The Hardness of Learning Quantum Circuits and its Cryptographic Applications

Bill Fefferman*[1], Soumik Ghosh†[1], Makrand Sinha‡[2], and Henry Yuen§[3]

[1]University of Chicago
[2]University of Illinois Urbana-Champaign
[3]Columbia University

Abstract

We show that concrete hardness assumptions about learning or cloning the output state of a random quantum circuit can be used as the foundation for secure quantum cryptography. In particular, under these assumptions we construct secure one-way state generators, digital signature schemes, quantum bit commitments, and private key encryption schemes. We also discuss evidence for these hardness assumptions by analyzing the best-known quantum learning algorithms, as well as proving black-box lower bounds for cloning and learning given state preparation oracles.

Our random circuit-based constructions provide concrete instantiations of quantum cryptographic primitives whose security do not depend on the existence of one-way functions. The use of random circuits in our constructions also opens the door to NISQ-friendly quantum cryptography. We discuss noise tolerant versions of our OWSG and digital signature constructions which can potentially be implementable on a noisy quantum computer. On the other hand, they are still secure against noiseless quantum adversaries, raising the intriguing possibility of a useful implementation of an end-to-end cryptographic protocol by a near-term quantum computer. Finally, our explorations suggest that the rich interconnections between learning theory and cryptography in classical theoretical computer science also extend to the quantum setting.

*wjf@uchicago.edu
†soumikghosh@uchicago.edu
‡msinha@illinois.edu
§henry.yuen@columbia.edu

1 Introduction

The hardness of learning and modern cryptography are inextricably linked in the world of classical computing. On one hand, hard cryptographic problems have served as the basis of showing that various learning tasks are intractable (see e.g., [KV94, KS06, DLSS14]). Conversely, the hardness of various learning tasks have been used to construct fundamental cryptographic primitives, like pseudorandom functions [IL90, BFKL93, OS16], and practical cryptographic protocols [Reg24, LPR10]. Furthermore, this connection between learning and cryptography extends further and has shed light on fundamental questions in the related areas of pseudorandomness [NW94] and circuit lower bounds [LMN93, CIKK16].

In the quantum world, our understanding of the analogous interconnections is quite lacking. In one direction, some prior works have used classical cryptographic assumptions, like quantum-secure one-way functions or LWE, to argue about the hardness of learning quantum states and circuits in different contexts (see [AGS20, ZLK+24]), but the complexity of many fundamental quantum learning tasks remain open. However, the *converse* direction — using hardness of quantum learning as a foundation for cryptography — has not received much attention, unlike the classical case[1]. Arguably, one difficulty in establishing such a connection is that classical cryptographic primitives appear insufficient to fully capture the inherently quantum nature of tasks involving learning quantum states or unitaries [CLS24]. This raises an interesting question that forms the basis of this paper:

Can we base quantum cryptography on the hardness of quantum learning?

Our results

We propose several fine-grained assumptions about quantum learning tasks related to random quantum circuits, give evidence for it in the black-box model, and use these assumptions to construct several quantum cryptographic primitives. Incidentally, this exploration also leads to several interesting results such as "NISQ-friendly" quantum cryptography and a deeper understanding of various quantum learning tasks and cryptographic primitives. In more detail:

- **Quantum hardness of learning assumptions.** We posit that it is computationally intractable to learn the output states of *unknown* random quantum circuits (Computational No-Learning Assumption) or to clone them (Computational No-Cloning Assumption). We formulate these conjectures and discuss the differences between them in Section 1.1. We also give evidence for these conjectures by proving lower bounds in the black-box model.

- **Cryptography from hardness of quantum learning.** We show that these hardness of learning assumptions can be used to construct quantum cryptographic primitives. In particular, we construct one way state generators (OWSGs) and quantum-secure digital signatures based on the Computational No-Learning Assumption, and we construct quantum bit commitments based on the Computational No-Cloning Assumption. Recent results have uncovered the tantalizing possibility of reducing the hardness assumptions behind such primitives beyond what is possible in classical cryptography [AQY22, MY22b, BCQ23, KQST23]. Our hardness

[1] As of the last several months, this is starting to change; see Section 1.5 for discussion of recent work on this topic.

assumptions about random circuits do not appear to imply the existence of classical one-way functions, and thus our constructions concretely instantiate this possibility.

- **Quantum cryptography for the NISQ era.** The use of random circuits in our constructions also opens the door to *NISQ-friendly quantum cryptography*. We demonstrate that our OWSG and digital signature constructions can be made noise tolerant and thus implementable on a noisy quantum computer, provided we have a quantum channel to pass quantum states as cryptographic keys. On the other hand, they are still secure against *noiseless* quantum adversaries. This raises the intriguing possibility of a near-term demonstration of *quantum cryptographic advantage*: using a NISQ device to run a cryptographic protocol whose security relies on fewer assumptions than what is possible in classical cryptography.

Organization

In Section 1.1, we introduce our hardness of learning assumptions. Section 1.3 gives a summary of the constructions of various cryptographic primitives. Section 1.6 provides a discussion of other implications of our results. Section 1.5 and Section 1.7 briefly discuss recent works that are related to this paper as well as mention some interesting directions that arise for future explorations. Section 2 introduces the notation and some preliminaries. The black-box lower bound to support our hardness assumptions are given in Section 3. Section 4 includes details of the cryptographic constructions.

1.1 Our hardness of learning assumptions

We propose two main assumptions regarding the hardness of quantum learning: the Computational No-Learning Assumption and the Computational No-Cloning Assumption. We precisely define these conjectures and discuss evidence for them.

The Computational No-Learning Assumption

In what follows, we let $\mathscr{C}$ denote a class of quantum circuits; for example, a concrete choice of $\mathscr{C}$ is the ensemble of 1D brickwork circuits with depth $d = \log^2 n$.[2]

Conjecture 1.1 (Computational No-Learning)**.** *Let $\mathscr{C}$ be a class of circuits. The $(\mathscr{C}, \varepsilon, \delta)$-Computational No-Learning Assumption stipulates the following. For all polynomial-time quantum algorithms A, for all sufficiently large n, we have that*

$$\mathbb{P}\left[\begin{array}{c} D \in \mathscr{C}_n \\ |\langle C|D\rangle|^2 \geq \varepsilon \end{array} : \begin{array}{c} C \leftarrow \mathscr{C}_n \\ D \leftarrow A(|C\rangle^{\otimes \mathrm{poly}(n)}) \end{array}\right] \leq \delta$$

Here, $C \leftarrow \mathscr{C}_n$ denotes that C is sampled uniformly from the subset of circuits in $\mathscr{C}$ that act on n qubits, and D is sampled from the output of the algorithm A given polynomially-many copies of $|C\rangle = C|0^n\rangle$. The output D is interpreted as a description of an n-qubit quantum circuit, and $|D\rangle = D|0^n\rangle$.

[2] We pick out $d = \log^2 n$ here because it happens to be a depth at which the existing state-of-the-art learning algorithms, like those in [HLB+23, LL24], fail to be efficient. Even if the learning algorithms were to be improved, any depth beyond which the efficiency of those learning algorithms fail is a good design choice for our assumptions and cryptographic constructions.

In words, the Computational No-Learning Assumption says that it a polynomial-time quantum algorithm A, given only polynomially-many copies of the output $|C\rangle$ of a randomly-chosen circuit C from the circuit ensemble $\mathscr{C}$, cannot produce a classical description of another circuit D from the same class $\mathscr{C}_n$ and whose output state $|D\rangle$ approximates $|C\rangle$. We note that the algorithm A only has access to copies of the state[3] generated by the random circuit and the circuit description itself is *not* known to the algorithm. Furthermore, Conjecture 1.1 actually refers to an entire *family* of assumptions, one for each choice of circuit class $\mathscr{C}$ and parameters (ε, δ). Throughout this paper we consider a fixed circuit family for convenience (e.g., 1D brickwork circuits with depth $\log^2 n$ unless otherwise specified). We elaborate on the different parameter settings for (ε, δ) below. For clarity we abbreviate "Computational No-Learning" as "No-Learning". Oftentimes we will abbreviate "$(\varepsilon, \varepsilon)$-No-Learning" as "$\varepsilon$-No-Learning".

Relationship between parameter values. (ε, δ)-No-Learning implies (ε', δ')-No-Learning for all $\varepsilon' \geq \varepsilon$ and $\delta' \geq \delta$. This is argued by taking the contrapositive: if there exists an efficient algorithm A that produced an ε'-fidelity approximation $|D\rangle$ of $|C\rangle$ with probability at least δ', then A *also* produced (with the same D) a ε-fidelity approximation with probability at least δ.

A conservative assumption would be to conjecture $(1 - 1/\operatorname{poly}(n))$-No-Learning (i.e., it is hard for a polynomial-time algorithm to produce a high-fidelity approximation of a state with high probability). On the other hand, given our current understanding of quantum state learning techniques, it also seems plausible to conjecture $2^{-\Omega(n)}$-No-Learning (i.e., it is hard for a polynomial-time algorithm to produce even an *exponentially bad* approximation with *exponential small* probability). That said, we remark that for the purposes of implementing the cryptographic constructions in this paper, the parameters we require are fairly mild. For instance, we merely need the $(1 - 1/\operatorname{poly}(n))$-No-Learning assumption for the cryptographic constructions that assume a noise-free circuit, and even for our NISQ-friendly constructions, the ε-No-Learning assumption for any negligible[4] function $\varepsilon(n)$ is sufficient.

Edge cases. The hardness assumption cannot hold for $\varepsilon \ll 2^{-n}$; this is because a quantum algorithm can output a uniformly random circuit D, and $|D\rangle$ will have fidelity at least 2^{-n} with $|C\rangle$ in expectation. Similarly, the hardness assumption cannot hold for $\delta \leq |\mathscr{C}_n|^{-1} = 2^{-\Omega(nd)}$, because the algorithm can always simply guess the description C of the state $|C\rangle$.

Proper vs. improper learning. We point out that the assumption is about the hardness of producing a circuit description D that *comes from the same class* $\mathscr{C}$ as C. In the learning theory literature this is known as the setting of *proper* learning, where the learner has to output a hypothesis from the same concept class as the true underlying concept. One can also consider the setting of *improper* learning, where the output description D can come from a more general class of circuits. For example, the circuit learning algorithms of [HPS24, LL24] are improper learners: given copies of output states of depth d circuits, they produce descriptions of circuits with depth greater than d. One can consider hardness of learning conjectures in the improper setting as well; ours is the more

[3] We stress that $|C\rangle$ does not mean the classical description of C, but rather the state generated by the circuit C on the all zero input.

[4] A function is negligible if it goes to zero faster than any inverse polynomial.

conservative assumption (because the learner has the additional constraint of outputting a circuit from $\mathscr{C}$).

The Computational No-Cloning assumption

In this section, we postulate a computational version of the famous No-Cloning principle of quantum mechanics. At a high level, it stipulates that given polynomially copies of the output state of a random quantum circuit, it is intractable to produce an additional copy of the state, or even an approximation of it.

Conjecture 1.2 (Computational No-Cloning)**.** *Let $\mathscr{C}$ denote a class of circuits. The $(\mathscr{C}, \varepsilon, \delta)$-Computational No-Cloning Assumption stipulates the following. For all polynomial-time quantum algorithms A, for all polynomials $k(n)$, for all sufficiently large n, we have that*

$$\mathbb{P}\left[|\langle C|^{\otimes(k+1)}|\phi\rangle|^2 \geq \varepsilon : \begin{array}{c} C \leftarrow \mathscr{C}_n \\ |\phi\rangle \leftarrow A(|C\rangle^{\otimes k}) \end{array}\right] \leq \delta$$

Here, $C \leftarrow \mathscr{C}_n$ denotes that C is sampled uniformly from the subset of circuits in $\mathscr{C}$ that act on n qubits, and a $(k+1)n$ qubit state $|\phi\rangle$ is sampled from the output of the algorithm A, given polynomially-many copies of $|C\rangle = C|0^n\rangle$.

As before, the algorithm A only has access to copies of the state generated by the random circuit and not the circuit description itself. We also abbreviate "Computational No-Cloning" as "No-Cloning" and also abbreviate "$(\varepsilon, \varepsilon)$-No-Cloning" as "$\varepsilon$-No-Cloning".

Corollary 1.3. *The ε-No-Cloning Assumption (Conjecture 1.2) implies the ε^c-No-Learning Assumption (Conjecture 1.1) for some $0 < c < 1$.*

(Proof Sketch). This follows from taking the contrapositive. If the description of the circuit C could be approximately learned with squared fidelity ϵ^c by some quantum polynomial-time algorithm A (i.e, quantum state learning is easy), we can devise an efficient cloning algorithm B that first runs A in a coherent fashion on the k copies to learn an approximation D, synthesizes an extra copy of $|D\rangle$ (which is supposed to represent the $(k+1)$st copy of $|C\rangle$), and then uncomputes A to recover the original k copies of $|C\rangle$. The fact that this strategy obtains ε fidelity with $|C\rangle^{\otimes(k+1)}$ follows from a calculation virtually identical to that of Claim 4.10, which is part of the proof that the No-Cloning Assumption implies the existence of quantum commitments. This violates the ε-No-Cloning Assumption. □

Cloning is potentially easier than learning. Corollary 1.3 means that learning is potentially an *easier* task than learning. It could be that we can clone just one more copy of the state but we still cannot learn the state.

There is some suggestive evidence of this fact from the work of Nehoran and Zhandry [NZ24], who construct a quantum oracle with respect to which a collection of states is efficiently clonable, but it is not efficiently "telegraphable," given only one quantum sample. Informally, telegraphing means getting a classical string out of one copy of a quantum state by "deonconstructing" it, from which one copy of the state can again be "reconstructed." While the notion is qualitatively reminiscent of a learning task, this is formally different from the notion of learning in our paper.

1.2 Barriers and evidence for the hardness assumptions

1.2.1 The barriers

What are the prospects of proving Conjecture 1.1 or Conjecture 1.2 outright? First, we note that the conjectures are necessarily *computational*, meaning that they are intrinsically about the limits of efficient quantum computation. This is because it *is* possible for an exponential-time quantum algorithm to learn a circuit description with high probability, given only polynomially-many quantum samples (and thus also solve the cloning task with similar sample complexity). One method is to use the classical shadows protocol of [HKP20]; we describe this in more detail in Section 2. We note that Morimae and Yamakawa already observed that there is always an exponential-time attack on one-way state generators [MY22a] via shadow tomography; this is essentially the same observation.

Furthermore, we note that proving our hardness conjectures would have dramatic implications for *classical* complexity theory. The classical shadows-based algorithm described in Section 2 can be implemented in polynomial time assuming the complexity inclusion $\mathrm{NP}^{\#\mathrm{P}} \subseteq \mathrm{BQP}$.[5] Therefore Conjecture 1.1 implies $\mathrm{P} \neq \mathrm{PSPACE}$. While this doesn't imply (say) $\mathrm{P} \neq \mathrm{NP}$, for all intents and purposes this would represent a similar breakthrough in mathematics and complexity theory. The longstanding difficulty of proving such complexity separations pose a barrier to proving our hardness assumptions.

