnostic learning – Classical and/or quantum). *Let $\mathcal{B} \subseteq \{0, 1\}^{\mathcal{X}_n}$ be a benchmark class. Let \mathcal{D} be a family of probability distributions over $\mathcal{X}_n \times \{0, 1\}$. Let $\alpha \geq 1$. We say that \mathcal{B} is α -agnostic verifiable with respect to \mathcal{D} using classical or quantum oracles \mathcal{O}_V and \mathcal{O}_P if there exists a pair of classical or quantum algorithms (V, P) with access to the oracles $\mathcal{O}_V(\mathcal{D})$ and $\mathcal{O}_P(\mathcal{D})$ respectively that satisfy the following conditions for every input accuracy parameter $\varepsilon \in (0, 1)$ and for every confidence parameter $\delta \in (0, 1)$:*

- **Completeness:** *For any $\mathcal{D} \in \mathcal{D}$, if V interacts with the honest prover P , then, with probability $\geq 1 - \delta$, V accepts the interaction with P and outputs a hypothesis h that satisfies the agnostic learning criterion (Equation (1)).*
- **Soundness:** *For any $\mathcal{D} \in \mathcal{D}$ and for any (possibly unbounded) dishonest prover P' , if V interacts with P' , then, with probability $\leq \delta$, V accepts the interaction with P' and outputs a hypothesis h that does not satisfy the agnostic learning criterion (Equation (1)).*

Moreover:

- *If the above can be achieved with computationally efficient V and P , then we say that \mathcal{B} is efficiently α -agnostic verifiable with respect to \mathcal{D} using oracles \mathcal{O}_V and \mathcal{O}_P .*
- *If V either rejects or outputs a hypothesis $h \in \mathcal{B}$, then we say that \mathcal{B} is proper α -agnostic verifiable with respect to \mathcal{D} using oracles \mathcal{O}_V and \mathcal{O}_P .*

1.2 Overview of the Main Results

Our first contribution is proposing a new resource for agnostic quantum learning, namely mixture-of-superpositions states $\rho_{\mathcal{D}} = \mathbb{E}_{f \sim F_{\mathcal{D}}} [|\psi_{(\mathcal{U}_n, f)}\rangle \langle \psi_{(\mathcal{U}_n, f)}|]$ (see Definition 5). With this proposal, we return to the fundamental question of quantum learning theory: Do quantum versions of classical data access models enlarge the class of feasible learning problems? In particular, while quantum superposition examples have been widely adopted as the canonical “quantization” of classical random examples, it is of fundamental interest to understand what other consistent quantizations of classical data oracles exist, and how access to such oracles influences the complexity of different learning problems. To this end, we note that our mixture-of-superpositions examples are indeed *consistent*, in the sense that they reduce to classical random examples upon measurements in the computational basis, and to the established quantum superposition examples in the functional agnostic case. Additionally, our definition is well-motivated by a natural operational interpretation of classical random examples for arbitrary distributions, which has previously been used to provide reductions from classical distributional to functional agnostic learning (see the

discussion in [24, Appendix B]). More specifically, each time a mixture-of-superpositions oracle for the distribution \mathcal{D} is queried, it responds by first choosing a random function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ according to the distribution $F_{\mathcal{D}}$ induced by \mathcal{D} and then sending a copy of $|\psi_{(\mathcal{U}_n, f)}\rangle$. Finally, our mixture-of-superpositions examples can be viewed as enriching quantum learning by an analogue of randomized quantum oracles, which, as discussed in Section 1.3, have recently received attention in quantum complexity theory [50, 37, 71, 12]. Indeed, our motivation here is similar to these recent works – namely to understand the effect of different oracle models on the landscape of quantum sample/query complexity.

Quantum Fourier sampling [14] is a central subroutine in most existing quantum learning algorithms. However, while it is known how to do quantum Fourier sampling from quantum superposition examples for functional agnostic learning, it is unknown whether quantum superposition examples suffice to perform quantum Fourier sampling in the distributional agnostic setting. Our first main result shows that mixture-of-superpositions examples allow for an approximate version of quantum Fourier sampling in the distributional agnostic setting and are thus a valuable resource for distributional agnostic quantum learning algorithms:

► **Theorem 2** (Distributional agnostic approximate quantum Fourier sampling and learning – Informal). *Let $\mathcal{D} = (\mathcal{U}_n, \varphi)$ be an unknown probability distribution over $\{0, 1\}^n \times \{0, 1\}$, with (known) uniform marginal over $\{0, 1\}^n$ and with (unknown) conditional label expectation $\varphi : \{0, 1\}^n \rightarrow [0, 1]$.*

1. **Distributional agnostic quantum Fourier sampling:** *There is an efficient quantum algorithm that, given a single copy of $\rho_{\mathcal{D}}$, with success probability $1/2$ outputs a sample from a probability distribution over $\{0, 1\}^n$ that is inverse-exponentially close to the squares of the Fourier coefficients of $\phi = 1 - 2\varphi$.*
2. **Distributional agnostic proper quantum parity learning:** *There is an efficient quantum algorithm that properly 1-agnostically learns parities from an efficient number of copies of $\rho_{\mathcal{D}}$.*
3. **Distributional 2-agnostic improper quantum Fourier-sparse learning:** *There is an efficient quantum algorithm that improperly 2-agnostically learns Fourier-sparse functions from an efficient number of copies of $\rho_{\mathcal{D}}$.*

Theorem 2, proved in Section 2, constitutes the first general progress on distributional agnostic quantum learning w.r.t. uniform input marginal. It achieves this by generalizing quantum Fourier sampling from the functional to the distributional setting (see Theorem 7). In proving Theorem 2, we establish agnostic learning guarantees from Fourier spectrum approximation that, to the best of our knowledge, also improve upon the best known analogous classical result in terms of the achieved α . Moreover, in the full version [24], we prove that, based on a version of the Goldreich-Levin/Kushilevitz-Mansour algorithm [44, 65], agnostic parity and Fourier-sparse learning remain possible efficiently even in a weaker data access model of distributional agnostic quantum statistical queries, which we introduce as an extension of the classical statistical query model [61] and its functional quantum variant [8]. In addition, we provide a variety of results establishing the feasibility of Fourier sampling, finding heavy Fourier coefficients, and agnostic learning in the functional setting, when given access to different types of quantum oracles.

In our second main result, we identify an agnostic learning problem that a classical learner cannot solve on their own, but that becomes feasible for a classical verifier interacting with a quantum prover who has access to mixture-of-superpositions examples.

► **Theorem 3** (Verifying distributional agnostic quantum learning – Informal). *There is a class \mathfrak{D} of probability distributions over $\{0, 1\}^n \times \{0, 1\}$ with (known) uniform marginal over $\{0, 1\}^n$ such that:*

- (a) *Distributional 1-agnostic parity learning is classically hard from SQs or random examples, even if the unknown distribution is promised to lie in \mathfrak{D} .*
- (b) *When promised that the unknown distribution lies in \mathfrak{D} , there is an efficient interactive verification procedure that allows a classical verifier, with SQ or random example access, to verify a distributional 1-agnostic quantum parity learner, who has mixture-of-superpositions example or QSQ access.*
- (c) *When promised that the unknown distribution lies in \mathfrak{D} , there is an efficient interactive verification procedure that allows a classical verifier, with SQ or random example access, to verify a distributional 2-agnostic quantum Fourier-sparse learner, who has mixture-of-superpositions example or QSQ access.*

Theorem 3, which collects the statements of several theorems in [24, Section 6], shows that our new notion of quantum data not only enables distributional agnostic quantum learning but does so in a classically efficiently verifiable manner. Thereby, Theorem 3 establishes a separation between what a classical learner can achieve on their own and what they can achieve when interacting with an untrusted quantum prover. This separation is unconditional for SQ access and conditional on the hardness of Learning Parity with Noise (LPN) for random example access. Moreover, we show that the distribution class \mathfrak{D} used in Theorem 3 cannot be meaningfully enlarged without significant losses in the efficiency of the classical verifier. All of this is proved in Section 3.

Theorems 2 and 3 show that mixture-of-superpositions examples serve as a powerful resource that can change the learning landscape in a positive way, by allowing us to solve learning problems for which we have so far been lacking quantum learners. Crucially, however, our proposed model of quantum data access is not all-powerful: Just like their established superposition counterpart, mixture-of-superpositions examples do not allow for relevant sample complexity advantages over classical learners when considering distribution-independent agnostic learning.

► **Theorem 4** (Sample Complexity Lower Bound for Distribution-Independent Distributional Agnostic Quantum Learning – Informal Version). *The quantum sample complexity of distribution-independent distributional agnostic learning a function class $\mathcal{F} \subseteq \{0, 1\}^{\{0, 1\}^n}$ from mixture-of-superpositions examples does not improve upon the classical sample complexity, up to logarithmic factors.*

Classically, it is well-established that the sample complexity of distribution-independent distributional agnostic learning \mathcal{F} behaves as $\Theta\left(\frac{\text{VCdim}(\mathcal{F}) + \log(1/\delta)}{\epsilon^2}\right)$ [77, 15, 76]. Here, the VC-dimension $\text{VCdim}(\mathcal{F})$ is a combinatorial complexity measure for the function class \mathcal{F} [77]. While we prove a quantum sample complexity lower bound that matches the classical upper bound up to factors logarithmic in $\text{VCdim}(\mathcal{F})$, we in fact conjecture that quantum and classical sample complexities for distribution-independent learning coincide up to constant factors. In addition, we show that also the optimal sample complexity lower bound for verifying distribution-independent agnostic classical learning from [70] carries over to agnostic quantum learning with mixture-of-superpositions examples. Thus, whereas Theorem 2 and Theorem 3 exhibit the power of our newly proposed quantum resource, Theorem 4 demonstrates that, from an information-theoretic perspective, mixture-of-superpositions examples do not change the landscape of distribution-independent learning.

1.3 Related Work

Interactive verification of learning

Recently, Refs. [45, 46] introduced a complexity-theoretic framework for reasoning about the verification of delegated learning tasks via interactive proofs for PAC learning. Since then, the complexity of verifying PAC learning has been further characterized [70], and the framework has been extended to consider both statistical learning algorithms and arbitrary benchmark learning algorithms [70], as well as the setting of limited communication complexity between prover and verifier [73]. Our work initiates the study of the natural setting in which the untrusted prover is quantum, with access to a quantum data oracle.

Quantum and agnostic learning

Quantum learning theory is aimed at understanding the potential and limitations of quantum learning algorithms with access to different notions of quantum data [6], such as quantum examples [17], quantum membership queries [75, 69, 5] and quantum statistical queries [8]. A wide variety of results have shown both the limitations of quantum learning algorithms in the distribution-independent setting [9, 79, 7], as well as the advantages offered by quantum learning algorithms for distribution-dependent problems [17, 57, 10, 60, 19] and for learning from noisy quantum examples [32, 49, 19]. However, despite a rich history of work on *agnostic* learning in the classical setting [47, 58, 59, 40, 38, 39, 48, 52], the notion of quantum agnostic learning is relatively undeveloped, with only one recent work providing a *functional* agnostic quantum learning algorithm for decision trees [13]. Against this backdrop, our work makes a variety of contributions. Firstly, we broaden the scope of quantum learning theory through both the introduction of the mixture-of-superpositions quantum example for agnostic learning, as well as the initiation of delegated quantum learning. Additionally, by using the mixture-of-superpositions oracle, we give the first efficient *distributional* quantum agnostic learning algorithms. We achieve this by developing the toolbox for quantum Fourier sampling [14], which extends a long and active line of classical work on Fourier-based learning algorithms [44, 65, 66, 35].

Randomized quantum oracles

The mixture-of-superpositions oracle that we propose is similar in spirit to a variety of “non-standard” randomized quantum oracles that have recently been used to provide oracle separations between QMA and QCMA, and to expose the subtle effects of oracle design on the quantum query complexity of testing problems [50, 37, 71, 12]. Our work introduces such oracles to the setting of quantum learning theory.

Verification of quantum computation

This work can be seen as a learning-theoretic analogue to a long line of research aimed at providing protocols via which efficient classical verifiers (BPP machines) can verify the results of efficient quantum provers (BQP machines) [43, 67]. As discussed in Ref. [46], we note that the relation between verification of *learning* and verification of *computation* is non-trivial. Additionally, there is a large body of work on *privacy-preserving* delegation of both quantum computations [16, 41] and classical learning [18]. While similar in spirit to this work, we do not enforce any notion of privacy.

2 Mixture-of-Superpositions Examples and Distributional Agnostic Quantum Learning

2.1 Mixture-of-Superpositions Examples

Despite successes in using quantum superposition examples for functional agnostic learning, their power for distributional agnostic learning remains unclear. Here, we introduce a new form of quantum data for distributional agnostic quantum learning:

► **Definition 5** (Mixture-of-superpositions quantum examples for distributional agnostic learning). *Let \mathcal{D} be a probability distribution over $\mathcal{X}_n \times \{0, 1\}$. Let $F_{\mathcal{D}}$ be the probability distribution over $\{0, 1\}^{\mathcal{X}_n}$ defined by sampling $f(x)$ from the conditional label distribution independently for each $x \in \mathcal{X}_n$. That is, for any $\tilde{f} : \mathcal{X}_n \rightarrow \{0, 1\}$,*

$$\mathbb{P}_{f \sim F_{\mathcal{D}}}[f = \tilde{f}] = \prod_{z \in \mathcal{X}_n} \mathbb{P}_{(x,y) \sim \mathcal{D}}[\tilde{f}(z) = y | x = z] = \prod_{z \in \mathcal{X}_n} ((1 - \varphi(z))(1 - \tilde{f}(z)) + \varphi(z)\tilde{f}(z)). \quad (3)$$

A mixture-of-superpositions quantum example for \mathcal{D} is a copy of the $(n + 1)$ -qubit state

$$\rho_{\mathcal{D}} = \mathbb{E}_{f \sim F_{\mathcal{D}}} [|\psi_{(\mathcal{D}_{\mathcal{X}_n}, f)}\rangle \langle \psi_{(\mathcal{D}_{\mathcal{X}_n}, f)}|]. \quad (4)$$

Accordingly, a mixture-of-superpositions quantum example oracle for \mathcal{D} is an oracle that when queried outputs a copy of $\rho_{\mathcal{D}}$.

Randomized quantum oracles similar in spirit to Definition 5 have previously appeared in a complexity-theoretic context, see for example [50, 37, 71, 12]. While standard quantum oracles (as for example that of Equation (2)) can be viewed in terms of black box unitaries, randomized quantum oracles correspond to black box mixed unitary channels. Note that Definition 5 reproduces the definition of a noisy functional quantum example from [49] when applied to a distribution \mathcal{D}_{η} that is obtained by adding i.i.d. label noise of strength $\eta \geq 0$ to a distribution $(\mathcal{D}_{\mathcal{X}_n}, f)$ with a deterministic labeling function $f : \mathcal{X}_n \rightarrow \{0, 1\}$. In particular, Definition 5 reproduces the functional superposition example of [17] for distributions of the form $(\mathcal{D}_{\mathcal{X}_n}, f)$ with Boolean f . Importantly, however, Definition 5 covers more general distributions, for example distributions arising from adding correlated labeling noise to a deterministic labeling. Definition 5 also reproduces the standard notion of a classical distributional agnostic random example under computational basis measurements:

► **Lemma 6.** *Let \mathcal{D} be a probability distribution over $\mathcal{X}_n \times \{0, 1\}$. Performing computational basis measurements on all $n + 1$ qubits of a copy of $\rho_{\mathcal{D}}$ produces a sample from \mathcal{D} .*

Proof. By definition of $\rho_{\mathcal{D}}$, the probability of observing an output string $(x, b) \in \mathcal{X}_n \times \{0, 1\}$ when measuring all $n + 1$ qubits in the computational basis is given by

$$\langle x, b | \rho_{\mathcal{D}} | x, b \rangle = \mathbb{E}_{f \sim F_{\mathcal{D}}} [|\langle x, b | \psi_{(\mathcal{D}_{\mathcal{X}_n}, f)} \rangle|^2] = \mathbb{E}_{f \sim F_{\mathcal{D}}} [\mathcal{D}_{\mathcal{X}_n}(x) \delta_{b, f(x)}] \quad (5)$$

$$= \mathcal{D}_{\mathcal{X}_n}(x) \mathbb{P}_{f \sim F_{\mathcal{D}}}[f(x) = b] = \mathcal{D}(x, b), \quad (6)$$

as claimed. ◀

Thus, Definition 5 constitutes a generalization of established definitions, both classical and quantum. Moreover, the probability distribution $F_{\mathcal{D}}$ over functions is an object that naturally appears in classical learning-theoretic proofs of reductions between distributional and functional agnostic learning, compare [47, Appendix A] and [24, Appendix B].

2.2 Distributional Agnostic Approximate Quantum Fourier Sampling

It is not known how to use the conventional superposition quantum examples $|\psi_{\mathcal{D}}\rangle$ to speed up distributional agnostic learning. As we show below, the key advantage of our newly introduced mixture-of-superpositions quantum examples $\rho_{\mathcal{D}} = \mathbb{E}_{f \sim F_{\mathcal{D}}} [|\psi_{(\mathcal{U}_n, f)}\rangle \langle \psi_{(\mathcal{U}_n, f)}|]$ is that they enable approximate quantum Fourier sampling in the distributional setting under uniform input marginal. To achieve this approximate Fourier sampling, we use the same simple, standard quantum algorithm that is known to work in the realizable setting: applying a layer of single-qubit Hadamard gates to a single copy of $\rho_{\mathcal{D}}$ followed by a measurement in the computational basis and post-selecting on the outcome 1 in the last qubit.

► **Theorem 7** (Formal statement of Theorem 2, Point 1). *Let \mathcal{D} be a probability distribution over $\mathcal{X}_n \times \{0, 1\}$ with $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$. Consider the following quantum algorithm: Given a copy of $\rho_{\mathcal{D}}$, first apply (the unitary channel for) the unitary $H^{\otimes(n+1)}$, then measure all $n+1$ qubits in the computational basis. The measurement outcomes of this procedure satisfy the following:*

- (i) *The computational basis measurement on the last qubit gives outcome 0 with probability $1/2$ and outcome 1 with probability $1/2$.*
- (ii) *Conditioned on having observed outcome 1 for the last qubit, the computational basis measurement on the first n qubits outputs a string $s \in \{0, 1\}^n$ with probability*

$$\frac{1}{2^n} (1 - \mathbb{E}_{x \sim \mathcal{U}_n} [(\phi(x))^2]) + (\hat{\phi}(s))^2. \quad (7)$$

The squares of the Fourier coefficients of ϕ in general do not form a probability distribution, because in general $\mathbb{E}_{x \sim \mathcal{U}_n} [(\phi(x))^2] < 1$. Thus, it does not make sense to speak of exact sampling from the “distribution formed by squares of Fourier coefficients” in this distributional agnostic case. However, by Parseval, we know that $\{\frac{1}{2^n} (1 - \mathbb{E}_{x \sim \mathcal{U}_n} [(\phi(x))^2]) + (\hat{\phi}(s))^2\}_{s \in \{0, 1\}^n}$ does form a probability distribution. It is exactly this probability distribution that Theorem 7 allows us to sample from (with success probability $1/2$).