1.2.2 The evidence

We now discuss the evidence for our hardness assumptions. First we discuss the No-Learning assumption.

Examining best known learning algorithms. Recently there has been progress on obtaining efficient algorithms for learning states of bounded circuit complexity [HLB+23, FGZ24, LL24]. The algorithm of Fefferman, Ghosh, and Zhan [FGZ24] requires the use of oracle access to the circuit itself, rather than only having copies of the output state $C|0^n\rangle$, which we do not consider in this paper. For algorithms that learn using copies of the output state alone, the current state-of-the-art is due to Landau and Liu [LL24], who show the following:

Theorem 1.4 ([LL24])**.** *Fix an integer $k > 0$. There is an algorithm that, for all $\varepsilon > 0, \delta > 0$, given copies of an unknown quantum state $|C\rangle = C|0^n\rangle$ for some depth-d circuit acting on a k-dimensional lattice with two-qubit gates, outputs the description of a depth $(2k+1)d$ circuit D such that $\||C\rangle - |D\rangle\| \leq \varepsilon$ with probability $1 - \delta$. Furthermore, the algorithm uses*

$$\frac{n^4 \cdot 2^{O(d^k)}}{\varepsilon^4} \log \frac{n}{\delta}$$

copies of $|C\rangle$, and runs in time

$$\frac{n^4 \cdot 2^{O(d^k)}}{\varepsilon^4} \log \frac{n}{\delta} + \left(\frac{nd}{\varepsilon}\right)^{O(d^{k+1})}.$$

Here, the $O(\cdot)$ suppresses dependence on k, which we treat as a constant.

[5] It was recently shown by Hiroka and Hsieh [HH24] that efficient state learning is possible if $\mathrm{PP} \subseteq \mathrm{BQP}$, which represent an improved complexity upper bound.

For intuition, consider some parameter settings. Suppose ε, δ are some fixed constants (like 0.001).

1. When $d = O(1)$, then the sample complexity is $O(n^4 \log n)$, and the time complexity is $n^{O(d^k)}$.
2. When $d = O(\log^{1/k} n)$, then the sample complexity is still $\mathrm{poly}(n)$, but the time complexity becomes quasipolynomial $2^{\mathrm{polylog}\, n}$.
3. When $d \gg \log n$, then both the sample and time complexity become superpolynomial.

In more detail, the learning algorithms of both Huang, et al. [HLB+23] and Landau and Liu [LL24] revolve around the idea of learning so-called "local inversions" of the circuit C, which are small subcircuits V that "undo" part of the overall circuit to be learned: $(V^\dagger \otimes \mathbb{I})|C\rangle \approx |0^k\rangle \otimes |\theta\rangle$ for some $n-k$ qubit state $|\theta\rangle$. In other words, some small number k of qubits have been disentangled from the state $|\psi\rangle$.

If the circuit depth d is small, then the local inversions have size $O(d^k)$ (assuming a k-dimensional circuit architecture), and the learning algorithm can brute force over all possible subcircuits of size $O(d^k)$. Once the local inversions have been learned, the challenge is to "stitch" all of the local inversions (which overlap with each other) in a consistent way.

The complexity of learning a single local inversion takes time at least $2^{O(d^k)}$. Thus with a randomly chosen circuit of depth that asymptotically grows faster than $O(\log n)$ (e.g., $d = \log^2 n$), there are many more possibilities than is possible to consider in polynomial time. Furthermore, searching through candidate local inversions of a random circuit appears to be an "all or nothing" task: either a candidate successfully inverts a local patch, or it will likely scramble the state even further. Thus it does not seem possible to make the learning algorithms of [HLB+23, LL24] "gracefully fail" by dialing back their runtime to being polynomial in n. Looking at the prototypical algorithmic ideas in these papers suggests that if one is in fact limited to polynomial time algorithms, then the typical fidelity will likely be exponentially small.

Just to give a crude sense of the time and sample complexity cost of these algorithms, let us take $n = 70$ qubits, depth $d = 24$, and a 2D geometry, where the values are taken from Google's latest quantum advantage experiment [MVM+23]. Then, from Theorem 1.4, the algorithm by Landau and Liu [LL24] needs at least

$$\frac{n^4 \cdot 2^{d^k}}{\epsilon^4} \log \frac{n}{\delta}$$

samples to get better than $1-\epsilon$ fidelity with more than $1-\delta$ success probability. Hence, to get 0.99 fidelity with 0.99 success probability, the algorithm needs *at least* 5.26×10^{177} trillion quantum samples. Similarly, by plugging in values to the formula in Theorem 1.4 to compute the time complexity, the algorithm needs *at least* 10^{174} trillion hours. This shows that both the sample and time complexity are too large to be practical, which is heuristic evidence that the task of learning the state is hard. However, an important caveat is that this is just one algorithm and it is very plausible that there is a much more optimized version of these learners for more specialized ensembles, like the one Google is using, by exploiting unique properties of the same. We leave it for future work to rigorously analyze the tradeoffs between performance and runtime of existing learning algorithms. Such analysis is crucial for obtaining further evidence for the validity of our hardness assumptions.

Worst-case hardness from post-quantum assumptions. The No-Learning Assumption is an *average-case* hardness assumption, because it is about learning over the uniform distribution over circuits. One can also wonder about the hardness of *worst-case* learning. Zhao, et al. [ZLK+24] showed that, assuming that subexponential-time quantum computers cannot solve the RingLWE problem, any quantum algorithm that learns the circuit description given copies of the n-qubit state requires at least $\exp(\Omega(\min\{G, n\}))$ time, where G is size of the quantum circuit to be learned. RingLWE is a version of the Learning With Errors (LWE) problem, whose assumed hardness underlies most proposals for post-quantum cryptography (i.e., cryptography that can be run on classical computers, but are secure against quantum computers). When the number of gates G is superlogarithmic, then the time complexity lower bound is superpolynomial.

One may wonder why, given the results of Zhao, et al. [ZLK+24], we need to make a separate assumption about the hardness of learning quantum circuits. Given the widespread belief in the security of various versions of LWE (for example this underlies much of NIST's recent standardization of recommended post-quantum cryptosystems [FIP23, Nat24]), it would seem that hardness of (Ring)LWE automatically implies the hardness of learning circuits from [ZLK+24] and therefore all our applications of Conjecture 1.1 follow.

There are two issues with this reasoning. The first is that the result of Zhao, et al. [ZLK+24] implies the hardness of learning under a very particular distribution of quantum circuits: namely, ones that encode the RingLWE problem, which are quite structured[6] and are statistically far from truly random quantum circuits. The second and most important issue is that, while the RingLWE hardness can be viewed as evidence for our hardness conjectures, it appears to be a significantly harder statement to prove. For one, the hardness of RingLWE trivially implies P $\neq$ NP, one of the major open questions in mathematics. On the other hand, P $\neq$ NP is *not* known to be implied by Conjecture 1.1. Indeed, there is emerging evidence that Conjecture 1.1 is a reduced assumption compared to P $\neq$ NP (i.e., there are mathematical worlds in which P = NP but tasks related to learning quantum circuits is still hard [KQST23]). Since one of our primary motivations is to base cryptography on the fewest mathematical assumptions possible, we treat our hardness assumptions as being more basic and plausible than any post-quantum hardness assumption.

Evidence for No-Cloning assumption. We can similarly examine the evidence for the No-Cloning assumption. As mentioned above, since the cloning is potentially an easier task, the No-Cloning conjecture is a *stronger* assumption than No-Learning. However, as far as we are aware, there are no known algorithms for cloning that perform significantly better than ones for learning. One could consider, for example, the optimal pure-state cloning map that was analyzed by Werner [Wer98]. This map, which is essentially to project the input state onto the $(k+1)$-fold symmetric subspace, is provably the most sample-efficient procedure for taking copies $|\psi\rangle^{\otimes k}$ of an *arbitrary* input state $|\psi\rangle$ (not necessarily one generated by a polynomial-size circuit) and producing an approximation of $|\psi\rangle^{\otimes k+1}$. Werner [Wer98] showed the best achievable fidelity in general is

$$\frac{\binom{2^n+k-1}{k}}{\binom{2^n+k}{k+1}} = \frac{k+1}{2^n}$$

which is exponentially small for $k = \text{poly}(n)$.

[6] In more detail, these are based on constructions of *pseudorandom states* from one-way functions [JLS+18].

Aside from Werner's optimal cloning map, as far as we are aware the best algorithms for the cloning task are based on first solving the learning task. As we discussed before, such algorithmic ideas, if we restrict them to run in polynomial time, seem unable to achieve anything better than exponentially small fidelity. We leave it is an interesting open problem to come up with an algorithms – even ones that run in subexponential time – for cloning states of random circuits that do not learn the circuit description first.

Black-box lower bounds. We give more evidence to support the No-Learning and No-Cloning assumptions by proving lower bounds in the black-box model. We desire a black-box model which captures the analogous properties in the white-box setting. For instance, one property we would like to capture is the typical distribution of the output state of a random quantum circuit, which mimics the distribution of a Haar-random state. Another property we would like to capture in the black-box model is the existence of a learning algorithm that uses only polynomially-many copies of the state but is allowed to make exponentially many black-box queries – this would be analogous to the existence of the exponential-time, polynomial-sample complexity learning algorithm.

With such considerations in mind, we formalize a *state preparation oracle* model where the oracle $\mathcal{O}$ takes as input a state $|i\rangle|0^n\rangle$ and outputs $|i\rangle|\psi_i\rangle$ for some Haar-random state $|\psi_i\rangle$. The oracle can be accessed in superposition, and there is no guarantee about what the oracle does when the second register is initialized to something other than all zeroes. Intuitively, each index i corresponds to a different polynomial-size circuit, but the algorithm is not allowed to exploit the structure of the circuit except to prepare the resulting output state. We prove the following:

Theorem 1.5 (Black-box lower bounds for cloning)**.** *There exists a state preparation oracle $\mathcal{O}$ such that all T-query quantum query algorithms getting k copies of $|\psi_J\rangle$ for a uniformly random index $J \in [2^n]$ satisfy*

$$\mathrm{F}(\rho, |\psi_J\rangle\langle\psi_J|^{\otimes k+1}) \leq 2^{-n/4}(2T + k + 1)$$

where ρ is the output of the query algorithm and $\mathrm{F}(\cdot)$ denotes the fidelity function.

In other words, unless either the number of queries T or the number of copies k are $2^{\Omega(n)}$, the expected fidelity of cloning is exponentially small. Since the ability to learn implies the ability to clone, this also shows a similar lower bound for the learning task. We elaborate on the model and prove Theorem 1.5 in Section 3.

1.3 Cryptography from our hardness assumptions

We now summarize our main cryptographic applications using the hardness of quantum learning assumptions from Section 1.1.

1.3.1 One-way state generators from hardness of learning

A *one way state generator* (OWSG) is an efficient algorithm that takes as input a classical key k, and outputs a quantum state $|\psi_k\rangle$ from that key. Anyone with the key k can efficiently verify that $|\psi_k\rangle$ is the correct output. Furthermore, given polynomially many copies of the $|\psi_k\rangle$ for a randomly chosen k, it should be computationally hard for an adversary to produce another key k' such that the corresponding state $|\psi_{k'}\rangle$ is close to the original output $|\psi_k\rangle$. (For a formal definition, see Section 4.1). Introduced by Morimae and Yamakawa [MY22b], OWSGs are a quantum

Protocol 1.6. Random Circuit OWSG

Generation algorithm: Given input key $C \in \{0,1\}^{r(n)}$, interpret it as a description of an n-qubit circuit C from the ensemble $\mathscr{C}_n$. Output $|C\rangle = C|0^n\rangle$.

Verification procedure: Given input $C \in \{0,1\}^{r(n)}$ and a state $|D\rangle$ on n qubits, apply $C^\dagger$ to the state, and measure. Accept if the result is all zeroes, and reject otherwise.

Figure 1: Construction of one-way state generator from random circuits.

analogue of one-way functions, which are functions efficiently computable in the forwards direction but computationally difficult to invert.

The No-Learning Assumption (Conjecture 1.1) is essentially *equivalent* to the existence of a OWSG, namely the Random Circuit OWSG described below in Figure 1. For a range of parameters, this is immediate: for negligible ε (i.e., ε goes to 0 faster than any inverse polynomial), the ε-No-Learning Assumption is easily seen to be equivalent to the security of the Random Circuit OWSG. When ε is larger, say even up to $1 - 1/\operatorname{poly}(n)$, the equivalence still holds; this relies on hardness amplification techniques for OWSGs [MY22a, BQSY24]. We note that this equivalence was also independently observed by Hiroka and Hsieh in a recent preprint [HH24].

Although a OWSG is not immediately cryptographically useful by itself, it is now known that OWSGs can be used as a primitive building block for a variety of quantum cryptosystems. In their papers defining OWSGs [MY22b, MY22a], Morimae and Yamakawa showed that OWSGs can be used to build bounded-time-secure digital signature schemes. In a breakthrough work, Khurana and Tomer showed that OWSGs imply the existence of quantum bit commitments [KT24a], which in turn imply other functionalities such as quantum zero knowledge proofs for NP and secure multiparty computation [BCQ23]. Therefore, using the Random Circuit OWSG, we obtain concrete implementations of these cryptosystems on quantum computers, without relying on the use of one-way functions.

An attractive feature of our Random Circuit OWSG is that it seems potentially amenable to implementation on noisy quantum computers, and thus the corresponding cryptosystems may be realizable in the near- and medium-term. We discuss noise tolerant versions of our random circuit-based cryptographic protocols in Section 1.3.1.

1.3.2 Simple quantum commitments from the hardness of cloning

A commitment scheme enables two parties (known as a "committer" and a "receiver") to perform the cryptographic equivalent of putting a message in a sealed envelope that is opened later. In a quantum commitment scheme, a committer upon getting a bit b generates a bipartite pure state $|\psi_b\rangle_{\mathrm{AB}}$, and sends the register B to the receiver; this is the "commitment phase" of the protocol and is the analogue of sending the sealed envelope. At this point the receiver should not be able to tell what the bit b is.

Later, in the "reveal phase" of the protocol, the committer announces the bit b and sends the remaining register A of $|\psi_b\rangle$ to the receiver, who can uncompute the state to check its validity. The security of the commitment scheme ensures that the committer cannot "change his mind" in

Protocol 1.7. Commitment scheme based on random circuits

Commitment phase: To commit to bit $b = 0$, the committer prepares the state

$$|\psi_0\rangle_{\mathrm{AB}} := \frac{1}{\sqrt{|\mathscr{C}_n|}} \sum_{C \in \mathscr{C}_n} \left(|C\rangle^{\otimes k} \otimes |0^n\rangle \right)_{\mathrm{A}} \otimes |\hat{C}\rangle_{\mathrm{B}}$$

where $|\hat{C}\rangle$ represents the *classical description* of the circuit C.

To commit to bit $b = 1$, then prepare the state

$$|\psi_1\rangle_{\mathrm{AB}} := \frac{1}{\sqrt{|\mathscr{C}_n|}} \sum_{C \in \mathscr{C}_n} |C\rangle_{\mathrm{A}}^{\otimes (k+1)} \otimes |\hat{C}\rangle_{\mathrm{B}} \ .$$

The committer sends register B of the state $|\psi_b\rangle$ to the receiver.

Reveal phase. The committer reveals the bit b to the receiver, and also sends the remaining register A of $|\psi_b\rangle$. To verify, the receiver will uncompute the unitary that synthesizes $|\psi_b\rangle$ and check that the all zeroes state is obtained.

Figure 2: Commitment scheme based on random circuits

between the commit and reveal phases to convince the receiver he had committed to the opposite bit $1 - b$.

Recently, quantum commitments have become a centerpiece of the zoo of quantum cryptographic primitives [AQY22, MY22b, BCQ23, KT24a]. As mentioned, Khurana and Tomer showed that OWSGs can be used to construct quantum bit commitments in a generic way [KT24a]. Therefore our No-Learning Assumption implies the existence of secure quantum commitments. However the construction of commitments in [KT24a] is quite involved, requiring an intricate sequence of transformations mimicking the classical transformation from one-way functions to pseudorandom generators [HILL99].

We construct a simple quantum bit commitments based on the Computational No-Cloning Assumption (Conjecture 1.2), and furthermore the analysis is fairly direct and straightforward. We describe the construction below. As discussed in Section 1.1, the No-Cloning Assumption implies the No-Learning Assumption. The ease of obtaining a commitment scheme from the No-Cloning Assumption suggests that the No-Cloning Assumption could be strictly stronger than the No-Learning Assumption (because it appears that obtaining commitments from OWSGs requires an intricate analysis [KT24a]).

We present and analyze the bit commitment scheme in detail in Section 4.2.

1.4 NISQ-friendly quantum cryptography

An important challenge in the field of quantum computing is to find a practical use case for noisy, near-term quantum (i.e., NISQ) computers. Although great strides have been made recently in demonstrating principles of error-correction and fault-tolerance on quantum devices [BEG+24, AABA+24], large-scale implementations of many quantum algorithms of interest (e.g., Shor's,

Grover's, Hamiltonian simulation, etc) still seem quite a ways off. Thus there is significant interest in finding an application that (a) can be implemented on an a NISQ device, (b) has some advantage over classical computers/protocols, and (c) is practically useful. We show that our hardness assumptions yield quantum cryptosystems that satisfy these three criteria.

Our random circuits-based cryptosystems are arguably NISQ-friendly. Random quantum circuits have been investigated extensively on real hardware platforms ever since Google's original quantum supremacy announcement in 2019 [AABea19]. The lack of structure in random quantum circuits is advantageous for maximizing quantum advantage while minimizing the burden on the NISQ device. Furthermore, the effective noise model when executing random quantum circuits often becomes quite simple [AABea19, DHJB24].

Furthermore, implementing cryptosystems based on our hardness assumptions would concretely realize the possibility of having secure quantum cryptography using reduced assumptions as compared to classical cryptography (for example, we do not need to assume P $\neq$ NP or that one-way functions exist) [Kre21, AQY22, MY22b, BCQ23, KT24b]. This would represent what we call "quantum cryptographic advantage."

We presented protocols that solve useful cryptographic tasks: digital signatures, bit commitments, encryption, and more. Although these protocols are not immediately NISQ-friendly out of the box, we show how to "NISQ-ify" some of them so that they are.

1.4.1 NISQ-friendly one-way state generators

While implementing the Random Circuit OWSG on a realistic quantum computer, noise can degrade the fidelity of the output state. For certain choices of noise parameters and depth regimes, the fidelity can be inverse polynomial in the number of qubits. If that happens, the success probability of any verification procedure would degrade similarly. Although the OWSG may be secure against any polynomial-time adversary, it may not be very useful if the "honest user" (e.g., someone with the key) tries to run it on a noisy quantum computer. On the other hand, making a OWSG tolerant to noise may give greater leeway to break its security.

However, if we make a sufficiently strong assumption (e.g., the ε-No Learning assumption for negligible ε), there is a noticeable gap between the success probability of a *noisy* verifier (who has the circuit description) and any *noiseless* polynomial-time adversary (who doesn't have the circuit description): the adversary cannot produce any approximation to the state except with negligible fidelity. We can amplify this gap to obtain a NISQ-friendly OWSG, where both the generation and verification algorithms can be run on a noisy quantum computer, but the OWSG also retains its security against noiseless adversaries. We achieve this amplification via a *computational Chernoff bound*.

Computational Chernoff bounds. A standard way to amplify the security of a OWSG is via *parallel repetition* [MY22a, BQSY24]: the key to the amplified OWSG corresponds to a t-tuple of independently chosen random circuits $(C_1, \ldots, C_t)$, and the output is the tensor product of the corresponding states $|C_1\rangle \otimes \cdots \otimes |C_t\rangle$. Verification proceeds by checking that the i'th block of n qubits is in the state $|C_i\rangle$ for each $i = 1, \ldots, n$. Intuitively, if it is somewhat hard for the adversary to learn the output of one random quantum circuit, then it should be *very* hard for the adversary to simultaneously learn the output of *many* random quantum circuits. However, because the noisy fidelity is small, standard parallel repetition theorems do not directly work because the fidelity of

the overall state on nt qubits would degrade exponentially with t, making it $\mathrm{negl}(n)$, which means even honest verifiers would fail to see a non-trivial signal.

Our technical innovation is to use *threshold parallel repetition* for amplification—this is a OWSG $G^{t,k}$ consisting of t independent copies of G, but instead of verifying that all t copies have been inverted, the verification algorithm checks that at least k out of the t have been inverted. To bound its security we prove new *computational Chernoff bounds* for OWSGs. A detailed discussion, with the theorem statements, can be found in Section 5.1.

1.4.2 NISQ-friendly digital signatures

As a direct application of our NISQ-friendly OWSG we obtain a NISQ-friendly quantum digital signature scheme. At a high level, a digital signature scheme (with quantum public keys) is a method for a *signer* to generate a *signature* for a message in a way that a third party *verifier* (using a quantum public key posted by the user beforehand) can verify that the signature belongs to the message (and in particular, the message or the signature have not been changed).

Morimae and Yamakawa [MY22b] showed that such quantum digital signature schemes can be directly constructed from OWSGs by adapting the famous Lamport construction of digital signatures [Lam79]. We show that by plugging in the NISQ-friendly random circuit OWSG, their digital signature scheme becomes NISQ-friendly as well. This gives example of an end-to-end cryptographic task for NISQ-devices whose hardness relies on an innately quantum conjecture.

We present the scheme and the analysis in detail in Section 5.2.

1.4.3 On noise assumptions and asymptotics

Note that the only noise assumption we need is that the fidelity of the signal should at most be inverse polynomially large. However, an observant reader would notice that $\mathscr{C}_n$ is an ensemble of sufficiently deep circuits. For certain noise models, for e.g. constant rate of depolarizing noise per gate, or more generally, constant rate of unital noise per gate, the output converges to the maximally mixed state [AB96, DNS+22] exponentially fast in the depth of the circuit, which would cause inverse superpolynomially large signal decay at large depths. However, there are three perspectives on why our proposal is still relevant to near-term experiments.

Firstly, the results proving convergence to the maximally mixed state [DNS+22, AGL+23] are asymptotic statements, whereas real experiments have finite system sizes. Thus, there is a discrepancy between what theoretically happens when we scale up the system size and what is experimentally observed for a fixed system size. For example, in quantum advantage demonstrations with random circuits, the experimentalists have observed signatures of long range entanglement and evidence of convergence to the Porter-Thomas distribution when the output state is measured in the standard basis [AABea19, MVM+23]. Note that Porter-Thomas distribution is far in total variation distance from the uniform distribution, where the latter is what we get when we measure a maximally mixed state in the standard basis. If the system were indeed close to being maximally mixed, we would neither see long-range entanglement nor Porter-Thomas type behavior. Thus, sampling from the uniform distribution is not a good approximation to the realistic output distribution, even though the distributions are close asymptotically [AB96, DNS+22].

There are other classical samplers, such as the recent one proposed in [AGL+23], which differ from simply sampling the uniform distribution. However, the sampler in [AGL+23] achieves a

smaller total variation distance than the uniform distribution only at a depth of approximately $\sim \log n$. Properties of real experiments—such as the presence of long-range entanglement—indicate that they do not operate at logarithmic depth but rather in a much deeper regime. Thus, in the same way as the uniform distribution, it is unclear if the sampler of [AGL+23] is directly applicable to real experiments, even though its classical spoofing distribution is again asymptotically close to the noisy target distribution.

Secondly, note that the fidelity of the output state of a noisy circuit, comprised of two qubit gates and a single-qubit, uncorrelated noise channel acting upon each qubit after the application of each gate, is

$$F = (1-\epsilon)^{2s},$$

or the probability that no errors occurred anywhere in the system. Here, where s is the circuit size and ϵ is the noise rate per qubit. If ϵ is at most $\sim \frac{1}{n}$, $F \approx e^{-\Theta(s\cdot\epsilon)}$. Hence, one valid regime for inverse polynomial decay in fidelity is when ϵ is $\sim \frac{1}{n \log n}$ and the system size s is $\sim n \log^2 n$. In certain models, the structure of the output state becomes even simpler. One of the a noise models for which this is manifestly true is the white noise model [AABea19, BEG+24, DHJB24]. According to this model, if the noise per gate is unital, and if the noise rate ϵ is at most $\sim \frac{1}{n}$, then the output state of the circuit can be written as

$$\rho_{\mathsf{out}} = F\rho_{\mathsf{noiseless}} + (1-F)\frac{\mathbb{I}}{2^n},$$

that is, as a linear combination of $\rho_{\mathsf{noiseless}}$, which is what the output state would have been if there were no noise, and the maximally mixed state. While the noise rate per gate going down with n is unrealistic for extremely large system sizes, it is nonetheless a reasonable model of real experiments as structural properties of experimental output states match the white noise output state. In fact, judging by the recent progress in random quantum circuit experiments, e.g. [AABea19, MVM+23, RABFF+23], it seems quite reasonable to model realistic noise as going down with system size.

Thirdly, note that in all of the previous discussions, we have assumed noise to be unital. But real noise is complicated and it is unclear if any of the above results (e.g., convergence to the maximally mixed state or the white noise model) are realistic for near-term experiments. As an example of surprising behavior with more general noise models, researchers have recently studied the effects of non-unital noise channels in random quantum circuits [FGG+24, MAG+24, OJF23]. Non-unital noise is ubiquitous in real world experiments, as witnessed by, e.g., readout errors, T_1 decay for superconducting systems, and photon loss in bosonic systems [S+23, W+22, ZL+21]. In particular, certain structural properties that are true for unital noise at some regimes, like anti-concentration or convergence to the maximally mixed state, fail in the presence of any constant rate non-unital noise channel [FGG+24], and hardness or easiness results that assume these properties, like [AA11, Aha03, BFNV19, AGL+23], do not seem to work.

1.5 Related work

We discuss the relationship between our work and some concurrent, independent works that recently appeared.

Cryptography from assumptions about random circuits: Khurana and Tomer [KT24b] also study the quantum cryptographic implications of hardness assumptions about random cir-

cuits. Their hardness assumption posits that it is #P-hard to estimate the output probabilities of a random quantum circuit, given a classical description of the circuit. This is a well-studied hardness assumption and is the theoretical basis for many quantum supremacy proposals (e.g., [AA11, BFNV19, AABea19]). Combined with the complexity-theoretic assumption that $\mathrm{P}^{\#\mathrm{P}} \not\subseteq \mathrm{ioBQP/qpoly}$, [KT24b] show how to construct *quantum one-way puzzles*, which in turn implies the existence of quantum bit commitments through their previous work [KT24a].

While random circuits are at the core of the hardness assumptions of our paper as well as [KT24b], the similarities quickly end. Khurana and Tomer are positing the hardness of a *classical-input, classical-output* task: given the description of a quantum circuit and a string, estimate the probability of outputting the string. Our hardness assumptions, on the other hand, are about *quantum-input* tasks: given *quantum* copies of the output state of a random circuit, either learn or clone it. Importantly, the circuit description is *not* known to the adversary.

Furthermore, our motivations have some differences: they are motivated by basing quantum cryptography on separations between *decision* complexity classes, whereas we are primarily motivated by the connection between quantum learning problems (which involve quantum inputs) and quantum cryptography.

On a different note, in Bostanci, Haferkamp, Hangleiter, and Poremba [BHHP24], the authors construct quantum cryptography from a suite of assumptions about random IQP circuits. Depending on the type of assumption, the authors get quantum trapdoor functions, quantum pseudoentanglement, and candidate constructions of efficient pseudorandom unitaries. There are large differences between this work and our work, in terms of the nature of assumptions, the justifications for hardness, the flavours of cryptography that one gets, and the query lower bounds. These two works represent complementary explorations into cryptography from two different random ensembles, ours involving brickwork random circuits, and theirs involving IQP circuits.

Cryptography from assumptions about quantum states, protocols, and noise: Qian, Raizes, and Zhandry [QRZ24] study the quantum cryptographic implications of a new "search-type" assumption they call *classical* $\rightarrow$ *quantum extrapolation*, where the goal is to extrapolate the rest of a bipartite pure state given the first register measured in the computational basis. They show that the hardness of this extrapolation task implies the existence of quantum bit commitments and is implied by the existence of various quantum public-key primitives. We view their work as studying the cryptographic implications of a (conceptually new) "generic" assumption, where they do not specify how exactly to generate the hard, inextrapolable bipartite states. On the other hand, we are focused on the cryptographic implications of a "concrete" assumption, where we instantiate the underlying primitive (OWSG, quantum commitment) with a concrete algorithmic implementation. This is similar to the difference between assuming that *some* one-way function exists, versus assuming that a *specific* one-way function exists (e.g., the RSA function or the LWE function).

Morimae, Shirakawa, and Yamakawa [MSY24] give a characterization of the complexity assumptions needed for a class of protocols for proofs of quantumness; in particular they show that one-way puzzles (the same primitive constructed by [KT24b]) are necessary and sufficient. Their goal is squarely aimed at understanding the complexity of proofs of quantumness in the abstract, and less by having concrete instantiations of quantum cryptographic primitives. For related papers on the interplay between one-way puzzles, proofs of quantumness, and quantum cryptography, also

see [CFSY23, HM24, CGGH24].