Proof. As $\rho_{\mathcal{D}}$ is a probabilistic mixture, we have, for any $s \in \{0, 1\}^n$ and $b \in \{0, 1\}$,

$$\langle s, b | H^{\otimes(n+1)} \rho_{\mathcal{D}} H^{\otimes(n+1)} | s, b \rangle = \mathbb{E}_{f \sim F_{\mathcal{D}}} \left[\left| \langle s, b | H^{\otimes(n+1)} |\psi_{(\mathcal{U}_n, f)}\rangle \right|^2 \right]. \quad (8)$$

Thus, standard (functional agnostic) quantum Fourier sampling [14] (see also [24, Lemma 2]) immediately gives (i) and tells us that, conditioned on having observed outcome 1 for the computational basis measurement on the last qubit, the computational basis measurement on the first n qubits produces string $s \in \{0, 1\}^n$ with probability $\mathbb{E}_{f \sim F_{\mathcal{D}}} [(\hat{g}_f(s))^2]$, where $g_f(s) = (-1)^f$. Using the definition of $F_{\mathcal{D}}$ via an independent sampling of labels, we can rewrite this quantity as

$$\mathbb{E}_{f \sim F_{\mathcal{D}}} [(\hat{g}_f(s))^2] \quad (9)$$

$$= \frac{1}{4^n} \sum_{x, y \in \{0, 1\}^n} \chi_s(x) \chi_s(y) \mathbb{E}_{f \sim F_{\mathcal{D}}} [(-1)^{f(x)} (-1)^{f(y)}] \quad (10)$$

$$= \frac{1}{4^n} \sum_{x, y \in \{0, 1\}^n} \chi_s(x) \chi_s(y) \cdot \begin{cases} \mathbb{E}_{f \sim F_{\mathcal{D}}} [(-1)^{f(x)}] \cdot \mathbb{E}_{f \sim F_{\mathcal{D}}} [(-1)^{f(y)}] & \text{if } x \neq y \\ 1 & \text{if } x = y \end{cases}. \quad (11)$$

Next, recall that $\mathbb{E}_{f \sim F_{\mathcal{D}}} [(-1)^{f(x)}] = 1 - 2\varphi(x) = \phi(x)$ holds by definition of $F_{\mathcal{D}}$. Using that $\chi_s^2 = 1$ holds for any $s \in \mathcal{X}_n$, this allows us to further rewrite

$$\mathbb{E}_{f \sim F_{\mathcal{D}}} [(\hat{g}_f(s))^2] = \frac{1}{4^n} \sum_{x \in \{0, 1\}^n} (\chi_s(x))^2 + \frac{1}{4^n} \sum_{\substack{x, y \in \{0, 1\}^n \\ x \neq y}} \chi_s(x) \phi(x) \chi_s(y) \phi(y) \quad (12)$$

$$= \frac{1}{2^n} + \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} \chi_s(x) \phi(x) \right)^2 - \frac{1}{4^n} \sum_{x \in \{0,1\}^n} (\chi_s(x))^2 (\phi(x))^2 \quad (13)$$

$$= \frac{1}{2^n} (1 - \mathbb{E}_{x \sim \mathcal{U}_n} [(\phi(x))^2]) + (\hat{\phi}(s))^2. \quad (14)$$

This finishes the proof. \blacktriangleleft

To see that Theorem 7 indeed implies Point 1 of Theorem 2, note that ϕ is $[-1, 1]$ -valued, so $0 \leq 1 - \mathbb{E}_{x \sim \mathcal{U}_n} [(\phi(x))^2] \leq 1$. Therefore, we can, with success probability $1/2$, produce a sample from a distribution that is $(1/2^n)$ -close in ∞ -norm to the sub-normalized distribution formed by the squares of the Fourier coefficients of ϕ as follows: First, perform $n + 1$ single-qubit Hadamard gates on $\rho_{\mathcal{D}}$. Second, measure the last qubit in the computational basis. If the outcome is 0, the sampling attempt fails. If the outcome is 1, then measure the first n qubits in the computational basis and output the observed string of bits.

Now equipped with a distributional agnostic analogue of quantum Fourier sampling, we can approximate the Fourier spectrum of the conditional label expectation relying on the Dvoretzky-Kiefer-Wolfowitz (DKW) Theorem [34, 68, 64] (compare also [60, Lemma 4]):

► **Corollary 8.** *Let \mathcal{D} be a probability distribution over $\mathcal{X}_n \times \{0, 1\}$ with $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$. Let $\delta, \varepsilon \in (0, 1)$. Assume that $\varepsilon > 2^{-\binom{n}{2}-2}$. Then, there exists a quantum algorithm that, given $\mathcal{O}\left(\frac{\log(1/\delta\varepsilon^2)}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, uses $\mathcal{O}\left(n\frac{\log(1/\delta\varepsilon^2)}{\varepsilon^4}\right)$ single-qubit gates, classical computation time $\tilde{\mathcal{O}}\left(n\frac{\log(1/\delta\varepsilon^2)}{\varepsilon^4}\right)$, and classical memory of size $\tilde{\mathcal{O}}\left(n\frac{\log(1/\delta\varepsilon^2)}{\varepsilon^4}\right)$, and outputs, with success probability $\geq 1 - \delta$, a succinctly represented $\tilde{\phi} : \mathcal{X}_n \rightarrow [-1, 1]$ such that $\|\tilde{\phi} - \hat{\phi}\|_{\infty} \leq \varepsilon$ and $\|\tilde{\phi}\|_0 \leq \frac{16\mathbb{E}_{x \sim \mathcal{U}_n} [(\phi(x))^2]}{\varepsilon^2} \leq \frac{16}{\varepsilon^2}$.*

Note that Corollary 8 imposes an additional assumption compared to the functional case, namely a lower bound on the desired accuracy ε . However, this lower bound is inverse-exponential in n and thus satisfied (for large enough n) for the inverse-polynomial accuracies that are usually of interest.

Proof. Our proof is similar to that of [60, Theorem 5]. Theorem 7 gives a procedure that, using a single copy of $\rho_{\mathcal{D}}$ and $n + 1$ single-qubit quantum gates, produces a sample from the probability distribution $q : \{0, 1\}^{n+1} \rightarrow [0, 1]$ defined via

$$q(s, 1) = \frac{1}{2} \left(\frac{1}{2^n} (1 - \mathbb{E}_{x \sim \mathcal{U}_n} [(\phi(x))^2]) + (\hat{\phi}(s))^2 \right), \quad q(0^n, 0) = \frac{1}{2}. \quad (15)$$

Hence, according to [24, Lemma 3] applied for the probability distribution q , confidence $\delta > 0$ and accuracy $\tau = \varepsilon^2/\delta$, we see that $m = \mathcal{O}\left(\frac{\log(1/\delta)}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$ are sufficient to obtain, with success probability $\geq 1 - \frac{\delta}{2}$, a succinctly represented estimate \tilde{q}_m such that $\|\tilde{q}_m\|_0 \leq \mathcal{O}\left(\frac{\log(1/\delta)}{\varepsilon^4}\right)$ and $\|q - \tilde{q}_m\|_{\infty} \leq \frac{\varepsilon^2}{8}$. Moreover, the estimate \tilde{q}_m can be obtained using $\mathcal{O}(nm) = \mathcal{O}\left(n\frac{\log(1/\delta)}{\varepsilon^4}\right)$ single-qubit Hadamard gates, classical computation time $\tilde{\mathcal{O}}\left(n\frac{\log(1/\delta)}{\varepsilon^4}\right)$, and classical memory of size $\tilde{\mathcal{O}}\left(n\frac{\log(1/\delta)}{\varepsilon^4}\right)$.

Starting from the estimate \tilde{q}_m for q , we output a list L of strings $s \in \{0, 1\}^n$ such that $\tilde{q}_m(s, 1) \geq \varepsilon^2/4$. As we have a succinct representation of \tilde{q}_m with at most $\mathcal{O}\left(\frac{\log(1/\delta)}{\varepsilon^4}\right)$ non-zero entries, the list L can be compiled by brute-force search in classical computation time $\mathcal{O}\left(n\frac{\log(1/\delta)}{\varepsilon^4}\right)$. By our approximation guarantee, with success probability $\geq 1 - \frac{\delta}{2}$, we have the following:

- If $|\hat{\phi}(s)| \geq \varepsilon$, then $\tilde{q}(s, 1) \geq \frac{\varepsilon^2}{2} - \frac{\varepsilon^2}{8} \geq \frac{\varepsilon^2}{4}$. So, if s is an ε -heavy Fourier coefficient of $\hat{\phi}$, then $s \in L$.
- If $s \in L$, that is, if $\tilde{q}_m(s, 1) \geq \frac{\varepsilon^2}{4}$, then $(\hat{\phi}(s))^2 \geq 2 \left(\frac{\varepsilon^2}{8} - \frac{1}{2^n} (1 - \mathbb{E}_{x \sim \mathcal{U}_n}[(\phi(x))^2]) \right) \geq \frac{\varepsilon^2}{16}$, thus also $|\hat{\phi}(s)| \geq \frac{\varepsilon}{4}$ and s is an $(\frac{\varepsilon}{4})$ -heavy Fourier coefficient of $\hat{\phi}$. In particular, combining this with Parseval's equality and the fact that $\mathbb{E}_{x \sim \mathcal{U}_n}[(\phi(x))^2] \leq 1$, we see that $|L| \leq \frac{16 \mathbb{E}_{x \sim \mathcal{U}_n}[(\phi(x))^2]}{\varepsilon^2} \leq \frac{16}{\varepsilon^2}$.

Now, for each of the at most $\frac{16}{\varepsilon^2}$ strings in L , we estimate the corresponding Fourier coefficient. For any single such string s , by Hoeffding's inequality, we know that $\mathcal{O}\left(\frac{\log(1/\delta)}{\varepsilon^2}\right)$ classical samples from \mathcal{D} suffice to produce an empirical estimate $\tilde{\phi}(s)$ that matches $\hat{\phi}(s)$ up to accuracy ε , with success probability $\geq 1 - \frac{\delta}{2}$. By a union bound over L , this implies that $m_2 = \mathcal{O}\left(|L| \frac{\log(1/\delta)}{\varepsilon^2}\right) = \mathcal{O}\left(\frac{\log(1/\delta \varepsilon^2)}{\varepsilon^4}\right)$ classical samples from \mathcal{D} suffice to estimate all $\hat{\phi}(s)$ with $s \in L$ simultaneously up to accuracy ε , with success probability $\geq 1 - \frac{\delta}{2}$. As $\hat{\phi}(s) \in [-1, 1]$ for all $s \in \{0, 1\}^n$, these estimates can only improve if we project them to $[-1, 1]$. Moreover, building these empirical estimates can be done using classical computation time $\tilde{\mathcal{O}}\left(n \frac{\log(1/\delta \varepsilon^2)}{\varepsilon^4}\right)$ and classical memory of size $\tilde{\mathcal{O}}\left(n \frac{\log(1/\delta \varepsilon^2)}{\varepsilon^4}\right)$. It remains to observe that a single copy of $\rho_{\mathcal{D}}$ can be measured in the computational basis to obtain a sample from \mathcal{D} (recall Lemma 6), and that, by one more union bound, the produced estimate $\tilde{\phi}$ has the desired properties with probability $\geq 1 - \delta$. ◀

With this subroutine for obtaining a succinctly represented approximation to the Fourier spectrum of interest, we can now obtain distributional agnostic quantum learning algorithms.