Hiroka and Hsieh [HH24] studies the hardness of learning efficiently generatable pure states. The main focus of this paper is a PP upper bound on this task. They also base some cryptography on their hardness assumptions, like the existence of one-way state generators. The crucial difference between this work and ours is that they consider one particular learning assumption, whereas we consider a suite of learning and distinguishing assumptions, basing different flavours of cryptography on each. We also discuss in detail how to make our protocols NISQ-implementable.

Poremba, Quek, and Shor [PQS24] put forward a new quantum-inspired primitive called Learning Stabilizers with Noise (LSN), which deals with decoding a random stabilizer code in the presence of local depolarizing noise. Their primitive implies (statistically hiding and computationally binding) bit commitments. Their goal is to construct a new natively quantum assumption for quantum cryptography, as opposed to NISQ-friendliness of their protocols.

NISQ-friendly cryptography: Finally, we comment on the relationship between the proposals by [AH23, BBF+24] to use NISQ devices to perform certifiable randomness generation. Similarly to our work, they propose a cryptographic task that can be performed on a NISQ device, and whose security can be based on hardness assumptions about random circuits.

However a significant difference is that the verification procedure in their protocols are inherently inefficient; it requires exponential time even using a quantum computer as it requires approximating output probabilities of a random quantum circuit. In contrast, our NISQ-friendly digital signature scheme is efficiently implementable.

1.6 Summary

Our exploration also uncovers a deeper understanding of various quantum learning tasks and cryptographic primitives.

Fine-grained distinctions in learning and cryptography. Our work connects fine-grained learning tasks to understanding the relative power of different cryptographic primitives. This opens up a potential new approach to understand both. For instance, cloning is potentially an easier task than learning[7]. Our work shows that the hardness of learning assumptions is essentially equivalent to the existence of OWSGs from random circuits, while the hardness of cloning can be used to construct a fairly simple bit commitment scheme. This indicates that commitments are likely to be a more minimalistic cryptographic primitive than OWSGs. Our conclusion is consistent with oracular evidence in [BMM+24, BCN24].

Practical applications of random quantum circuits. Random circuits have been extensively studied in the context of quantum advantage in the NISQ-era. Several experimental groups around the world (for e.g, [AABea19, MVM+23, ZWD+20, ZCY+21, ZL+21] have claimed practical demonstrations of quantum advantage with sampling tasks involving random circuits. Furthermore, there also has been extensive effort to build theoretical foundations of quantum advantage based on such sampling tasks (see e.g., [AC17, BFNV19, AGL+23, FGG+24]), but even if we are able to demon-

[7]There is some suggestive evidence here in the form of a black-box separation for a related task [NZ24].

strate practical advantage with such circuits, one major challenge that remains is to use NISQ devices or random quantum circuits to solve practically useful problems.

By proposing useful cryptographic applications of random circuits that are "NISQ-friendly", our work takes one further step in addressing this challenge and complements the recent work on certified random number generation with random circuits ([AH23, BBF+24]).

Minimal assumptions for cryptography. Our work contributes to understanding the minimal theoretical assumptions needed for quantum cryptography. While in the classical world the existence of one-way functions is widely believed to be necessary for cryptography [Imp95], in the quantum context this may not be the case. In particular, recent work has given black-box evidence in which P = NP and yet single-copy secure pseudorandom quantum states still exist [KQST23, Kre21]. This suggests that certain quantum cryptographic primitives are possible even without the existence of one-way functions, for e.g., see [AQY22, BCQ23, LMW24]. On the other hand, in the white-box setting, all currently known constructions of such quantum pseudorandom states require the existence of quantum-secure one-way functions [JLS+18]. Consequently, a major challenge that remains is to construct quantum cryptographic primitives which do not rely on the existence of one-way functions and are based on concrete hardness assumptions. Our work addresses this by constructing cryptography based on the hardness of quantum learning.

1.7 Future directions

The connections between the hardness of quantum learning and cryptography leads to several interesting directions that require more exploration:

- **Relations between learning and cloning.** We posed two concrete assumptions about the hardness of quantum learning for random circuits. The natural open question that remains is to understand the differences between these different learning tasks (No-Learning versus No-cloning), their relative hardness, and to understand the security parameters needed for the hardness assumptions.

- **Quantum cryptography from concrete hardness assumptions.** Innately quantum hardness of learning assumptions, like the No-Learning and No-Cloning assumptions, give a natural direction to give concrete instantiations of other cryptographic primitives on assumptions that might be weaker than one-way functions.

- **NISQ-friendly cryptography and practical applications.** While several cryptographic constructions presented in this paper are NISQ-friendly, others, such as for quantum bit commitments, involve operations that are not realistic for near-term devices, for instance, coherently implementing a superposition over all quantum circuits. This leads to the tantalizing possibility of finding other NISQ-friendly cryptographic primitives.

- **Quantum pseudorandomness and connections to complexity.** There are fundamental open questions regarding pseudorandomness properties of states produced by random quantum circuits. Using hardness of learning to probe such questions is an interesting open direction. Along this line of inquiry, one might further expect to find deeper connections between learning, cryptography, pseudorandomness, state complexity classes and circuit lower bounds.

2 Preliminaries

In this section, we give an overview of our notation, collect some useful facts, theorems, and lemmas, that will be used in the rest of the paper.

Notation. We write $[t]$ to denote the set $\{1, 2, \ldots, t\}$. For an integer n we write 1^n to denote its unary representation. We write $\mathrm{poly}(n)$ to denote $p(n)$ for some polynomial p. We write $\mathrm{negl}(n)$ to denote a negligible function, that is, some function $\delta(n)$ such that for all polynomials $p(n)$, for all sufficiently large n, $\delta(n) \leq \frac{1}{p(n)}$. In other words, $\delta(n)$ goes to 0 faster than any inverse polynomial.

The identity operator is denoted by $\mathbb{I}$. For an operator A we write $\|A\|_1$ to denote its trace norm, i.e., the sum of its singular values. For two density matrices ρ, σ we write $\mathrm{F}(\rho, \sigma) := \mathrm{Tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2$ to denote the fidelity between them.

Quantum circuits. All quantum circuits in this paper use single- and two-qubit gates from some discrete universal gate set that includes the Clifford group, i.e., the set of unitaries generated by $\mathrm{CNOT}, \mathrm{H}, \mathrm{S} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$. The size of a circuit is the number of gates in it. We write $\mathscr{C}_{n,d}$ to denote the set of all n-qubit, depth-d circuits where the gates are arranged in a 1D brickwork architecture.

We write $\mathscr{C}_n$ to denote the set $\mathscr{C}_{n,d}$ for $d = \log^2(n)$. Sometimes we will omit the subscript n and write $\mathscr{C}$ when the number of qubits is clear from context. We write $C \leftarrow \mathscr{C}_n$ to denote sampling a uniformly random circuit C from $\mathscr{C}_n$. For a circuit $C \in \mathscr{C}_{n,d}$, we write $|C\rangle$ to denote the state resulting from applying C to the all zeroes input, i.e.,

$$|C\rangle := C|0^n\rangle \ .$$

For a circuit C, we write $\hat{C}$ to denote its *classical description* (to distinguish it from the unitary operator corresponding to C).

A *quantum polynomial-time (QPT)* algorithm A is a uniform family of circuits $\{C_n\}_{n\in\mathbb{N}}$ such that there is a polynomial $p(n)$ such that the size of C_n is bounded by $p(n)$ for all n. Here, uniform means that there is a polynomial-time classical algorithm that, on input 1^n, outputs the classical description of C_n. In a QPT algorithm, we also allow the circuits to initialize some number of ancilla qubits to $|0\rangle$ and trace them out at the end of the computation (and thus a QPT algorithm in general corresponds to a quantum channel).

Classical shadows. The classical shadows protocol of [HKP20] gives a method to perform measurements on a small number of copies of a quantum state ρ, and use the measurement outcomes to estimate the expectation values of ρ with respect to a much larger number of observables. The method is sample efficient, but not necessarily computationally efficient. We first summarize their protocol.

Protocol 2.1. Classical shadows protocol

Parameters: k, t integers such that t divides k.
Observables: $A_1, \ldots, A_M$.
Input: k copies of an n-qubit state ρ.

1. Sample k random Clifford circuits $S_1, \ldots, S_k$. Applying S_j to the j'th copy of ρ and measure in the standard basis to obtain sample $x_j \in \{0,1\}^n$.

2. For each $j \in [k]$, compute the classical description of the Hermitian matrix called a *shadow*:
$$\hat{\rho}_j = (2^n + 1) S_j^\dagger |x_j\rangle\langle x_j| S_j - \mathbb{I}.$$

3. Divide the k samples into t groups of k/t, and for each group $r \in [t]$, and each observable $i = 1, \ldots, M$, compute the following estimator:
$$\hat{a}_i^{(r)} = \frac{t}{k} \sum_{j=1+k(r-1)/t}^{kr/t} \mathrm{Tr}(A_i \hat{\rho}_j) \ .$$

4. For each $i = 1, \ldots, M$, compute the final estimator
$$\hat{a}_i = \mathrm{median}\left\{ \hat{a}_1^{(1)}, \ldots, \hat{a}_1^{(t)} \right\} \tag{2.1}$$

Lemma 2.2 (Performance of the classical shadows protocol)**.** *Let $\{A_1, \ldots, A_M\}$ denote a set of n-qubit observables, i.e., each A_j is a Hermitian matrix. Then the classical shadows protocol of Protocol 2.1 will with probability at least $1 - \delta$ produce estimates $\{\hat{a}_1, \ldots, \hat{a}_M\}$ such that $|\hat{a}_i - \mathrm{Tr}(A_i \rho)| \leq \varepsilon$ provided that*

$$k \geq \frac{204}{\varepsilon^2} \log\left(\frac{2M}{\delta}\right) B \qquad \text{and} \qquad t \geq 2 \log\left(\frac{2M}{\delta}\right)$$

where $B = \max_i \mathrm{Tr}((A_i - 2^{-n} \mathrm{Tr}(A_i)\mathbb{I})^2)$.

Proof. This is proved in the the Supplementary Information of [HKP20]. □

Corollary 2.3. *Let $\mathscr{C}_{n,d}$ denote the circuit ensemble described in Section 2. There exists a quantum algorithm that, given input $|C\rangle^{\otimes k}$ where $|C\rangle = C|0^n\rangle$ for some circuit $C \in \mathscr{C}_{n,d}$, with probability at least $1 - \delta$ outputs a classical description of a circuit $D \in \mathscr{C}_{n,d}$ such that*

$$|\langle C|D\rangle|^2 \geq 1 - \varepsilon$$

provided that

$$k \geq \mathcal{O}\left(\frac{1}{\varepsilon^2} \log\left(\frac{|\mathscr{C}_{n,d}|}{\delta}\right)\right) \ .$$

Furthermore, if $\mathrm{NP}^{\#\mathrm{P}} \subseteq \mathrm{BQP}$ this task can be done in quantum polynomial time.

Proof. This follows directly from Lemma 2.2 where, enumerating the circuits as $\mathscr{C}_{n,d} = \{C_1, C_2, \ldots\}$, we define the observable

$$A_i = C_i|0^n\rangle\langle 0^n|C_i^\dagger \ .$$

The quantity B in the statement of Lemma 2.2 can be upper-bounded by a constant $\mathcal{O}(1)$, leading to the stated sample complexity bounds.

On input $|C\rangle^{\otimes k}$, the quantum algorithm will run the classical shadows protocol, and obtain estimates $\{\hat{a}_i\}$. With probability $1-\delta$ there is at least one index i such that $\hat{a}_i \geq 1-\varepsilon$ (namely, the one corresponding to the circuit C that generated the input state), so the algorithm can pick one arbitrarily (e.g. randomly selecting one) and outputting the corresponding circuit description.

We now consider the complexity of this algorithm. Observe that, as stated, the algorithm uses exponential time, simply for computing the estimates $\hat{a}_i$ for all observables $A_1, \ldots, A_M$, of which there are exponentially many. We can reformulate this algorithm so that it runs in polynomial-time, assuming that $\mathrm{NP}^{\#\mathrm{P}} \subseteq \mathrm{BQP}$.

Consider the following decision problem: given shadows $(S_1, x_1), \ldots, (S_k, x_k)$ (i.e. descriptions of n-qubit Clifford circuits along with an n-bit string) and integers $1 \leq x < y \leq M$ represented in binary, decide if there exists an $x \leq i \leq y$ such that the corresponding estimator $\hat{a}_i$ defined in Equation (2.1) is at least $1-\varepsilon$. Note that each $i \in [M]$ the estimator $\hat{a}_i$ can be computed in polynomial-time given an oracle for #P. Therefore a nondeterministic polynomial-time Turing machine with an oracle to #P can nondeterministically guess an index i such that $\hat{a}_i \geq 1-\varepsilon$.

Thus if $\mathrm{NP}^{\#\mathrm{P}} \subseteq \mathrm{BQP}$, the quantum algorithm can perform the shadow measurements, and then perform binary search in polynomial time to identify such an index i with high probability. This concludes the "Furthermore" part of the corollary. □

Note that by the results of an independent recent work, by Hiroka and Hsieh [HH24], the complexity theoretic inclusion can be improved to $\mathrm{PP} \subseteq \mathrm{BQP}$.

3 Black-box lower bounds for quantum learning and cloning

In this section, we give evidence for our hardness of learning conjectures by proving lower bounds in the black-box model that amongst other attack rule out efficient shadow tomography type attacks. We first introduce the black-box setting where we model output states of sufficiently deep random circuits by a Haar random state.

Query algorithm with a state preparation oracle

1. Independently sample $N = 2^n$ many n-qubit Haar random states $\mathcal{S} = \{|\psi_1\rangle, \ldots, |\psi_N\rangle\}$. Also, independently sample a uniformly random index $J \in [N]$.

2. The quantum algorithm is given some $k = \text{poly}(n)$ copies of the target state $|\psi_J\rangle$ and black-box access to a state preparation oracle that generates each $|\psi_i\rangle$ in the following way:

$$O_{\mathcal{S}}|i\rangle|0\rangle = |i\rangle|\psi_i\rangle.$$

   We can implement the above oracle unitarily by arbitrarily extending each $|\psi_i\rangle$ to an independent random basis of $(\mathbb{C}^2)^{\otimes n}$.

3. The output state of a T-query quantum algorithm in this model, just before the final measurement, can be expressed as

$$U_{T+1}(O_{\mathcal{S}} \otimes I) \cdots (O_{\mathcal{S}} \otimes I)U_2(O_{\mathcal{S}} \otimes I)U_1|\psi_J\rangle^{\otimes k}|0^m\rangle,$$

   where $m = \text{poly}(n)$ is the number of ancillas. The unitaries U_i's are arbitrary fixed unitaries that do not depend on J or $\mathcal{S}$.

We now formalize our learning tasks in the black-box setting.

Quantum learning task. The algorithm succeeds if it outputs an index $i \in [N]$ such that $|\langle\psi_i|\psi_J\rangle|^2$ is non-negligible in n.

Cloning task. The algorithm succeeds if it outputs a state $|\xi\rangle$ on $n(k+1)$ qubits, such that $|\langle\xi|\psi^{\otimes k}\rangle|^2$ is non-negligible in n.

To motivate the above black-box setting, we note that the Haar random state models the output state of the random circuit and the random index J above is the analog of the random circuit C. Furthermore, the inefficient shadow tomography based algorithm to learn the circuit in the white-box setting (see Section 2) has an analog in the black-box model as well. We only give a sketch of the algorithm which solves the state learning task: first take an epsilon net $\mathcal{N} = \{|\phi_1\rangle, |\phi_2\rangle, \cdots, |\phi_M\rangle\}$ of the complex unit sphere in $N = 2^n$ dimensions where $\epsilon = 1/\text{poly}(n)$ and $M = (C/\varepsilon)^N$ for some universal constant C. Then, using the classical shadows algorithm of [HKP20] with the observables $\{|\phi_i\rangle\langle\phi_i|\}_{i=1}^M$, learn an index r such that the overlap $|\langle\phi_r|\psi_J\rangle| \geq 1/\text{poly}(n)$. Note that only $k = \text{poly}(n)$ samples of the input state $|\psi_J\rangle$ are needed for this. Finally, by querying the oracle $O_{\mathcal{S}}$ exponentially many times and measuring the overlap with the state $|\phi_r\rangle$, learn the index J. After learning the index J, one can solve the cloning task with non-negligible probability as well. Note that this algorithm can even be implemented *non-adaptively*.

There are several more modifications one could make to the above black-box model to make it more in line with the white-box model. For example, the choice of the parameter N above could be changed to further model the fact that there are many more than 2^n circuits acting on n qubit states. Or one could plant a small number of states that are correlated with each of the Haar random states to model the fact that the output states of circuits may have some non-trivial overlaps with a

few other ones. The results proven below are robust to such changes, at least for the examples given above. We have mostly opted for the choices made here for the purposes of presenting a cleaner analysis.

Theorem 1.5 (Black-box lower bounds for cloning)**.** *There exists a state preparation oracle $\mathcal{O}$ such that all T-query quantum query algorithms getting k copies of $|\psi_J\rangle$ for a uniformly random index $J \in [2^n]$ satisfy*

$$\mathrm{F}(\rho, |\psi_J\rangle\langle\psi_J|^{\otimes k+1}) \leq 2^{-n/4}(2T + k + 1)$$

where ρ is the output of the query algorithm and $\mathrm{F}(\cdot)$ denotes the fidelity function.

To prove the above result, it suffices to look at the cloning task, since a query-efficient algorithm for the state learning task can be used to obtain a query-efficient algorithm for the former.

3.1 Proof of Theorem 1.5

The proof proceeds in two parts: the first part follows a hybrid argument similar to [BBBV97] — we show that up to a small error, the queries to the oracle can be replaced by queries to a different oracle that does not depend on $|\psi_J\rangle$. The second part then argues the base case for the algorithm which makes no queries and only uses the given copies of the input state. For the cloning lower bound, we appeal to the well-known result about the optimal cloning probability of a Haar random state [Wer98].

Hybrid argument. Let $N = 2^n$. A T-query algorithm starts in the initial state

$$|\phi^{(0)}\rangle = |\psi_J\rangle^{\otimes k} \otimes |0 \cdots 0\rangle,$$

and the state after $t \in [T]$ queries is given by

$$|\phi^{(t)}\rangle = (O_{\mathcal{S}} \otimes I)U_{t-1} \cdots (O_{\mathcal{S}} \otimes I)U_1(O_{\mathcal{S}} \otimes I)U_0|\phi^{(0)}\rangle,$$

where $U_1, \ldots, U_{T+1}$ are fixed unitaries.

We will show that calls to the state preparation oracle $O_{\mathcal{S}}$ can be replaced with calls to another oracle $O'_{\mathcal{S}}$. Towards this end, we consider the following oracle $O'_{\mathcal{S}}$:

$$\begin{aligned} O'_{\mathcal{S}}|i\rangle|0\rangle &= |i\rangle|\psi_i\rangle, \text{ for all } i \neq J \\ O'_{\mathcal{S}}|J\rangle|0\rangle &= |i\rangle|\psi'\rangle, \end{aligned}$$

where $|\psi'\rangle$ is a Haar random state sampled independently of J and $\mathcal{S}$. Note that all the random variables $J, \mathcal{S} = \{|\psi_i\rangle\}_{i=1}^N$ and $|\psi'\rangle$ are independent and the oracle $O'_{\mathcal{S}}$ above can be implemented unitarily as before, by extending $|\psi'\rangle$ to a random basis independent of $\{|\psi_i\rangle\}_{i=1}^N$ and J.

Defining

$$|\phi'^{(t-1)}\rangle = (O'_{\mathcal{S}} \otimes I)U_{t-1} \cdots (O'_{\mathcal{S}} \otimes I)U_1(O'_{\mathcal{S}} \otimes I)U_0|\phi^{(0)}\rangle,$$

we will show that on average over the choice of $J, \mathcal{S}, |\phi'\rangle$, the following holds

$$\mathbb{E}\left[\left\||\phi^{(t-1)}\rangle - |\phi'^{(t-1)}\rangle\right\|\right] \leq \epsilon_t \text{ where } \epsilon_t = \frac{2t}{\sqrt{N}}. \tag{3.1}$$

The statement is trivially true when no queries are made, so consider any $t \in [T]$. Then, using the triangle inequality,

$$\mathbb{E}\left[\left\||\phi^{(t)}\rangle - |\phi'^{(t)}\rangle\right\|\right] \leq \mathbb{E}\left[\left\||\phi^{(t)}\rangle - (O_{\mathcal{S}} \otimes I)U_t|\phi'^{(t-1)}\rangle\right\|\right] + \mathbb{E}\left[\left\|(O_{\mathcal{S}} \otimes I)U_t|\phi'^{(t-1)}\rangle - |\phi'^{(t)}\rangle\right\|\right], \tag{3.2}$$

Note that $|\phi^{(t)}\rangle = (O_{\mathcal{S}} \otimes I)U_t|\phi^{(t-1)}\rangle$, thus by the unitary invariance of the Euclidean norm, the induction hypothesis implies that the first term is at most ε_{t-1}. We now bound the second term

$$\mathbb{E}\left[\left\|(O_{\mathcal{S}} \otimes I)U_t|\phi'^{(t-1)}\rangle - |\phi'^{(t)}\rangle\right\|\right] = \mathbb{E}\left[\left\|(O_{\mathcal{S}} \otimes I)U_t|\phi'^{(t-1)}\rangle - (O'_{\mathcal{S}} \otimes I)U_t|\phi'^{(t-1)}\rangle\right\|\right].$$

Let us write

$$|\xi^{(t-1)}\rangle := U_t|\phi'^{(t-1)}\rangle = \sum_{i=1}^{N} |i\rangle|\xi'^{(t-1)}_i\rangle,$$

for some sub-normalized states $|\xi'^{(t-1)}_i\rangle$ satisfying $\sum_{i=1}^{N} \||\xi'^{(t-1)}_i\rangle\|^2 = 1$. Note that

$$\left\|(O_{\mathcal{S}} \otimes I)U_t|\phi'^{(t-1)}\rangle - (O'_{\mathcal{S}} \otimes I)U_t|\phi'^{(t-1)}\rangle\right\| \leq 2\left\||\xi'^{(t-1)}_J\rangle\right\|,$$

since $O_{\mathcal{S}} \otimes I$ and $O'_{\mathcal{S}} \otimes I$ act the same on all states of the form $|i\rangle|\phi\rangle$ where $i \neq J$.

Thus, plugging this in (3.2),

$$\mathbb{E}\left[\left\||\phi^{(t)}\rangle - |\phi'^{(t)}\rangle\right\|\right] \leq \varepsilon_{t-1} + 2 \cdot \mathbb{E}\left[\left\||\xi'_J{}^{(t-1)}\rangle\right\|\right].$$

Note that J was sampled uniformly from $[N]$ and independently of $\mathcal{S}$ and $|\psi'\rangle$ which are i.i.d. Haar random states. By symmetry, it follows that J is uniformly distributed in $[N]$ conditioned on $O'_{\mathcal{S}}$ and $|\psi'\rangle$. Since the state $|\xi'^{(t-1)}\rangle$ is determined by the choice $O'_{\mathcal{S}}$ and $|\psi'\rangle$, it follows that

$$\begin{aligned}
\mathbb{E}\left[\left\||\phi^{(t)}\rangle - |\phi'^{(t)}\rangle\right\|\right] &\leq \varepsilon_{t-1} + \frac{2}{N} \cdot \mathbb{E}\left[\sum_{i=1}^{N}\left\||\xi'^{(t-1)}_i\rangle\right\|\right] \\
&\leq \varepsilon_{t-1} + \frac{2}{N} \cdot \sqrt{N} \cdot \mathbb{E}\left[\sqrt{\sum_{i=1}^{N}\left\||\xi'^{(t-1)}_i\rangle\right\|^2}\right] \\
&\leq \varepsilon_{t-1} + \frac{2}{\sqrt{N}} = \frac{2t}{\sqrt{N}} = \varepsilon_t.
\end{aligned}$$

Note that the expectation on the right hand side is only taken over the choice of the oracle $O'_{\mathcal{S}}$ and $|\psi'\rangle$.

Recalling $N = 2^n$, it follows by Markov's inequality that with probability at least $1 - 2^{-n/4}$, the states after T queries are $2T \cdot 2^{-n/4}$-close in the Euclidean norm after replacing the oracle $O_{\mathcal{S}}$ with $O'_{\mathcal{S}}$.

Base case. Now we consider algorithms for the cloning task that receive k copies of the input state $|\psi_J\rangle$ and query the oracle $O'_{\mathcal{S}}$. Since $|\psi_J\rangle$ is independent of $O'_{\mathcal{S}}$ and the index J, observe that any such quantum algorithm defines a cloning channel for a Haar random state, i.e., a quantum channel that takes k copies of an n-qubit Haar random state and outputs a mixed state ρ on $(k+1)n$ qubits that should be close to $(k+1)$ copies of the input state. By the well-known results about the optimal cloning of Haar random states [Wer98], it follows that

$$\mathrm{F}(\rho, |\psi_J\rangle\langle\psi_J|^{\otimes(k+1)}) \leq \frac{\binom{2^n+k-1}{k}}{\binom{2^n+k}{k+1}} \leq \frac{k+1}{2^n+k} \leq k \cdot 2^{-n}.$$

Combining the hybrid argument and the base case, it follows that for the cloning task, the success probability is at most $2T \cdot 2^{-n/4} + 2^{-n/4} + k \cdot 2^{-n} \leq 2^{-n/4}(2T + k + 1)$. This completes the proof of Theorem 1.5.

4 Cryptography from hardness of quantum learning

In this section, we lay out the cryptographic objects we can construct from our hardness of learning conjectures.

4.1 One-way state generators

The conjecture about the hardness of learning is essentially *equivalent* to the existence of one-way state generators (OWSGs). These are a quantum analogue of one-way functions, which are functions efficiently computable in the forwards direction but computationally difficult to invert. We first recall the formal definition of a OWSG, a primitive first introduced by Morimae and Yamakawa [MY22b, MY22a].

Definition 4.1 (One-way state generator)**.** A *one-way state generator* G is a pair of QPT algorithms $(\mathrm{Gen}, \mathrm{Ver})$ such that there exists polynomials $r(n), m(n)$ such that

- Gen (called the *generator*) takes as input the security parameter 1^n in unary, a string $k \in \{0,1\}^{r(n)}$ called a *key*, and outputs a $m(n)$-qubit pure quantum state $|\psi_k\rangle$.
- Ver (called the *verification*) takes as input the security parameter 1^n in unary, a key $k \in \{0,1\}^{r(n)}$ and an $m(n)$-qubit state $|\psi\rangle$, and accepts or rejects.

We say that a OWSG G satisfies *correctness* if for all security parameters n, for all keys $k \in \{0,1\}^{r(n)}$,

$$\mathbb{P}\left[\mathrm{Ver}(1^n, k, \mathrm{Gen}(1^n, k)) \text{ accepts}\right] \geq 1 - \mathrm{negl}(n) \ .$$

We say that G *has security error* $\gamma(n)$ if for all polynomials $q(n)$, for all QPT algorithms A, for all sufficiently large n,

$$\mathbb{P}\left[\mathrm{Ver}(1^n, k', |\psi_k\rangle) \text{ accepts} : \begin{array}{c} k \leftarrow \{0,1\}^{r(n)} \\ |\psi_k\rangle \leftarrow \mathrm{Gen}(1^n, k) \\ k' \leftarrow A(|\psi_k\rangle^{\otimes q(n)}) \end{array}\right] \leq \gamma(n) \ .$$

We say that G is *cryptographically secure* if it has security error that is negligible in n (i.e., it goes to zero faster than $1/\operatorname{poly}(n)$).

Remark 4.2. For simplicity we often omit the security parameter 1^n as an input to Gen, Ver when it is clear from context.

Let $r(n)$ be a polynomial such that $r(n)$ bits are sufficient to describe a circuit from the family $\mathscr{C}_n$. We define a OWSG $G = (\text{Gen}, \text{Ver})$ based on random circuits.

Protocol 4.3. Random circuit OWSG

Gen: Given input key $C \in \{0,1\}^{r(n)}$, interpret it as a description of an n-qubit circuit C from the ensemble $\mathscr{C}_n$ (defined in Section 2). Output $|C\rangle = C|0^n\rangle$.

Ver: Given input $C \in \{0,1\}^{r(n)}$ and a state $|D\rangle$ on n qubits, apply $C^\dagger$ to the state, and measure. Accept if the result is all zeroes, and reject otherwise.

It is clear that the algorithms Gen, Ver run in polynomial time. It is also easy to see that G has perfect correctness. We now argue that the security of the OWSG is essentially *equivalent* to Computational No-Learning Assumption (Conjecture 1.1). The level of security corresponds to the strength of the hardness conjecture.