2.3 Distributional Agnostic Quantum Learning Parities and Fourier-Sparse Functions

First, we show how to apply Corollary 8 as a subroutine for distributional agnostic quantum parity learning.

► **Corollary 9** (Formal statement of Theorem 2, Point 2). *Let \mathcal{D} be a probability distribution over $\mathcal{X}_n \times \{0, 1\}$ with $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$. Let $\delta, \varepsilon \in (0, 1)$. Assume that $\varepsilon > 2^{-\binom{n}{2}-2}$. There is a quantum algorithm that, given $\mathcal{O}\left(\frac{\log(1/\delta \varepsilon^2)}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, uses $\mathcal{O}\left(n \frac{\log(1/\delta \varepsilon^2)}{\varepsilon^4}\right)$ single-qubit gates, classical computation time $\tilde{\mathcal{O}}\left(n \frac{\log(1/\delta \varepsilon^2)}{\varepsilon^4}\right)$, and classical memory of size $\tilde{\mathcal{O}}\left(n \frac{\log(1/\delta \varepsilon^2)}{\varepsilon^4}\right)$, and outputs, with success probability $\geq 1 - \delta$, a bit string $s \in \{0, 1\}^n$ such that*

$$\mathbb{P}_{(x,b) \sim \mathcal{D}}[b \neq s \cdot x] \leq \min_{t \in \{0,1\}^n} \mathbb{P}_{(x,b) \sim \mathcal{D}}[b \neq t \cdot x] + \varepsilon. \quad (16)$$

Thus, this quantum algorithm is a distributional agnostic proper quantum parity learner up to inverse-exponentially small accuracies, assuming a uniform marginal over inputs.

Proof. As parity learning corresponds to learning a largest Fourier coefficient (compare [24, Lemma 11]), it suffices to show that there is a quantum algorithm with the claimed complexity bounds that, with success probability $\geq 1 - \delta$, outputs a (2ε) -approximately-largest Fourier coefficient of ϕ . To achieve this, first run the procedure from Corollary 8 to obtain, with probability $\geq 1 - \delta$, a succinctly represented $\tilde{\phi}$ such that $\|\tilde{\phi} - \hat{\phi}\|_{\infty} \leq \varepsilon$ and $\|\tilde{\phi}\|_0 \leq \frac{16}{\varepsilon^2}$. Now, let $s \in \operatorname{argmax}_{t \in \{0,1\}^n} \tilde{\phi}(t)$. Note that such an s can be found in time $\mathcal{O}\left(\frac{n}{\varepsilon^2}\right)$ since $\|\tilde{\phi}\|_0 \leq \frac{16}{\varepsilon^2}$. This s now satisfies

$$\max_{t \in \{0,1\}^n} \hat{\phi}(t) - \hat{\phi}(s) = \max_{t \in \{0,1\}^n} \hat{\phi}(t) - \tilde{\phi}(t) + \tilde{\phi}(t) - \tilde{\phi}(s) + \tilde{\phi}(s) - \hat{\phi}(s) \quad (17)$$

$$\leq \|\tilde{\phi} - \hat{\phi}\|_\infty + 0 + \|\tilde{\phi} - \hat{\phi}\|_\infty \leq 2\varepsilon, \quad (18)$$

as needed. The bounds on copy complexity, quantum gate complexity, classical runtime, and classical memory are all inherited from Corollary 8. \blacktriangleleft

In particular, Corollary 9 gives an efficient agnostic quantum parity learner for inverse-polynomial accuracy parameter ε and for inverse-exponential confidence parameter δ . In contrast, by reduction to the widely believed hardness of LPN, we do not expect an efficient classical procedure for the corresponding classical agnostic learning problem to exist.

In a similar vein, Corollary 8 can serve as a subroutine for distributional agnostic quantum learning of Fourier-sparse functions:

► Corollary 10 (Formal statement of Theorem 2, Point 3). *Let \mathcal{D} be a probability distribution over $\mathcal{X}_n \times \{0, 1\}$ with $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$. Let $\delta, \varepsilon \in (0, 1)$. Assume that $\varepsilon > 2^{-(\frac{n}{2}-2)}$. Then, there is a quantum algorithm that, given $\mathcal{O}\left(\frac{k^4 \log(k^2/\delta\varepsilon^2)}{\varepsilon^4}\right)$ copies of $\rho_{\mathcal{D}}$, uses $\mathcal{O}\left(n \frac{k^4 \log(1/\delta\varepsilon^2)}{\varepsilon^4}\right)$ single-qubit gates, classical computation time $\tilde{\mathcal{O}}\left(n \frac{k^4 \log(k^2/\delta\varepsilon^2)}{\varepsilon^4}\right)$, and classical memory of size $\tilde{\mathcal{O}}\left(n \frac{k^4 \log(k^2/\delta\varepsilon^2)}{\varepsilon^4}\right)$, and outputs, with success probability $\geq 1 - \delta$, a randomized hypothesis $h : \mathcal{X}_n \rightarrow \{0, 1\}$ such that*

$$\mathbb{P}_{(x,b) \sim \mathcal{D}} [b \neq h(x)] \leq 2 \min_{\substack{\tilde{f} : \mathcal{X}_n \rightarrow \{0,1\} \\ \text{Fourier-}k\text{-sparse}}} \mathbb{P}_{(x,b) \sim \mathcal{D}} [b \neq \tilde{f}(x)] + \varepsilon. \quad (19)$$

In particular, this quantum algorithm is a distributional 2-agnostic improper quantum Fourier-sparse learner up to inverse-exponentially small accuracies, assuming a uniform marginal over inputs.

Proof. By [24, Lemma 14] it suffices to show that there is a quantum algorithm with the claimed complexity bounds that, with success probability $\geq 1 - \delta$, outputs $(\varepsilon/2k)$ -accurate estimates of k $(\varepsilon/2k)$ -approximately-heaviest Fourier coefficients of ϕ . To achieve this, let $\tilde{\varepsilon} = \varepsilon/4k$ and run the procedure from Corollary 8 to obtain, with probability $\geq 1 - \delta$, a succinctly represented $\tilde{\phi} : \mathcal{X}_n \rightarrow [-1, 1]$ such that $\|\tilde{\phi} - \hat{\phi}\|_\infty \leq \tilde{\varepsilon}$ and $\|\tilde{\phi}\|_0 \leq \frac{16}{\tilde{\varepsilon}^2}$. Let $s_1 \in \operatorname{argmax}_{t \in \{0,1\}^n} |\tilde{\phi}(t)|$ and, for $2 \leq \ell \leq k$, let $s_\ell \in \operatorname{argmax}_{t \in \{0,1\}^n \setminus \{s_1, \dots, s_{\ell-1}\}} |\tilde{\phi}(t)|$. Note that such s_1, \dots, s_k can be found in time $\mathcal{O}\left(\frac{nk^4}{\varepsilon^2}\right)$ since $\|\tilde{\phi}\|_0 \leq \frac{16}{\tilde{\varepsilon}^2} \leq \mathcal{O}\left(\frac{k^4}{\varepsilon^2}\right)$. Let $t_1 \in \operatorname{argmax}_{t \in \{0,1\}^n} |\hat{\phi}(t)|$, and for $2 \leq \ell \leq k$, let $t_\ell \in \operatorname{argmax}_{t \in \{0,1\}^n \setminus \{t_1, \dots, t_{\ell-1}\}} |\hat{\phi}(t)|$. By the technical lemma ([24, Lemma 17]), $\|\tilde{\phi} - \hat{\phi}\|_\infty \leq \tilde{\varepsilon}$ implies that, for every $1 \leq \ell \leq k$, $\left| |\hat{\phi}(t_\ell)| - |\hat{\phi}(s_\ell)| \right| \leq 2\tilde{\varepsilon} \leq \varepsilon/2k$, so we can apply [24, Lemma 14]. The bounds on copy complexity, quantum gate complexity, classical runtime, and classical memory are all inherited from Corollary 8. \blacktriangleleft

As 1-agnostic Fourier-sparse learning is at least as hard as 1-agnostic parity learning, which in turn is at least as hard as LPN, this task is widely believed to be classically intractable from random examples. To the best of our knowledge, currently there are also no classical algorithms for 2-agnostic Fourier-sparse learning from examples. Thus, while Corollary 10 does not achieve 1-agnostic quantum Fourier-sparse learning, it serves as an

indication for the power of mixture-of-superpositions examples in learning Fourier-sparse functions w.r.t. uniformly random inputs. In [24, Section 5.3], we further introduce a notion of mixture-of-superpositions QSQs and demonstrate that also they are a powerful resource for distributional agnostic learning.

3 Classical Verification of Distributional Agnostic Quantum Learning

Section 2 has demonstrated the power of quantum data for agnostic parity learning and Fourier-sparse learning. Here, we show that classical verifiers interacting with quantum provers can make use of this power to solve similar learning problems. The results in this subsection, which serve to fully establish Theorem 3 when focusing on the regime $\varepsilon, \vartheta \geq \Omega(1/\text{poly}(n))$ and $\delta \geq \Omega(1/\exp(n))$, rely on two assumptions.

► **Definition 11** (Distributions with no small non-zero Fourier coefficients). *Let $\vartheta \in (0, 1)$. We denote the class of probability distributions $\mathcal{D} = (\mathcal{U}_n, \varphi)$ over $\mathcal{X}_n \times \{0, 1\}$ that have a uniform marginal over \mathcal{X}_n and whose $\{-1, 1\}$ -label expectation ϕ has no non-zero Fourier coefficients of magnitude $< \vartheta$ by*

$$\mathfrak{D}_{\mathcal{U}_n; \geq \vartheta} := \left\{ (\mathcal{U}_n, \varphi) \mid \hat{\phi} \neq 0 \Rightarrow |\hat{\phi}| \geq \vartheta \right\}. \quad (20)$$

Importantly, when considering learning problems under the promise $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; \geq \varepsilon}$, this includes scenarios in which the unknown distribution is a (noisy) parity or Fourier-sparse function. For the distributional agnostic setting considered in this subsection, we rely on the following additional assumption.

► **Definition 12** (Distributions with L_2 -bounded bias). *Let $0 \leq a \leq b \leq 1$. We denote the class of probability distributions $\mathcal{D} = (\mathcal{U}_n, \varphi)$ over $\mathcal{X}_n \times \{0, 1\}$ that have a uniform marginal over \mathcal{X}_n and whose $\{-1, 1\}$ -label expectation ϕ has squared L_2 norm in $[a^2, b^2]$ by*

$$\mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]} := \left\{ (\mathcal{U}_n, \varphi) \mid \mathbb{E}_{x \sim \mathcal{U}_n}[(\phi(x))^2] \in [a^2, b^2] \right\}. \quad (21)$$

Even with an added promise of this form, we still generalize beyond the noiseless and noisy functional agnostic cases. Namely, the noiseless functional case comes with the strong promise of $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]}$ for $a = b = 1$, and for the noisy functional case with noise rate η we can take $a = b = (1 - 2\eta)$. In particular, distributional agnostic parity learning under the promise $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; \geq (1-2\eta)} \cap \mathfrak{D}_{\mathcal{U}_n; [(1-2\eta)^2, (1-2\eta)^2]}$ is at least as hard as LPN.

We now state the distributional agnostic version of classical verification for quantum parity learning (see [24, Theorem 15] for a Fourier-sparse learning version).

► **Theorem 13.** *Let $\vartheta \in (2^{-(\frac{n}{2}-3)}, 1)$. Let $0 \leq a \leq b \leq 1$. Let $\delta \in (0, 1)$ and $\varepsilon \geq 2\sqrt{b^2 - a^2}$. The class of n -bit parities is efficiently proper 1-agnostic verifiable w.r.t. $\mathfrak{D}_{\mathcal{U}_n; \geq \vartheta} \cap \mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]}$ by a classical verifier V with access to classical random examples interacting with a quantum prover P with access to mixture-of-superpositions quantum examples. There is a verifier-prover pair (V, P) such that P uses $\mathcal{O}\left(\frac{\log(1/\delta\vartheta^2)}{\vartheta^4}\right)$ copies of $\rho_{\mathcal{D}}$, $\mathcal{O}\left(n \frac{\log(1/\delta\vartheta^2)}{\vartheta^4}\right)$ single-qubit gates, a classical memory of size $\tilde{\mathcal{O}}\left(n \frac{\log(1/\delta\vartheta^2)}{\vartheta^4}\right)$, and classical running time $\tilde{\mathcal{O}}\left(n \frac{\log(1/\delta\vartheta^2)}{\vartheta^4}\right)$, and such that V uses $\mathcal{O}\left(\frac{b^4 \log(1/\delta\vartheta^2)}{\varepsilon^4 \vartheta^4}\right)$ classical random examples, $\tilde{\mathcal{O}}\left(n \frac{b^4 \log(1/\delta\vartheta^2)}{\varepsilon^4 \vartheta^4}\right)$ classical running time, and a classical memory of size $\tilde{\mathcal{O}}\left(n \frac{b^4 \log(1/\delta\vartheta^2)}{\varepsilon^4 \vartheta^4}\right)$. Moreover, this can be achieved by a pair (V, P) that uses only a single round of communication consisting of at most $\mathcal{O}\left(\frac{n}{\vartheta^2}\right)$ classical bits.*

Proof sketch (see [24, Theorem 12] for a detailed proof). In our protocol, the prover P uses their quantum data access to produce a list $L = \{s_1, \dots, s_{|L|}\}$ corresponding to all non-negligible Fourier coefficients $\hat{\phi}(s)$ and sends it to the verifier V . Given such a list, the V can independently from P estimate the total squared Fourier weight of the list, $\sum_{\ell=1}^{|L|} (\hat{\phi}(s_\ell))^2$, using their classical data access only. Based on the promise $D \in \mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]}$, V can find out if P cheated by checking whether its estimate of the total Fourier weight of the list deviates too much from the a priori known total Fourier weight of the distribution. Moreover, the promise $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; \geq \vartheta}$ ensures that P can efficiently find all relevant Fourier coefficients and implies a bound on the length of the list L , which is important for the runtime of V .

Let us describe the protocol in more detail. Let $\delta, \varepsilon \in (0, 1)$. Let $0 \leq a \leq b \leq 1$. Let $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; \geq \vartheta} \cap \mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]}$. Assume that $\varepsilon \geq 2\sqrt{b^2 - a^2}$, with $\vartheta \in (2^{-(\frac{n}{2}-3)}, 1)$. The actions of the classical verifier V and the honest quantum prover P are as follows:

1. V asks P to provide a list $L = \{s_1, \dots, s_{|L|}\} \subset \{0, 1\}^n$ of length $|L| \leq 64b^2/\vartheta^2$ consisting of pairwise distinct n -bit strings whose associated Fourier coefficients are non-zero.
2. P follows the procedure in Corollary 8 to produce, with success probability $\geq 1 - \frac{\delta}{2}$, a succinctly represented $\tilde{\phi} : \mathcal{X}_n \rightarrow [-1, 1]$ such that $\|\tilde{\phi} - \hat{\phi}\|_\infty \leq \vartheta/2$ and $\|\tilde{\phi}\|_0 \leq \frac{64b^2}{\vartheta^2}$. If P obtains an output that violates the $\|\cdot\|_0$ -bound, then P declares failure and the interaction aborts. Otherwise, P then sends the list $L = \{s \in \{0, 1\}^n \mid |\tilde{\phi}(s)| \geq \vartheta/2\}$ to V .
3. If V receives a list L of length $|L| > 64b^2/\vartheta^2$, V rejects the interaction. Otherwise, V uses $\mathcal{O}\left(\frac{|L|^2 \log(|L|/\delta)}{\varepsilon^4}\right)$ classical random examples from \mathcal{D} to obtain simultaneously $(\varepsilon^2/16|L|)$ -accurate estimates $\hat{\xi}(s)$ of $\hat{\phi}(s)$ for all $s \in L$, with success probability $\geq 1 - \frac{\delta}{2}$, via Chernoff-Hoeffding combined with a union bound over L . (For $t \notin L$, the verifier's estimate $\hat{\gamma}(t)$ for $\hat{g}(t)$ is just 0.)
4. If $\sum_{\ell=1}^{|L|} (\hat{\xi}(s_\ell))^2 \geq a^2 - \frac{\varepsilon^2}{8}$, then V determines $s_{\text{out}} \in \operatorname{argmax}_{1 \leq \ell \leq |L|} \hat{\xi}(s)$ and outputs the hypothesis $h : \mathcal{X}_n \rightarrow \{0, 1\}$, $h(x) = s_{\text{out}} \cdot x$. If $\sum_{\ell=1}^{|L|} (\hat{\xi}(s_\ell))^2 < a^2 - \frac{\varepsilon^2}{8}$, then V outputs reject.

We now show that the pair (V, P) has the desired completeness and soundness properties. As a first step towards this goal, we show that V accepts an interaction with P with high probability. To this end, observe that, conditioned on P succeeding in Step 2, V never rejects in Step 3. If we then further condition on V succeeding in Step 3, we can use 2-Lipschitzness of $[-1, 1] \ni \xi \rightarrow \xi^2$, the promise $\mathcal{D} \in \mathfrak{D}_{\mathcal{U}_n; \geq \vartheta} \cap \mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]}$, and Parseval to show

$$\sum_{\ell=1}^{|L|} (\hat{\xi}(s_\ell))^2 = a^2 - \frac{\varepsilon^2}{8}. \quad (22)$$

Thus, if both Step 2 and Step 3 succeed, which by a union bound happens with probability $\geq 1 - \delta$, then V accepts in Step 4.

Moreover, whenever Step 3 is successful and V does not reject in Step 4, then the output string $s_{\text{out}} \in \operatorname{argmax}_{1 \leq \ell \leq |L|} \gamma(s)$ of V is as desired. To see this, we again use 2-Lipschitzness of $[-1, 1] \ni \xi \rightarrow \xi^2$ to show: If V does not reject in Step 4 and if Step 3 was successful, then this implies that for any $s \notin L$,

$$\left(\hat{\phi}(s)\right)^2 = (b^2 - a^2) + \frac{\varepsilon^2}{4}. \quad (23)$$

This tells us that $|\hat{\phi}(s)| \leq \sqrt{(b^2 - a^2) + \frac{\varepsilon^2}{4}} \leq \sqrt{b^2 - a^2} + \varepsilon/2 \leq \varepsilon$ holds for every $s \notin L$, which – writing the misclassification probability of a parity as $\mathbb{P}_{(x,y) \sim \mathcal{D}}[y \neq \chi_t(x)] = \frac{1}{2}(1 - \hat{\phi}(t))$ – now allows us to show that the output $s_{\text{out}} \in \operatorname{argmax}_{1 \leq \ell \leq |L|} \gamma(s)$ of V has the desired property

$\mathbb{P}_{(x,y) \sim \mathcal{D}}[y \neq \chi_{s_{\text{out}}}(x)] \leq \min_{t \in \{0,1\}^n} \mathbb{P}_{(x,y) \sim \mathcal{D}}[y \neq \chi_t(x)] + \varepsilon$. This last part of our reasoning only relies on V succeeding in Step 3 and accepting in Step 4, but is independent of the action of the quantum prover. Therefore, we have also established the desired soundness. \blacktriangleleft

► **Remark 14.** In the full version [24], we provide an alternative approach to classically verifying functional agnostic quantum learning, based on the interactive Goldreich-Levin algorithm laid out in [46]. Said verification scheme requires classical membership query access for the prover in order to answer the queries sent by the verifier. We observe that under certain Fourier-sparsity assumptions on the unknown function, the quantum prover can emulate this membership query access. Moreover, the full version [24] also contains variants of Theorem 13 for an SQ verifier and/or a mixture-of-superpositions QSQ prover.