Lemma 4.4 (Equivalence between hardness of learning and random circuit OWSG security)**.** *Assuming ε-No-Learning (Conjecture 1.1), the random circuit OWSG has security error $(2-\varepsilon)\varepsilon$. Conversely, if the random circuit OWSG has security error γ, then the $\sqrt{\gamma}$-No Learning Assumption holds.*

Proof. To prove (weak) security, assume for contradiction there exists a QPT adversary A and a polynomial $q(n)$ such that for infinitely many n,

$$\mathbb{P}\left[\text{Ver}(D,|C\rangle) \text{ accepts} : \begin{array}{c} C \leftarrow \mathscr{C}_n \\ D \leftarrow A(|C\rangle^{\otimes q(n)}) \end{array}\right] > (2-\varepsilon)\varepsilon$$

The acceptance probability of the verification circuit can be written as:

$$\mathop{\mathbb{E}}_{D \leftarrow A(|C\rangle^{\otimes q(n)})} |\langle C|D\rangle|^2 > (2-\varepsilon)\varepsilon \ .$$

This implies that the probability over the choice of D output by $A(|C\rangle^{\otimes q(n)})$ that $|\langle C|D\rangle|^2 \geq \varepsilon$ is greater than ε for infinitely many n. This contradicts Conjecture 1.1.

Conversely, assume for contradiction that the random circuit OWSG has security error γ, but there exists a QPT algorithm A such that

$$\mathbb{P}\left[|\langle C|D\rangle|^2 \geq \sqrt{\gamma} : \begin{array}{c} C \leftarrow \mathscr{C}_n \\ D \leftarrow A(|C\rangle^{\otimes \text{poly}(n)}) \end{array}\right] > \sqrt{\gamma} \ .$$

This implies that

$$\mathbb{P}\left[\text{Ver}(D,|C\rangle) \text{ accepts} : \begin{array}{c} C \leftarrow \mathscr{C}_n \\ D \leftarrow A(|C\rangle^{\otimes q(n)}) \end{array}\right] > \gamma$$

which contradicts the security of the OWSG.

□

When $\varepsilon \geq 1/\operatorname{poly}(n)$, then we consider the resulting OWSG from Lemma 4.4 to have *weak* security, as it implies that a QPT adversary could potentially produce a non-negligible approximation of $|C\rangle$ with non-negligible probability. To obtain a *cryptographically secure* OWSG, we need to amplify it so that all QPT adversaries can only succeed with at most negligible probability.

A standard way to amplify the security of a weak OWSG is via *parallel repetition*: now the key to the amplified OWSG corresponds to a t-tuple of independently chosen random circuits $(C_1, \ldots, C_t)$, and the output is the tensor product of the corresponding states $|C_1\rangle \otimes \cdots \otimes |C_t\rangle$. Verification proceeds by checking that the i'th block of n qubits is in the state $|C_i\rangle$ for each $i = 1, \ldots, n$. Intuitively, if it is somewhat hard to learn the output of one random quantum circuit, then it should be *very* hard to simultaneously learn the output of *many* random quantum circuits. This intuition holds true and the parallel repetition of OWSGs is formally analyzed in Morimae and Yamakawa [MY22a], who showed that if a OWSG G has security error γ, then the t-fold repetition of G, denoted by G^t, has security error $\approx \gamma^t$ (up to additive errors that are negligible in n). (This is also implied by the general quantum hardness amplification result of [BQSY24]).

Theorem 4.5. *Assuming ε-No-Learning for $\varepsilon \leq 1 - \frac{1}{\operatorname{poly}(n)}$, there exists a cryptographically secure one-way state generator.*

Proof. Assuming ε-No-Learning, by Lemma 4.4 there exists a weak OWSG G with security error $(2-\varepsilon)\varepsilon$. Let

$$t = \frac{\log^2(n)}{\log \frac{1}{(2-\varepsilon)\varepsilon}} \ .$$

When $\varepsilon \leq 1 - 1/\operatorname{poly}(n)$, the quantity t is at most $\operatorname{poly}(n)$. Consider the following OWSG $\hat{G}$, which is simply the original OWSG G repeated t times in parallel.

Protocol 4.6. (Strong) one-way state generator

Gen: Given input key $(C_1, \ldots, C_t) \in (\{0,1\}^{r(n)})^t$, interpret it as a description of a t-tuple of n-qubit circuits from the ensemble $\mathscr{C}_n$. Output $|C_1\rangle \otimes \cdots \otimes |C_t\rangle$.

Ver: Given input $(C_1, \ldots, C_t) \in (\{0,1\}^{r(n)})^t$ and a state $|D\rangle$ on nt qubits, apply $C_1^\dagger \otimes \cdots \otimes C_t^\dagger$ to the state, and measure. Accept if the result is all zeroes, and reject otherwise.

By the result on hardness amplification of OWSGs by Morimae and Yamakawa [MY22a] (alternatively, by the quantum parallel repetition theorem of [BQSY24]), the security error of $\hat{G}$ is at most

$$\Big((2-\varepsilon)\varepsilon\Big)^t + \operatorname{negl}(n) = 2^{-\log^2(n)} + \operatorname{negl}(n) \ .$$

Note that $2^{-\log^2(n)}$ goes to 0 faster than any inverse polynomial $1/\operatorname{poly}(n)$, and thus $\hat{G}$ has negligible security error.

□

4.2 Quantum bit commitments

In this section we explore the cryptographic implications of our second hardness assumption, the Computational No-Cloning Assumption (Conjecture 1.2). We show that No-Cloning directly implies the existence of secure quantum bit commitments.

We formally define quantum bit commitment schemes. In this paper we only define a special kind known as *noninteractive quantum commitments*; while quantum commitment schemes can be interactive in general, it was shown by [Yan22] that in the quantum setting they can always be generically compiled to a simple noninteractive protocol[8].

Definition 4.7 (Noninteractive quantum bit commitment)**.** A *noninteractive quantum bit commitment scheme* (or a *commitment scheme* for short) Com is a QPT algorithm that takes as input a security parameter 1^n and a bit $b \in \{0,1\}$, and behaves as follows: it applies a unitary $U_{n,b}$ to the all zeroes state, obtaining a bipartite pure state $|\psi_b\rangle$ on registers AB.

We say that a commitment scheme Com satisfies *correctness* if for all security parameters n, for all $b \in \{0,1\}$,

$$|\langle\psi_0|\psi_1\rangle|^2 \leq \mathrm{negl}(n)$$

where $|C_b\rangle$ is the output of $\mathrm{Com}(1^n, b)$. We say that Com satisfies ε-*statistical hiding* if

$$\mathrm{F}(\rho_0, \rho_1) \geq 1 - \varepsilon(n)$$

where $\mathrm{F}(\cdot,\cdot)$ is the fidelity between two density matrices, and ρ_b is the reduced density matrix of $|\psi_b\rangle$ on register B. We say that Com satisfies δ-*computational binding* if for all QPT adversaries A, for all sufficiently large n,

$$\mathrm{F}(|\psi_0\rangle\langle\psi_0|, (A_n \otimes \mathbb{I})(|\psi_1\rangle\langle\psi_1|)) \leq \delta(n)$$

where A_n denotes running A with security parameter 1^n and it takes as input register A of $|\psi_1\rangle$.

For the remainder of this section, we will simply refer to noninteractive commitment schemes as simply a commitment scheme. We only defined the notion of commitments with statistical hiding and computational binding; there is also the other "flavor" of commitments such as computational hiding and statistical binding. These flavors can be efficiently switched in a blackbox way; see [Yan22, HMY23] for a proof.

We note that at least one of the hiding or binding properties must rely on computational hardness assumptions [BCMS97]; in other words, there do not exist quantum commitment schemes that are both statistically hiding as well as statistically binding. For an exploration of the complexity-theoretic underpinnings of the security of quantum commitment schemes, see Bostanci, et al. [BEM+23].

We now define a commitment scheme based on random circuits. Recall that $\mathscr{C}_n$ is the ensemble of n-qubit quantum circuits as defined in Section 2. To distinguish between a classical description of a circuit and the state generated by the circuit, we use the following notation: $|\hat{C}\rangle$ denotes the standard basis state of $r(n) = \log|\mathscr{C}_n|$ qubits that represents the classical description of the circuit C, and $|C\rangle = C|0^n\rangle$ denotes the state output by the circuit on the all zeroes input. In what follows, we set $k = \mathcal{O}(\varepsilon^{-2}\log|\mathscr{C}_n|/\varepsilon) = \mathcal{O}(n\log^2 n)$ for some $\varepsilon = 1/\operatorname{poly}(n)$ to be determined later.

[8] It is worth noting that this transformation is not generically possible in the classical setting!

Protocol 4.8. Commitment scheme based on random circuits

$\mathrm{Com}(1^n, b)$: If $b = 0$, then prepare the state

$$|\psi_0\rangle_{\mathrm{AB}} := \frac{1}{\sqrt{|\mathscr{C}_n|}} \sum_{C \in \mathscr{C}_n} \left(|C\rangle^{\otimes k} \otimes |0^n\rangle \right)_{\mathrm{A}} \otimes |\hat{C}\rangle_{\mathrm{B}} \ .$$

If $b = 1$, then prepare the state

$$|\psi_1\rangle_{\mathrm{AB}} := \frac{1}{\sqrt{|\mathscr{C}_n|}} \sum_{C \in \mathscr{C}_n} |C\rangle_{\mathrm{A}}^{\otimes (k+1)} \otimes |\hat{C}\rangle_{\mathrm{B}} \ .$$

It should be clear that for each $b \in \{0, 1\}$ there is an efficiently computable unitary $U_{n,b}$ that prepares $|\psi_b\rangle$ from the all zeroes state. Therefore Com is a QPT algorithm.

We argue the correctness property of Com.

Claim 4.9. *The commitment scheme of Protocol 4.8 satisfies correctness.*

Proof. We evaluate the overlap between $|\psi_0\rangle, |\psi_1\rangle$:

$$\begin{aligned}
|\langle\psi_0|\psi_1\rangle|^2 &= \left| \frac{1}{|\mathscr{C}_n|} \sum_{C \in \mathscr{C}_n} \langle C|0^n\rangle \right|^2 \\
&\leq \left(\frac{1}{|\mathscr{C}_n|} \sum_{C \in \mathscr{C}_n} |\langle C|0\rangle| \right)^2 \\
&\leq \frac{1}{|\mathscr{C}_n|} \sum_{C \in \mathscr{C}_n} |\langle C|0\rangle|^2 \\
&= \mathrm{Tr}\left(\frac{1}{|\mathscr{C}_n|} \sum_{C \in \mathscr{C}_n} |C\rangle\langle C| \cdot |0^n\rangle\langle 0^n| \right) \\
&= \frac{1}{2^n}
\end{aligned}$$

where we used the property that the uniform distribution over $\mathscr{C}_n$ forms a 1-design; here we use the fact that the gate set includes all Cliffords and therefore all Pauli operators. The uniform distribution over $\mathscr{C}_n$ is then invariant under applying a layer of random Pauli operators on each qubit at the end of each circuit. □

We now analyze the statistical hiding property of the commitment scheme.

Claim 4.10. *The commitment scheme of Protocol 4.8 satisfies 4ε-statistical hiding.*

Proof. Let ρ_0, ρ_1 denote the reduced density matrices of $|\psi_0\rangle, |\psi_1\rangle$ on register B. To bound the fidelity between ρ_0, ρ_1, we use Uhlmann's theorem [Uhl76], which implies that for all unitary operators U acting on register A and an ancilla register E (consisting of r qubits), we have

$$\mathrm{F}(\rho_0, \rho_1) \geq |\langle 0^r, \psi_0|_{\mathrm{EAB}} \left(U_{\mathrm{EA}} \otimes \mathbb{I}_{\mathrm{B}} \right) |0^r, \psi_1\rangle_{\mathrm{EAB}}|^2 \ .$$

Here, $|0^r, \psi_b\rangle\rangle$ is simply shorthand for pre-pending r ancilla qubits to the state $|\psi_b\rangle$.

To give a lower bound it suffices to describe *some* local unitary operator U that, with some ancillas, maps $|\psi_0\rangle$ to have large overlap with $|\psi_1\rangle$. At a high level, the unitary U will coherently run the classical shadows protocol of Huang, Kueng, and Preskill [HKP20] as described in Section 2 on the k copies of $|C\rangle$ in order to obtain a classical description of a circuit D such that $|D\rangle \approx |C\rangle$. Controlled on this description, a copy of $|D\rangle$ is synthesized. Since $|D\rangle$ is close to $|C\rangle$ with very high probability, the intermediate work of the classical shadows protocol can be uncomputed with high fidelity.

In more detail, the unitary U behaves as follows. Given input $|C\rangle^{\otimes k}$ (plus some ancillas), the circuit-learning algorithm from Corollary 2.3 can be purified to be a unitary operator V that is run coherently to yield

$$V(|0^r\rangle \otimes |C\rangle^{\otimes(k-1)}) \otimes |C\rangle \otimes |0^n\rangle = \sum_{D\in\mathscr{C}_n} \sqrt{p_{C,D}}|\vartheta_{C,D}\rangle \otimes |\hat{D}\rangle \otimes |C\rangle \otimes |0^n\rangle$$

where $p_{C,D}$ is the probability that running the algorithm V on $|C\rangle^{\otimes(k-1)}$ outputs D, the state $|\hat{D}\rangle$ is the classical description of D, and the state $|\vartheta_{C,D}\rangle$ is the post-measurement state after measuring $\hat{D}$. Controlled on $|\hat{D}\rangle$, the inverse $D^\dagger$ can be applied to the remaining copy of $|C\rangle$, and a copy of $|D\rangle$ can be synthesized to yield

$$\sum_{D\in\mathscr{C}_n} \sqrt{p_{C,D}}|\vartheta_{C,D}\rangle \otimes |\hat{D}\rangle \otimes D^\dagger|C\rangle \otimes |D\rangle \ .$$

The last two n-qubit registers are swapped to yield:

$$\sum_{D\in\mathscr{C}_n} \sqrt{p_{C,D}}|\vartheta_{C,D}\rangle \otimes |\hat{D}\rangle \otimes |D\rangle \otimes D^\dagger|C\rangle \ .$$

Finally, applying the inverse $V^\dagger$ we get

$$\sum_{D\in\mathscr{C}_n} \sqrt{p_{C,D}}V^\dagger(|\vartheta_{C,D}\rangle \otimes |\hat{D}\rangle) \otimes |D\rangle \otimes D^\dagger|C\rangle \ .$$

This concludes the description of the unitary U. We can now compute the inner product between $U|\psi_0, 0\cdots0\rangle$ and $|\psi_1, 0\cdots0\rangle$. We rearrange the ancillas for notational convenience:

$$\begin{aligned}
&\langle 0^r, \psi_1, 0^n| \Big(\frac{1}{\sqrt{|\mathscr{C}_n|}} \sum_{C,D\in\mathscr{C}_n} \sqrt{p_{C,D}}V^\dagger(|\vartheta_{C,D}\rangle \otimes |\hat{D}\rangle) \otimes |D\rangle \otimes D^\dagger|C\rangle \Big) \otimes |\hat{C}\rangle \\
&= \frac{1}{|\mathscr{C}_n|} \sum_{C,D\in\mathscr{C}_n} \sqrt{p_{C,D}}(\langle 0^r| \otimes \langle C|^{\otimes k})V^\dagger(|\vartheta_{C,D}\rangle \otimes |\hat{D}\rangle) \langle C|D\rangle \langle 0^n|D^\dagger|C\rangle \\
&= \frac{1}{|\mathscr{C}_n|} \sum_{C,D\in\mathscr{C}_n} p_{C,D} \, |\langle C|D\rangle|^2 \ .
\end{aligned}$$

Note that $\sum_{D\in\mathscr{C}_n} p_{C,D}|\langle C|D\rangle|^2$ is the expected overlap between the state of the circuit D output by the classical shadows protocol and the state $|C\rangle$; by Corollary 2.3 this is at least $(1-\varepsilon)^2$. Thus the fidelity between ρ_0, ρ_1 is at least $(1-\varepsilon)^4 \geq 1-4\varepsilon$. This concludes the proof of Claim 4.10.