In Theorem 13, the achievable accuracy is limited by $2\sqrt{b^2 - a^2}$. Next, we show that such a limitation is necessary for interactive classical-quantum verification of learning with a sublinear-in- n sample complexity for the classical verifier:

► **Theorem 15.** *Let $\eta \in [0, 1/6)$. Define $a = 0$ and $b = \vartheta = 1 - 2\eta$. Let $\delta = 1/3$ and $\varepsilon = (1 - 2\eta)/3 = \frac{1}{3} \cdot \sqrt{b^2 - a^2}$. Proper 1-PAC verification for the class of n -bit parities w.r.t. $\mathfrak{D}_{\mathcal{U}_n; \geq \vartheta} \cap \mathfrak{D}_{\mathcal{U}_n; [a^2, b^2]}$ by a classical verifier V with access to classical random examples interacting with a quantum prover P with access to mixture-of-superpositions quantum examples requires the verifier to use at least $\Omega(n)$ classical examples.*

Here, we consider η to be a constant and focus on the scaling with n . Theorem 15 tells us that the accuracy lower bound $\varepsilon \geq 2\sqrt{b^2 - a^2}$ in Theorem 13 cannot be significantly improved without at the same time worsening the number of examples used by the classical verifier from n -independent to linear-in- n .

Proof sketch (see [24, Theorem 13] for a detailed proof). We adapt the proof strategy of [70, Theorem 8] to our setting. That is, we use the assumed pair (V, P) of a classical verifier and a quantum prover to construct a testing algorithm T that can distinguish between $\mathcal{D} = \mathcal{U}_{n+1}$ and $\mathcal{D} \in \{(\mathcal{U}_n, (1 - 2\eta)\chi_s)\}_{s \in \{0,1\}^n}$ using $m_T = m_V + \mathcal{O}(1)$ classical random examples of the unknown distribution. This distinguishing task is known to require $\Omega(n)$ classical examples (see [24, Lemma 18] for a proof). T first simulates the interaction of V and P , where V has access to classical examples drawn from \mathcal{D} and P has access to copies of $\rho_{\mathcal{U}_{n+1}}$. Importantly, as \mathcal{U}_{n+1} is a fixed distribution, T can simulate P without requiring any quantum access to \mathcal{D} . After this simulation step, T decides between $\mathcal{D} = \mathcal{U}_{n+1}$ and $\mathcal{D} \in \{(\mathcal{U}_n, (1 - 2\eta)\chi_s)\}_{s \in \{0,1\}^n}$ using $m_T = m_V + \mathcal{O}(1)$ based how the output of V performs on an additional classical test data set drawn from \mathcal{D} . \blacktriangleleft

► **Remark 16.** While Theorem 15 focuses on sample complexities, the proof has immediate computational complexity implications. To see this, notice that (by Theorem 7) it is trivial to classically simulate distributional agnostic quantum Fourier sampling if $\mathcal{D} = \mathcal{U}_{n+1}$ and thus $\phi \equiv 0$. Namely, we first toss a fair coin to decide whether the sampling attempt succeeds or fails, and in the case of success we then sample a uniformly random n -bit string s . Thus, a classical T can efficiently simulate the actions of a quantum P with access to copies of $\rho_{\mathcal{U}_{n+1}}$. Consequently, with the same parameter choices as in Theorem 15, a computationally efficient V would lead to a computationally efficient classical tester T able to distinguish between the uniform distribution and random noisy parities. Therefore, assuming that this distribution testing version of LPN is hard, we cannot meaningfully improve the accuracy lower bound $\varepsilon \geq 2\sqrt{b^2 - a^2}$ in Theorem 13 without losing computational efficiency of V .

4 Distribution-Independent Agnostic Quantum Learning and its Verification

So far, we have focused on agnostic learning under a promise on the input marginal. Namely, we assumed that $\mathcal{D}_{\mathcal{X}_n} = \mathcal{U}_n$ is the uniform distribution. While this focus on distribution-independent learning is common in computational learning theory (to avoid computational infeasibility results), statistical learning theory also often considers a setting of distribution-independent learning, where no prior assumptions on the input marginal of the unknown distribution are made. The sample complexity of learning in this distribution-independent agnostic model has long been fully characterized in the classical case [77, 15, 76]. Moreover, [7] recently established that the optimal quantum sample complexity when using superposition examples coincides with the classical one up to constant factors. Here, we demonstrate that such a limitation of quantum learning also applies to our mixture-of-superpositions examples.

► **Theorem 17** (Formal statement of Theorem 4). *Let $\mathcal{F} \subseteq \{0, 1\}^{\mathcal{X}_n}$ be a benchmark class with VC-dimension $\text{VC}(\mathcal{F}) = d \geq 1$. Then at least*

$$m \geq \tilde{\Omega} \left(\frac{d + \log(1/\delta)}{\varepsilon^2} \right) \quad (24)$$

copies of $\rho_{\mathcal{D}}$, with \mathcal{D} an unknown probability distribution over $\{0, 1\}^n \times \{0, 1\}$, are necessary for distribution-independent quantum agnostic learning of \mathcal{F} with accuracy $\varepsilon \in (0, \frac{1}{4})$ and confidence parameter $\delta \in (0, \frac{1}{2})$. Here, the $\tilde{\Omega}$ hides prefactors logarithmic in d .

Proof sketch (see [24, Theorem 17] for a detailed proof). The d -independent part of the lower bound, $m \geq \Omega \left(\frac{\log(1/\delta)}{\varepsilon^2} \right)$, can be proved for any non-trivial benchmark class \mathcal{F} with a reasoning similar to [7, Lemma 12] and [20, Lemma 5.1]. This argument relies only on basic tools from quantum information, namely on the characterization of the optimal success probability for distinguishing between two quantum states in terms of their trace distance (compare, e.g., [72]), on the Fuchs-van de Graaf inequalities [42], and on the strong concavity of the fidelity [72, Theorem 9.7].

Next, we prove the d -dependent part of the lower bound. For this, we adapt the information-theoretic proof strategy from [7]. Let $\varepsilon \in (0, \frac{1}{4})$. As $\text{VC}(\mathcal{F}) = d$, we can find a set $S = \{x_1, \dots, x_d\} \subset \mathcal{X}_n$ of d distinct points that is shattered by \mathcal{F} . That is, for every $a \in \{0, 1\}^d$, there exists $f_a \in \mathcal{F}$ such that $f_a(x_i) = a_i$ holds for all $1 \leq i \leq d$. Now, for each $a \in \{0, 1\}^d$, we define the probability distribution \mathcal{D}_a over $\mathcal{X}_n \times \{0, 1\}$ as follows:

$$\mathcal{D}_a(x, b) = \begin{cases} \frac{1}{2d} (1 + (-1)^{a_i+b} \cdot 4\varepsilon) & \text{if } x = x_i \\ 0 & \text{else} \end{cases}. \quad (25)$$

By construction, for every $a \in \{0, 1\}^d$ and for every $f \in \mathcal{F}$, we have

$$\mathbb{P}_{(x,b) \sim \mathcal{D}_a} [b \neq f(x)] = \frac{1}{2d} \sum_{i=1}^d ((1 + 4\varepsilon) \delta_{f(x_i), a_i \oplus 1} + (1 - 4\varepsilon) \delta_{f(x_i), a_i}) \quad (26)$$

Thus, $f \in \mathcal{F}$ is a minimum-error concept in \mathcal{F} w.r.t. \mathcal{D}_a if and only if $f|_S(x_i) = a_i$ holds for all $1 \leq i \leq d$. Moreover, if $f|_S(x_i) = c_i$ for some $c \in \{0, 1\}^d$ with $c \neq a$, then such an f incurs excess risk $\frac{4\varepsilon}{d} \cdot d_H(a, c)$, where $d_H(\cdot, \cdot)$ denotes the Hamming distance. Accordingly, any quantum algorithm for distribution-independent quantum agnostic learning \mathcal{F} from m copies of $\rho_{\mathcal{D}_a}$, $a \in \{0, 1\}^d$ unknown, has to output a hypothesis that, when restricted to S ,

becomes a d -bit string that is $\frac{d}{4}$ -close to a in Hamming distance, with success probability $\geq 1 - \delta$. This still holds if the quantum learner is promised in advance that the unknown distribution is supported on S .

Let us consider the CQ state $\rho := \frac{1}{2^d} \sum_{a \in \{0,1\}^d} |a\rangle \langle a| \otimes \rho_{\mathcal{D}_a}^{\otimes m}$ where m is the training data size. We will refer to its classical subsystem as the A -subsystem and to the m quantum registers as subsystems B_1, \dots, B_m . Using that the output of the quantum agnostic learner upon input of $\rho_{\mathcal{D}_a}^{\otimes m}$ is $\frac{d}{4}$ -close to a , one can show that the mutual information between the classical subsystem and the quantum subsystems in ρ satisfies $I(A; B_1, \dots, B_m)_\rho \geq \Omega(d)$, compare [7, Proof of Theorem 12]. Next, since $\rho_{\mathcal{D}_a}^{\otimes m}$ is a tensor power for every a , we have $I(A; B_1, \dots, B_m)_\rho \leq m \cdot I(A; B_1)_\rho$, compare again [7, Proof of Theorem 12]. Thus, the remainder of the proof is concerned with upper bounding $I(A; B_1)_\rho$. As ρ_{AB_1} is a CQ-state, $I(A; B_1)_\rho$ equals the Holevo information of the ensemble $\{(\frac{1}{2^d}, \rho_{\mathcal{D}_a})\}_{a \in \{0,1\}^d}$, compare [78, Exercise 11.6.9]. That is,

$$I(A; B_1)_\rho = S\left(\frac{1}{2^d} \sum_{a \in \{0,1\}^d} \rho_{\mathcal{D}_a}\right) - \frac{1}{2^d} \sum_{a \in \{0,1\}^d} S(\rho_{\mathcal{D}_a}) = S(\bar{\rho}) - \frac{1}{2^d} \sum_{a \in \{0,1\}^d} S(\rho_{\mathcal{D}_a}), \quad (27)$$

where we defined $\bar{\rho} := \frac{1}{2^d} \sum_{a \in \{0,1\}^d} \rho_{\mathcal{D}_a}$ and $S(\rho)$ denotes the von Neumann entropy of a state ρ . It turns out that we can diagonalize both $\bar{\rho}$ as well as $\rho_{\mathcal{D}_a}$ exactly and hence obtain their respective spectra and then also their von Neumann entropies. In fact, the $\rho_{\mathcal{D}_a}$ for different a are related via permutations. Thus, they share the same spectrum. For more details on the diagonalization, we refer the reader to the full version [24, Theorem 17]. The respective entropies are given by

$$S(\bar{\rho}) = 1 + \frac{1}{2} \log(d), \quad S(\rho_{\mathcal{D}_a}) = 1 + \frac{1}{2} \log(d) - \frac{d}{2(d-1)} \log(d) \epsilon^2 + O(\epsilon^4). \quad (28)$$