□

Finally, we prove that the commitment scheme is secure under the No-Cloning Assumption.

Claim 4.11. *The δ-No-Cloning Assumption (Conjecture 1.2) implies that the commitment scheme from Protocol 4.8 satisfies $(2-\delta)\delta$-computational binding.*

Proof. Suppose there was a QPT adversary A that for infinitely many $n \in \mathbb{N}$,

$$\mathrm{F}(|\psi_0\rangle\langle\psi_0|, (A_n \otimes \mathbb{I})(|\psi_1\rangle\langle\psi_1|)) > 2\delta(n) \ .$$

Purifying the algorithm A to include ancillas, there exists a unitary V and a pure state $|\vartheta\rangle$ such that

$$|\langle\vartheta, \psi_0|(V \otimes \mathbb{I})|0^r, \psi_1\rangle|^2 \geq (2-\delta(n))\delta(n) \ .$$

This however means that

$$\frac{1}{|\mathscr{C}_n|}\sum_{C\in\mathscr{C}_n}\left|(\langle\vartheta| \otimes \langle C|^{\otimes(k+1)})V(|0^r\rangle \otimes |C\rangle^{\otimes k} \otimes |0^n\rangle)\right|^2 > (2-\delta(n))\delta(n)$$

by Jensen's inequality. This implies that the algorithm A satisfies, for infinitely many n,

$$\mathbb{P}\left[|\langle C|^{\otimes(k+1)}|\phi\rangle|^2 \geq \delta(n) : \begin{array}{c} C \leftarrow \mathscr{C}_n \\ |\phi\rangle \leftarrow A(|C\rangle^{\otimes k}) \end{array}\right] > \delta(n)$$

which contradicts the No-Cloning Assumption. □

A converse? Does the computational binding security of Com *imply* the No-Cloning Assumption? Intuitively, an efficient algorithm to clone outputs of random circuits (i.e., break the No-Cloning Assumption) should be able to break the binding property of the commitment scheme Com. However, this requires that the cloning algorithm can be run *coherently*, without ancillary junk states that are entangled with the the underlying circuit C – this is because breaking binding requires coherently mapping $|C\rangle^{\otimes k}$ to as close to $|C\rangle^{\otimes(k+1)}$ as possible. This is not *a priori* guaranteed by the fact that the No-Cloning Assumption was broken.

We summarize the conclusions of Claims 4.9, 4.10, and 4.11 below.

Lemma 4.12. *The commitment scheme of Protocol 4.8 satisfies correctness, 4ε-statistical hiding, and (assuming δ-No-Cloning) $(2-\delta)\delta$-computational binding.*

One could ask whether the security guarantees of this commitment scheme could be improved. Just like how we were able to amplify a weakly-secure OWSG to a cryptographically secure OWSG via parallel repetition in Section 4.1, we would like to amplify the commitment scheme of Protocol 4.8 so that it has negligible error for both the hiding and binding properties.

Amplification of bit commitments is more subtle than with OWSGs, however, because there are two different security properties to handle. In general, trying to amplify one security property comes at the cost of degrading the other security property. For example, it was only recently shown by Bostanci, Qian, Spooner, and Yuen [BQSY24] that the t-fold parallel repetition of a δ-computational binding quantum commitment Com yields a new commitment scheme Com^t with roughly δ^t-computational binding. However, if the original commitment Com satisfied ε-statistical hiding, then Com^t satisfies $t\varepsilon$ hiding – it is now *easier* for the receiver to distinguish between commitments to 0 and 1 (because now it has t chances to do so). This would be fine if ε were a negligible quantity to begin with (i.e., the base commitment scheme had negligible hiding error),

but in our case ε is at best an inverse polynomial quantity (this is due to the sample complexity of the classical shadows protocol).

To perform the amplification, we take advantage of the "flavor switching" transformation for quantum bit commitments, which allows us to take a quantum commitment of one flavor (e.g., statistical hiding, computational binding), and generically obtain a quantum commitment of the other flavor (e.g., computational hiding, statistical binding) with only a small loss in parameters.

Lemma 4.13 (Flavor switching)**.** *If* Com *is a ε-statistical (resp. computational) hiding and δ-computational (resp. statistical) binding quantum commitment, then there exists a commitment* Com$'$ *that satisfies $\sqrt{\delta}$-computational (resp. statistical) hiding and ε-statistical (resp. computational) binding.*

Proof. This is proved in [HMY23, Theorem 7]. □

With this in hand we obtain the following amplification result:

Lemma 4.14. *Assuming δ-No-Cloning for some $\delta(n) \leq 1-1/p(n)$ for some polynomial p, there exists a cryptographically-secure quantum commitment scheme satisfying correctness,* negl(n)*-statistical hiding and* negl(n)*-computational binding.*

Proof. First, we choose $\varepsilon(n) = (2np(n))^{-2}$ where p satisfies $\delta(n) \leq 1 - \frac{1}{p(n)}$. Then assuming δ-No-Cloning, Lemma 4.12 implies that the commitment scheme Com of Protocol 4.8 (with the $k(n)$ parameter chosen as a function of $\varepsilon(n)$) has 4ε-statistical hiding and $(2-\delta)\delta$-computational binding.

Let Com$'$ denote the $np(n)^2$-fold parallel repetition of the commitment Com, which means that the committer and receiver run $np(n)^2$ parallel independent instances of Com. The parallel repetition theorem of Bostanci, Qian, Spooner, and Yuen [BQSY24] implies that Com$'$ has $\frac{1}{n}$-statistical hiding and the computational binding security error is at most

$$((2-\delta(n))\delta(n))^{np(n)^2} + \mathrm{negl}(n) \leq \left(1 - \frac{1}{p(n)^2}\right)^{np(n)^2} + \mathrm{negl}(n) \leq e^{-\Omega(n)} + \mathrm{negl}(n) = \mathrm{negl}(n) \ .$$

Switching flavors (using Lemma 4.13), we get a commitment Com$''$ with negl(n)-computational hiding and $\frac{1}{n}$-statistical binding. We then perform parallel repetition once again, repeating the commitment Com$''$ for n times in parallel to obtain a commitment Com$'''$ where the computational hiding security error is at most $n \cdot \mathrm{negl}(n) = \mathrm{negl}(n)$ and the statistical binding security error is at most $n^{-n} = \mathrm{negl}(n)$. The computational hiding bound is argued via a hybrid argument: if there were an efficient algorithm A that could distinguish between $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$ with non-negligible advantage α where ρ_0, ρ_1 are the reduced density matrices seen by the receiver in Com$''$, then via the triangle inequality there exists a $j \in [n]$ such that A can distinguish between $\rho_0^{\otimes j} \otimes \rho_0 \otimes \rho_1^{\otimes(n-j-1)}$ and $\rho_0^{\otimes j} \otimes \rho_1 \otimes \rho_1^{\otimes(n-j-1)}$ with advantage at least α/n, which contradicts the negligible hiding security error of Com$''$.

The statistical binding bound is argued by viewing the binding security game of a commitment scheme as a 2-message interactive protocol, and using the fact that parallel repetition reduces the soundness error of such protocols at an exponential rate [KW00, Theorem 6].

Finally, we can switch flavors once more to obtain the final commitment Com$''''$, which satisfy negl(n)-statistical hiding and negl(n)-statistical binding. □

5 NISQ-friendly quantum cryptography

5.1 NISQ-friendly one-way state generators

We take a step further and ask whether it is possible to obtain a OWSG that is both cryptographically secure and *NISQ-friendly*. NISQ-friendliness of a quantum cryptographic primitive generally means that the correctness property of the primitive still holds even when the quantum algorithms in the primitive (e.g., the key generation or verification algorithms) suffer from noise. We define this notion for a OWSG.

Definition 5.1 (Noise-robust OWSG)**.** Let $\mathcal{N}$ denote a noise model for quantum computers. We say that a OWSG $G = (\mathrm{Gen}, \mathrm{Ver})$ is $\eta(n)$*-robust against noise model* $\mathcal{N}$ if for all security parameters n, for all keys $k \in \{0,1\}^{r(n)}$,

$$\mathbb{P}\left[\widetilde{\mathrm{Ver}}(1^n, k, \widetilde{\mathrm{Gen}}(1^n, k)) \text{ accepts}\right] \geq \eta(n)$$

where $\widetilde{\mathrm{Gen}}$ and $\widetilde{\mathrm{Ver}}$ denote the quantum channels corresponding to running the algorithms Gen, Ver on a quantum computer with noise model $\mathcal{N}$.

In realistic devices, noise rapidly degrades the fidelity of the signal. The fidelity between the states produced by the noisy circuit and an ideal circuit could be $1/p(n)$ for some polynomial $p(n)$, which implies $\eta(n)$ is also inverse polynomially small. This gives adversaries a greater leeway to break the security of the construction as generating a state that only has inverse polynomial overlap with the output is enough to pass the verification step. In this work, we show how to amplify the security of the one way state generator even in this case. As discussed in Section 1, there are experimentally realistic depth regimes and noise rates for which inverse polynomial fidelity is reasonable due to the white noise phenomenon—for instance, see [AABea19, MVM+23, DHJB24].

What can we say about the security of the the OWSG in this regime? We have not changed the OWSG construction, so the security guarantee still holds with respect to *any* polynomial-time adversary, including noise-free ones. Suppose we assume ε-No-Learning for $\varepsilon \ll 1/p(n)$. This assumption is consistent with what we know about the complexity of the learning task; as discussed in Section 1.1, we do not know of an efficient algorithm that can, given polynomially-many copies of the state $|C\rangle$, to output a circuit description D that has fidelity any better than $2^{-\Omega(n)}$.

Even though the noisy quantum computer can only verify the outputs of the OWSG with small probability (even when given the key), any efficient adversary – even a noise-free one – has a *much smaller* probability of being able to invert the outputs of a OWSG. We now have an exploitable gap: we can amplify a less noise robust OWSG to have high noise robustness, yet preserve security.

Parallel repetition is no longer a good amplification technique: although the amplified OWSG G^t is secure against polynomial time adversaries, it may not be possible for honest adversaries to successfully run the verification procedure of G^t on a noisy quantum computer: if the success probability of a single verification of G is at η, then the success probability of t parallel verifications is η^t, an exponentially small quantity.

We instead take the *threshold repetition* of $G = (\mathrm{Gen}, \mathrm{Ver})$; this is a OWSG $G^{t,k}$ consisting of t independent copies of G, but instead of verifying that all t copies have been inverted, the verification algorithm checks that at least k out of the t have been inverted. We formally define the threshold repetition OWSG $G^{t,k} = (\mathrm{Gen}^t, \mathrm{Ver}^{t,k})$ next. Let $|\psi_k\rangle$ denote the output of Gen on input k (we omit mention of the security parameter n for convenience).

Protocol 5.2. Threshold one-way state generator $G^{t,k}$

Gen^t: Given input $(k_1, \ldots, k_t) \in (\{0,1\}^{r(n)})^t$, output $|\psi_{k_1}\rangle \otimes \cdots \otimes |\psi_{k_t}\rangle$.

$\mathrm{Ver}^{t,k}$: Given input $(k_1, \ldots, k_t) \in (\{0,1\}^{r(n)})^t$ and a state $|D\rangle$ on nt qubits, for all $i \in [t]$, run the verification procedure Ver on the i'th block of n qubits with input k_i. Accept iff at least k of the individual verifications accept.

We now show that for an appropriate choice of t, k, the threshold repetition $G^{t,k}$ has good noise robustness and also good security. This proof relies on two types of Chernoff bounds. One is the standard one (that the sum of independent random variables concentrates around their mean); this is used to obtain the improved noise robustness. The other is a *computational Chernoff bound*, which argues that if an efficient adversary has at most γ probability of inverting the output of a OWSG, then an efficient adversary has an exponentially small probability of inverting significantly more than γ fraction of t independent instances of the OWSG. In theoretical computer science and cryptography, such a result is also known as a *threshold direct product theorem* (see, e.g., [IK10]).

Lemma 5.3. *Let $\mathcal{N}$ denote a noise model for quantum computers such that independent, parallel computations (i.e., they do not share qubits) experience independent noise. Let $G = (\mathrm{Gen}, \mathrm{Ver})$ be a OWSG that is $\eta(n)$-robust against $\mathcal{N}$ and has security error $\gamma(n)$ such that $\eta(n) - \gamma(n) \geq 1/p(n)$ for some polynomial $p(n)$. Then for all sufficiently large polynomials $t(n)$, the threshold repetition $G^{t,k}$ for $k(n) = \left(\eta(n) - \sqrt{\frac{n}{2t(n)}}\right)t(n)$ is $(1 - \mathcal{O}(2^{-n}))$-robust against $\mathcal{N}$ and has negligible security error.*

Proof. We sometimes omit mention of n for notational clarity, and write $t = t(n), k = k(n), \eta = \eta(n)$, etc.

We upper bound the probability that fewer than k out of t of the verifications fail to accept in $G^{t,k}$. By the assumption on the noise model in the lemma statement, the acceptance of each verification is a Bernoulli random variable X_i with bias at least η. Therefore by the Chernoff-Hoeffding bound,

$$\mathbb{P}\left[X_1 + \cdots + X_t < k\right] = \mathbb{P}\left[X_1 + \cdots + X_t < \eta t - \sqrt{\frac{nt}{2}}\right] \leq 2\exp\left(-\Omega(n)\right) \ .$$

This establishes the noise robustness of $G^{t,k}$. We now argue about its security, using the following computational Chernoff bound:

Lemma 5.4 (Computational Chernoff bound for OWSGs)**.** *Let $\xi(n)$ denote an inverse polynomial, i.e., $\xi(n) = 1/p(n)$ for some polynomial $p(n)$. Let $G = (\mathrm{Gen}, \mathrm{Ver})$ be a OWSG with security error $\gamma(n)$. Then for all sufficiently large polynomials $t(n)$, the threshold OWSG $G^{t,k}$ has negligible security error, where $k(n) = (\gamma(n) + \xi(n))t(n)$.*

We prove Lemma 5.4 in Appendix A. Let $\xi(n) := \frac{k(n)}{t(n)} - \gamma(n)$. Then

$$\begin{aligned}\xi(n) &= \eta(n) - \sqrt{\frac{n}{2t(n)}} - \gamma(n) \\ &\geq \frac{1}{p(n)} - \sqrt{\frac{n}{2t(n)}}\end{aligned}$$

by our assumption on the gap $\eta(n) - \gamma(n)$. For sufficiently large polynomials $t(n)$, this is at least $1/p'(n)$ for some other polynomial $p'(n)$. The conditions of Lemma 5.4 are satisfied, and therefore for sufficiently large polynomial $t(n)$, the threshold repetition $G^{t,k}$ has negligible security error.