Plugging, these entropies into the mutual information $I(A; B_1)_\rho$ in Equation (27), we find that $I(A; B_1)_\rho = O(\epsilon^2 \log d)$. Hence, we have an overall upper bound on the mutual information

$$I(A; B_1, \dots, B_m)_\rho \leq O(m\epsilon^2 \log d). \quad (29)$$

Contrasting this with the lower bound found above that is required for agnostic learning, namely of $I(A; B_1, \dots, B_m)_\rho = \Omega(d)$, we hence find that the number of copies m necessary to learn must be at least

$$m = \Omega\left(\frac{d}{\epsilon^2 \log d}\right) = \tilde{\Omega}\left(\frac{d}{\epsilon^2}\right). \quad (30)$$

This proves the d -dependent part of the overall lower bound on the sample complexity of distribution-independent agnostic learning and hence completes the proof. ◀

With Theorem 17, we have seen that mixture-of-superpositions examples do not significantly impact the landscape of distribution-independent agnostic learning compared to their classical data counterpart when focusing on sample complexities. Our next result shows that mixture-of-superpositions examples are also not information-theoretically more powerful than classical examples for verification of learning. Namely, when we consider interactive verification of distribution-independent learning, we match the classical upper and lower bounds of [70] for this task.

► **Theorem 18.** Let $\mathcal{F} \subseteq \{0, 1\}^{\mathcal{X}_n}$ be a benchmark class with VC-dimension $\text{VC}(\mathcal{F}) = d \geq 1$. Assume that (V, P) is an interactive classical-quantum verifier-prover pair that 1-agnostic verifies \mathcal{F} with accuracy parameter $\varepsilon = 1/3$ and confidence parameter $\delta = 1/3$. If we assume that V uses m_V classical random examples and P uses m_P mixture-of-superpositions quantum examples, then $m_V \geq \Omega(\sqrt{d})$, independently of m_P .

Proof. As already argued in the proof of Theorem 15, the reduction strategy used in [70, Theorem 8] is also immediately applicable to a scenario with a quantum prover, because the relevant quantum mixture-of-superpositions state $\rho_{u_{n+1}}$ is completely known and because we can, due to the focus on sample complexity, ignore computational efficiency issues arising from classically simulating a quantum computation. Thus, we get the lower bound as in [70, Theorem 8]. ◀

5 Directions for Future Work

Our work opens up several directions for future research. Firstly, we have demonstrated that mixture-of-superpositions examples enable quantum Fourier sampling-based distributional agnostic learning – in the distribution-dependent setting – by giving explicit learning algorithms for parities and Fourier-sparse functions. At the same time, we have shown that mixture-of-superpositions examples *do not* give a sample complexity advantage over classical random examples in the distribution-independent setting. Thus, it is of natural interest to go beyond these initial results and to further understand both the potential and the limitations of mixture-of-superpositions examples for agnostic learning, for example exploring their use for other model classes. Additionally, one of the primary motivations for the mixture-of-superpositions examples introduced here is the difficulty in developing techniques for Fourier sampling from standard quantum superposition examples in the distributional agnostic setting. However, while no such techniques have been developed to date, there are no established hardness results. As such, it remains unclear whether mixture-of-superpositions examples are indeed strictly more powerful than standard quantum superposition examples. In light of this, it would be interesting to understand whether there is a separation between the power of the two oracle models.

Additionally, in this work we have explored the extent to which one can delegate problems of *supervised learning of Boolean functions* to untrusted quantum servers. However, there is a plethora of other learning problems whose delegation to quantum algorithms would be desirable to investigate. A natural first example would be the delegation and verification of *distribution learning* [62] problems to quantum servers. In particular, we note that, unlike for supervised learning, in the distribution learning context even the realizable setting seems non-trivial. Alternatively, there is a multitude of learning and testing problems for which the object to be learned or tested is inherently quantum. Examples include testing or learning to predict properties of quantum states [1, 2, 55], quantum measurements [30], or quantum processes [31, 20, 36, 21, 54]. For many of these problems there are known exponential separations between what can be achieved by quantum algorithms with or without access to a quantum memory (see [56, 4, 28, 53, 21, 27]). This prompts a natural question: Can quantum learning algorithms without a quantum memory efficiently delegate such learning or testing problems to untrusted quantum algorithms with access to a quantum memory?

Moreover, from a technical perspective, there are concrete ways in which our learning algorithms and verification procedures might be improved. Firstly, for distributional agnostic learning Fourier-sparse functions, our learning algorithms are 2-agnostic – i.e., they yield hypotheses whose risk is guaranteed to be at most twice the risk of the optimal model,

plus some desired tolerance ε . Ideally, however, one would like to give 1-agnostic learning algorithms. Secondly, our verification procedures do not work for arbitrary unknown functions or distributions, but require prior assumptions. The learning problems we consider remain classically hard under these assumptions, and are therefore still sufficient for demonstrating the existence of problems which can be efficiently delegated/verified by classical learning algorithms, although not efficiently solved without delegation. Nevertheless, it seems interesting to understand the extent to which our assumptions are truly necessary.

Finally, our work is motivated by a desire to understand the potential for classical clients to profit from the advantages of quantum learning algorithms in a (realistic) world where quantum computations are delegated to untrusted quantum servers with access to proprietary quantum data resources. However, at least currently and for the intermediate-term future, any quantum server will only have access to “noisy intermediate scale quantum” (NISQ) devices [74]. As such, to bring our results closer to immediate practical relevance, it is of interest to explore the extent to which classical clients can verify untrusted NISQ-friendly quantum machine learning algorithms based on the variational optimization of parameterized quantum circuits. Indeed, there has recently been progress on the statistical foundations of such hybrid quantum-classical learning algorithms [22, 3, 11, 23, 33, 25, 26], and it would be of significant interest to enrich this developing understanding with insight into the complexity of classical verification. Additionally, in our work we have explored the setting in which a classical client interacts with a quantum server, with access to a quantum data oracle. However, one may also consider quantum clients of limited complexity (e.g., NISQ clients) that interact with more powerful quantum servers. This would serve to enrich our growing understanding of the capability of NISQ algorithms from a complexity-theoretic perspective [29].

References

- 1 Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007. doi:10.1098/rspa.2007.0113.
- 2 Scott Aaronson. Shadow tomography of quantum states. *SIAM Journal on Computing*, 49(5):STOC18–368, 2019. doi:10.1137/18M120275X.
- 3 Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021. doi:10.1038/s43588-021-00084-1.
- 4 Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature Communications*, 13(1):1–9, 2022. doi:10.1038/s41467-021-27922-0.
- 5 Srinivasan Arunachalam, Sourav Chakraborty, Troy Lee, Manaswi Paraashar, and Ronald de Wolf. Two new results about quantum exact learning. *Quantum*, 5:587, November 2021. doi:10.22331/q-2021-11-24-587.
- 6 Srinivasan Arunachalam and Ronald de Wolf. Guest column: A survey of quantum learning theory. *SIGACT News*, 48, 2017. doi:10.1145/3106700.3106710.
- 7 Srinivasan Arunachalam and Ronald de Wolf. Optimal quantum sample complexity of learning algorithms. *Journal of Machine Learning Research*, 19(71):1–36, 2018. URL: <http://jmlr.org/papers/v19/18-195.html>.
- 8 Srinivasan Arunachalam, Alex B. Grilo, and Henry Yuen. Quantum statistical query learning, 2020. arXiv:2002.08240.
- 9 Alp Atıcı and Rocco A. Servedio. Improved bounds on quantum learning algorithms. *Quantum Information Processing*, 4(5):355–386, 2005. doi:10.1007/s11128-005-0001-2.
- 10 Alp Atıcı and Rocco A. Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2007. doi:10.1007/s11128-007-0061-6.

- 11 Leonardo Banchi, Jason Pereira, and Stefano Pirandola. Generalization in quantum machine learning: A quantum information standpoint. *PRX Quantum*, 2(4):040321, 2021. doi:10.1103/PRXQuantum.2.040321.
- 12 Roozbeh Bassirian, Bill Fefferman, and Kunal Marwaha. On the Power of Nonstandard Quantum Oracles. In Omar Fawzi and Michael Walter, editors, *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*, volume 266 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11:1–11:25, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.TQC.2023.11.
- 13 Debajyoti Bera and Sagnik Chatterjee. Efficient quantum agnostic improper learning of decision trees, 2022. arXiv:2210.00212.
- 14 Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. doi:10.1137/S0097539796300921.
- 15 Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965, 1989. doi:10.1145/76359.76371.
- 16 Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, October 2009. doi:10.1109/focs.2009.36.
- 17 Nader H. Bshouty and Jeffrey C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing*, 28(3):1136–1153, 1998. doi:10.1137/S0097539795293123.
- 18 Ran Canetti and Ari Karchmer. Covert learning: How to learn with an untrusted intermediary. In *Theory of Cryptography Conference*, pages 1–31. Springer, 2021. doi:10.1007/978-3-030-90456-2_1.
- 19 Matthias C. Caro. Quantum learning boolean linear functions w.r.t. product distributions. *Quantum Information Processing*, 19, 2020. doi:10.1007/s11128-020-02661-1.
- 20 Matthias C. Caro. Binary classification with classical instances and quantum labels. *Quantum Machine Intelligence*, 3, 2021. doi:10.1007/s42484-021-00043-z.
- 21 Matthias C. Caro. Learning quantum processes and hamiltonians via the pauli transfer matrix, 2022. arXiv:2212.04471.
- 22 Matthias C. Caro and Ishaun Datta. Pseudo-dimension of quantum circuits. *Quantum Machine Intelligence*, 2:14, 2020. doi:10.1007/s42484-020-00027-5.
- 23 Matthias C. Caro, Elies Gil-Fuster, Johannes Jakob Meyer, Jens Eisert, and Ryan Sweke. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum*, 5:582, 2021. doi:10.22331/q-2021-11-17-582.
- 24 Matthias C. Caro, Marcel Hinsche, Marios Ioannou, Alexander Nietner, and Ryan Sweke. Classical verification of quantum learning, 2023. arXiv:2306.04843.
- 25 Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles. Generalization in quantum machine learning from few training data. *Nature Communications*, 13, 2022. doi:10.1038/s41467-022-32550-3.
- 26 Matthias C. Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T. Sornborger, Lukasz Cincio, Patrick J. Coles, and Zoë Holmes. Out-of-distribution generalization for learning quantum dynamics. *Nature Communications*, 14, 2023. doi:10.1038/s41467-023-39381-w.
- 27 Kean Chen, Qisheng Wang, Peixun Long, and Mingsheng Ying. Unitarity estimation for quantum channels. *IEEE Transactions on Information Theory*, 69(8):5116–5134, 2023. doi:10.1109/TIT.2023.3263645.
- 28 Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585. IEEE, 2022. doi:10.1109/FOCS52979.2021.00063.
- 29 Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of nisq, 2022. arXiv:2210.07234.

- 30 Hao-Chung Cheng, Min-Hsiu Hsieh, and Ping-Cheng Yeh. The learnability of unknown quantum measurements. *Quantum Info. Comput.*, 16(7–8):615–656, May 2016. doi:10.5555/3179466.3179470.
- 31 Kai-Min Chung and Han-Hsuan Lin. Sample Efficient Algorithms for Learning Quantum Channels in PAC Model and the Approximate State Discrimination Problem. In Min-Hsiu Hsieh, editor, *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)*, volume 197 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:22, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.TQC.2021.3.
- 32 Andrew W Cross, Graeme Smith, and John A Smolin. Quantum learning robust against noise. *Physical Review A*, 92(1):012327, 2015. doi:10.1103/PhysRevA.92.012327.
- 33 Yuxuan Du, Zhuozhuo Tu, Xiao Yuan, and Dacheng Tao. Efficient measure for the expressivity of variational quantum algorithms. *Physical Review Letters*, 128(8):080506, 2022. doi:10.1103/PhysRevLett.128.080506.
- 34 Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 1956. URL: <https://www.jstor.org/stable/2237374>.
- 35 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.
- 36 Marco Fanizza, Yihui Quek, and Matteo Rosati. Learning quantum processes without input control, 2022. arXiv:2211.05005.
- 37 Bill Fefferman and Shelby Kimmel. Quantum vs. Classical Proofs and Subset Verification. In Igor Potapov, Paul Spirakis, and James Worrell, editors, *43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018)*, volume 117 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:23, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.MFCS.2018.22.
- 38 Vitaly Feldman. On the power of membership queries in agnostic learning. *Journal of Machine Learning Research*, 10(7):163–182, 2009. URL: <http://jmlr.org/papers/v10/feldman09a.html>.
- 39 Vitaly Feldman. Distribution-specific agnostic boosting. In *Proceedings Innovations in Computer Science – ICS 2010*, ICS 2010, pages 241–250, Tsinghua University, Beijing, China, 2010. Tsinghua University Press. arXiv:0909.2927.
- 40 Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. On agnostic learning of parities, monomials, and halfspaces. *SIAM Journal on Computing*, 39(2):606–645, 2009. doi:10.1137/070684914.
- 41 Joseph F Fitzsimons. Private quantum computation: an introduction to blind quantum computing and related protocols. *npj Quantum Information*, 3(1):23, 2017.
- 42 Christopher A Fuchs and Jeroen Van De Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999. doi:10.1109/18.761271.
- 43 Alexandru Gheorghiu, Theodoros Kapourniotis, and Elham Kashefi. Verification of quantum computation: An overview of existing approaches. *Theory of computing systems*, 63:715–808, 2019. doi:10.1007/s00224-018-9872-3.
- 44 Oded Goldreich and Leonid A Levin. A hard-core predicate for all one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 25–32, 1989. doi:10.1145/73007.73010.
- 45 Shafi Goldwasser, Guy N. Rothblum, Jonathan Shafer, and Amir Yehudayoff. Interactive proofs for verifying machine learning. *Electron. Colloquium Comput. Complex.*, 20(58), 2020. URL: <https://ecc.weizmann.ac.il/report/2020/058>.

- 46 Shafi Goldwasser, Guy N Rothblum, Jonathan Shafer, and Amir Yehudayoff. Interactive proofs for verifying machine learning. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ITCS.2021.41.
- 47 Parikshit Gopalan, Adam Tauman Kalai, and Adam R Klivans. Agnostically learning decision trees. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 527–536, 2008. doi:10.1145/1374376.1374451.
- 48 Parikshit Gopalan, Ryan O’Donnell, Rocco A Servedio, Amir Shpilka, and Karl Wimmer. Testing fourier dimensionality and sparsity. *SIAM Journal on Computing*, 40(4):1075–1100, 2011. doi:10.1137/100785429.
- 49 Alex B Grilo, Iordanis Kerenidis, and Timo Zijlstra. Learning-with-errors problem is easy with quantum samples. *Physical Review A*, 99(3):032314, 2019. doi:10.1103/PhysRevA.99.032314.
- 50 Aram W. Harrow and David J. Rosenbaum. Uselessness for an oracle model with internal randomness. *Quantum Inf. Comput.*, 14(7-8):608–624, 2014. doi:10.26421/QIC14.7-8-5.
- 51 David Haussler. Decision theoretic generalizations of the pac model for neural net and other learning applications. *Information and computation*, 100(1):78–150, 1992. doi:10.1016/0890-5401(92)90010-D.
- 52 Lunjia Hu and Charlotte Peale. Comparative Learning: A Sample Complexity Theory for Two Hypothesis Classes. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 72:1–72:30, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2023.72.
- 53 Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, et al. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022. doi:10.1126/science.abn7293.
- 54 Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes, 2023. arXiv:2210.14894.
- 55 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. doi:10.1038/s41567-020-0932-7.
- 56 Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021. doi:10.1103/PhysRevLett.126.190505.
- 57 Jeffrey C. Jackson, Christino Tamon, and Tomoyuki Yamakami. Quantum dnf learnability revisited. In *International Computing and Combinatorics Conference*, pages 595–604. Springer, 2002. doi:10.1007/3-540-45655-4_63.
- 58 Adam Tauman Kalai, Yishay Mansour, and Elad Verbin. On agnostic boosting and parity learning. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC ’08, pages 629–638, New York, NY, USA, 2008. Association for Computing Machinery. doi:10.1145/1374376.1374466.
- 59 Varun Kanade and Adam Kalai. Potential-based agnostic boosting. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc., 2009. URL: <https://proceedings.neurips.cc/paper/2009/file/13f9896df61279c928f19721878fac41-Paper.pdf>.
- 60 Varun Kanade, Andrea Rocchetto, and Simone Severini. Learning dnfs under product distributions via μ -biased quantum fourier sampling. *Quantum Information & Computation*, 19(15&16):1261–1278, 2019. doi:10.26421/QIC19.15-16.
- 61 Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998. doi:10.1145/293347.293351.

- 62 Michael Kearns, Yishay Mansour, Dana Ron, Ronitt Rubinfeld, Robert E Schapire, and Linda Sellie. On the learnability of discrete distributions. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 273–282, 1994.
- 63 Michael J. Kearns, Robert E. Schapire, and Linda M. Sellie. Toward efficient agnostic learning. *Mach. Learn.*, 17(2–3):115–141, November 1994. doi:10.1007/BF00993468.
- 64 Michael R Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer, 2008. doi:10.1007/978-0-387-74978-5.
- 65 Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. *SIAM Journal on Computing*, 22(6):1331–1348, 1993. doi:10.1137/0222080.
- 66 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993. doi:10.1145/174130.174138.
- 67 U. Mahadev. Classical verification of quantum computations, 2018. arXiv:1804.01082.
- 68 Pascal Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990. URL: <https://www.jstor.org/stable/2244426>.
- 69 Ashley Montanaro. The quantum query complexity of learning multilinear polynomials. *Information Processing Letters*, 112(11):438–442, 2012. doi:10.1016/j.ipl.2012.03.002.
- 70 Saachi Mutreja and Jonathan Shafer. Pac verification of statistical algorithms. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 5021–5043. PMLR, July 2023. URL: <https://proceedings.mlr.press/v195/mutreja23a.html>.
- 71 Anand Natarajan and Chinmay Nirkhe. A Distribution Testing Oracle Separating QMA and QCMA. In Amnon Ta-Shma, editor, *38th Computational Complexity Conference (CCC 2023)*, volume 264 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:27, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2023.22.
- 72 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- 73 Jack O’Connor. Delegating machine learning with succinct proofs. Master’s thesis, University of Warwick, 2021.
- 74 John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018. doi:10.22331/q-2018-08-06-79.
- 75 Rocco A. Servedio and Steven J. Gortler. Equivalences and separations between quantum and classical learnability. *SIAM Journal on Computing*, 33(5):1067–1092, 2004. doi:10.1137/S0097539704412910.
- 76 Michel Talagrand. Sharper bounds for gaussian and empirical processes. *The Annals of Probability*, pages 28–76, 1994. doi:10.1214/aop/1176988847.
- 77 Vladimir N. Vapnik and Alexei Ya. Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971. doi:10.1137/1116025.
- 78 Mark M. Wilde. From classical to quantum shannon theory, 2011. arXiv:1106.1445.
- 79 Chi Zhang. An improved lower bound on query complexity for quantum pac learning. *Information Processing Letters*, 111(1):40–45, 2010. doi:10.1016/j.ipl.2010.10.007.