□

Combining this with our Computational No-Learning assumption, we obtain the following:

Corollary 5.5 (NISQ-friendly random circuit OWSG)**.** *Let $\mathcal{N}$ denote a noise model where for some polynomial p, a n-qubit, depth-d circuit C can be run on the noisy quantum computer with fidelity at least $1/p(nd)$. Assuming* $\mathrm{negl}(n)$*-No-Learning, for a sufficiently large polynomial $t(n)$, setting $k(n) = \left(\frac{1}{p(nd)} - \sqrt{\frac{n}{2t(n)}}\right)t(n)$, the threshold repetition $G^{t,k}$ of the random circuit OWSG G from Protocol 4.3 is $(1 - \mathcal{O}(2^{-n}))$-robust against $\mathcal{N}$ and has negligible security error.*

5.2 NISQ-friendly quantum digital signatures

Although one-way functions are a fundamental primitive in classical cryptography, they are not very useful by themselves: their utility comes from being building blocks within cryptographic protocols such as encryption, or pseudorandomness generation [Gol01]. Similarly, one-way state generators are useful as building blocks within *quantum* cryptographic protocols, such as bit commitments [KT24a] or digital signatures [MY22b]. Thus, we would like to realize the utility of a NISQ-friendly OWSG by using it to obtain a NISQ-friendly quantum cryptographic protocol that is amenable to a real world implementation.

We illustrate this possibility with a NISQ-friendly *quantum digital signature scheme*. At a high level, a digital signature scheme is a method for a user to generate a *signature* for a message in a way that a third party (using a public key posted by the user beforehand) can verify that the signature belongs to the message (and in particular, the message or the signature have not been changed). While it has been long known that digital signatures are constructible from one-way functions [Lam79], Morimae and Yamakawa [MY22b] showed that one-way functions are not necessary, and one can use a "fully quantum" primitive instead – namely, one-way state generators.

Here we instantiate the Morimae-Yamakawa digital signature construction with the NISQ-friendly random circuit OWSG from Corollary 5.5. The security of the digital signature scheme follows directly from the security analysis of [MY22b] and the Computational No-Learning assumption. We furthermore argue that since the underlying OWSG is noise robust, so is the digital signature scheme, meaning that it can be implemented on noisy quantum computers.

We first present the formal definition of a signature scheme with quantum public keys:

Definition 5.6. A *signature scheme with quantum keys* is a tuple of algorithms $(\mathrm{SKGen}, \mathrm{PKGen}, \mathrm{Sign}, \mathrm{Ver})$ satisfying the following:

1. The classical randomized polynomial-time algorithm SKGen takes as input a security parameter 1^n, and then outputs a secret key sk, which is a classical string.

2. The QPT algorithm PKGen takes as input a secret key string sk, and deterministically outputs a quantum public-key state $|\mathrm{pk}\rangle$.

3. The classical randomized polynomial-time algorithm Sign takes as input a secret key sk and a message m and outputs a classical signature σ.

4. The QPT algorithm Ver takes as input a quantum public key $|\mathrm{pk}\rangle$, a message m and a candidate signature σ and accepts or rejects.

We say that such a signature scheme satisfies *correctness* if for all security parameters n, for all messages m,

$$\mathbb{P}\left[\begin{array}{c} \mathrm{sk} \leftarrow \mathrm{SKGen}(1^n) \\ \mathrm{Ver}(|\mathrm{pk}\rangle, m, \sigma) \text{ accepts} : \quad |\mathrm{pk}\rangle \leftarrow \mathrm{PKGen}(\mathrm{sk}) \\ \sigma \leftarrow \mathrm{Sign}(\mathrm{sk}, m) \end{array}\right] = 1 - \mathrm{negl}(n) \ .$$

We now define a notion of *one-time security* for a signature scheme with quantum public-keys. Intuitively, the security definition stipulates that a polynomial-time adversary, given copies of the quantum public-key $|\mathrm{pk}\rangle$, a message m (which the adversary can choose) and its corresponding signature σ, cannot produce a valid message-signature pair (m', σ') for some message $m' \neq m$ with non-negligible probability.

Definition 5.7. A signature scheme with quantum public keys $(\mathrm{SKGen}, \mathrm{PKGen}, \mathrm{Sign}, \mathrm{Ver})$ satisfies *one-time security* if for all polynomials $p(n)$, for all pairs of QPT algorithms A_1, A_2, the following holds for sufficiently large n:

$$\mathbb{P}\left[\begin{array}{c} \mathrm{sk} \leftarrow \mathrm{SKGen}(1^n) \\ |\mathrm{pk}\rangle \leftarrow \mathrm{PKGen}(\mathrm{sk}) \\ m' \neq m \wedge \mathrm{Ver}(|\mathrm{pk}\rangle, m', \sigma') \text{ accepts} : \quad m \leftarrow A_1(|\mathrm{pk}\rangle^{\otimes p(n)}) \\ \sigma \leftarrow \mathrm{Sign}(\mathrm{sk}, m) \\ (m', \sigma') \leftarrow A_2(|\mathrm{pk}\rangle^{\otimes p(n)}, m, \sigma) \end{array}\right] = \mathrm{negl}(n)$$

Operationally, a signature scheme with quantum keys is used as follows: there is one party called the *signer* and many *verifiers*. First, the signer will generate a secret key sk and many copies of the public key $|\mathrm{pk}\rangle$. The signer publishes the quantum public keys on some central website on the (quantum) internet.

To sign a message m, the signer computes the classical signature $\sigma \leftarrow \mathrm{Sign}(\mathrm{sk}, m)$ and publishes the message/signature pair (m, σ) on the (classical) internet. To verify that the signature is valid, anyone with a copy of $|\mathrm{pk}\rangle$ can run $\mathrm{Ver}(|\mathrm{pk}\rangle, m, \sigma)$. The one-time security property allows the verifier to have confidence that, if the signer only used the secret key sk once to sign a message, then the signature is valid.

With the definition of signature schemes and their security in hand, we now present our NISQ-friendly signature scheme. For concreteness, we base it on the random circuit OWSG from Corollary 5.5; let t, k be the parameters from the corollary. For simplicity we describe signatures for *single bit* messages; this can be extended to many-bit messages in a straightforward way.

Protocol 5.8. NISQ-friendly signature scheme with quantum public keys

SKGen. Sample $2t$ descriptions of quantum circuits $C_1^{(b)}, \ldots, C_t^{(b)}$ uniformly at random, for both $b = 0, 1$. Set $\mathrm{sk} = (C_i^{(b)})_{i\in[t],b\in\{0,1\}}$.

PKGen: Based on the value of sk, output the public key state

$$|\mathrm{pk}\rangle := \bigotimes_b |C_1^{(b)}\rangle \otimes \cdots \otimes |C_t^{(b)}\rangle \ .$$

Sign: To sign a bit b, output the classical descriptions of the the corresponding circuits. Let $\sigma = (C_1^{(b)}, \ldots, C_t^{(b)})$.

Ver: Given the public key $|\mathrm{pk}\rangle$, a bit b, and a candidate signature σ, first interpret σ as a tuple $(D_1, \ldots, D_t)$ where each D_j is a description of a circuit. For each $j \in [t]$, apply $D_j^\dagger$ to the n qubits of $|C_j^{(b)}\rangle$, and measure the n qubits. If the result is all zeroes, then set $E_j = 1$, otherwise set $E_j = 0$. If $E_1 + \cdots + E_t \geq k$, then accept. Otherwise, reject.

Lemma 5.9 (NISQ-friendly digital signatures)**.** *Let $\mathcal{N}$ denote the noise model from Corollary 5.5, and assume the same ε-No-Learning conjecture as in the corollary. Then, the digital signature scheme described in Protocol 5.8 satisfies correctness and one-time security. Furthermore, the scheme it is noise-robust against $\mathcal{N}$, in that the verification procedure succeeds with high probability, even on a noisy quantum computer:*

$$\mathbb{P}\left[\begin{array}{c} \widetilde{\mathrm{Ver}}(|\mathrm{pk}\rangle, m, \sigma) \text{ accepts}: \\ \end{array} \begin{array}{c} \mathrm{sk} \leftarrow \mathrm{SKGen}(1^n) \\ |\mathrm{pk}\rangle \leftarrow \widetilde{\mathrm{PKGen}}(\mathrm{sk}) \\ \sigma \leftarrow \mathrm{Sign}(\mathrm{sk}, m) \end{array}\right] = 1 - \mathrm{negl}(n) \ .$$

Here $\widetilde{\mathrm{Ver}}$ and $\widetilde{\mathrm{PKGen}}$ denote the noisy executions of the algorithms Ver *and* PKGen *on a quantum computer with noise model $\mathcal{N}$.*

We don't consider noisy versions of SKGen and Sign because these are entirely classical algorithms, which we can run on a noiseless classical computer.

Proof. The correctness property is straightforward to verify, and the security of the digital signature scheme follows from the security of the OWSG from Corollary 5.5 and the No-Learning Assumption. The noise robustness of the digital signature verification follows from the noise robustness of the underlying OWSG proved in Corollary 5.5. □

Acknowledgements

We thank John Bostanci, Jonas Helsen, Yihui Quek, and Abhinav Deshpande for helpful discussions. B.F. and S.G. acknowledge support from AFOSR (FA9550-21-1-0008). This material is based upon work partially supported by the National Science Foundation under Grant CCF-2044923 (CAREER), by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers (Q-NEXT) and by the DOE QuantISED grant DE-SC0020360. H.Y. acknowledges support from AFOSR award FA9550-23-1-0363, NSF CAREER award CCF-2144219, NSF award CCF-2329939, and the Sloan Foundation. M.S. acknowledges support from the NSF

award QCIS-FF: Quantum Computing & Information Science Faculty Fellow at the University of Illinois Urbana-Champaign (NSF 1955032). This work was done in part while some of the authors were visiting the Simons Institute for the Theory of Computing, supported by NSF QLCI Grant No. 2016245.

References

[AA11] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 333–342, 2011.

[AABA+24] Rajeev Acharya, Laleh Aghababaie-Beni, Igor Aleiner, Trond I Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Nikita Astrakhantsev, Juan Atalaya, et al. Quantum error correction below the surface code threshold. *arXiv preprint arXiv:2408.13687*, 2024.

[AABea19] Frank Arute, Kunal Arya, Ryan Babbush, and et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574:505–510, 2019.

[AB96] Dorit Aharonov and Michael Ben-Or. Polynomial simulations of decohered quantum computers. In *Proceedings of 37th Conference on Foundations of Computer Science*, page 46–55. IEEE Comput. Soc. Press, 1996.

[AC17] Scott Aaronson and Lijie Chen. Complexity-theoretic foundations of quantum supremacy experiments. In *Proceedings of the 32nd Computational Complexity Conference (CCC 2017)*, pages 22:1–22:67. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2017.

[AGL+23] Dorit Aharonov, Xun Gao, Zeph Landau, Yunchao Liu, and Umesh Vazirani. A polynomial-time classical algorithm for noisy random circuit sampling. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, pages 1307–1318. ACM, 2023.

[AGS20] Srinivasan Arunachalam, Alex B. Grilo, and Aarthi Sundaram. Quantum hardness of learning shallow classical circuits. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (STOC 2020)*, pages 291–304, New York, NY, USA, 2020. ACM.

[AH23] Scott Aaronson and Shih-Han Hung. Certified randomness from quantum supremacy. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 933–944. ACM, 2023.

[Aha03] Dorit Aharonov. A simple proof that Toffoli and Hadamard are quantum universal. *arXiv preprint quant-ph/0301040*, 2003.

[AQY22] Prabhanjan Ananth, Luowen Qian, and Henry Yuen. Cryptography from pseudorandom quantum states. In *Advances in Cryptology – CRYPTO 2022*, volume 13508 of *Lecture Notes in Computer Science*, pages 208–236. Springer, 2022.

[BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997.

[BBF+24] Roozbeh Bassirian, Adam Bouland, Bill Fefferman, Sam Gunn, and Avishay Tal. On certified randomness from fourier sampling or random circuit sampling, 2024.

[BCMS97] Gilles Brassard, Claude Crépeau, Dominic Mayers, and Louis Salvail. A brief review on the impossibility of quantum bit commitment. *arXiv preprint quant-ph/9712023*, 1997.

[BCN24] John Bostanci, Boyang Chen, and Barak Nehoran. Oracle separation between quantum commitments and quantum one-wayness. *arXiv preprint arXiv:2410.03358*, 2024.

[BCQ23] Zvika Brakerski, Ran Canetti, and Luowen Qian. On the computational hardness needed for quantum cryptography. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.

[BEG+24] Dolev Bluvstein, Simon J Evered, Alexandra A Geim, Sophie H Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, 2024.

[BEM+23] John Bostanci, Yuval Efron, Tony Metger, Alexander Poremba, Luowen Qian, and Henry Yuen. Unitary complexity and the Uhlmann transformation problem. *arXiv preprint arXiv:2306.13073*, 2023.

[BFKL93] Avrim Blum, Merrick Furst, Michael Kearns, and Richard J. Lipton. Cryptographic primitives based on hard learning problems. In Douglas R. Stinson, editor, *Advances in Cryptology — CRYPTO '93: 13th Annual International Cryptology Conference, Santa Barbara, California, USA, August 22–26, 1993, Proceedings*, volume 773 of *Lecture Notes in Computer Science*, pages 278–291. Springer Berlin Heidelberg, 1993.

[BFNV19] Adam Bouland, Bill Fefferman, Chinmay Nirkhe, and Umesh Vazirani. On the complexity and verification of quantum random circuit sampling. *Nature Physics*, 15(2):159–163, 2019.

[BHHP24] John Bostanci, Jonas Haferkamp, Dominik Hangleiter, and Alexander Poremba. Efficient quantum pseudorandomness from hamiltonian phase states, 2024.

[BMM+24] Amit Behera, Giulio Malavolta, Tomoyuki Morimae, Tamer Mour, and Takashi Yamakawa. A new world in the depths of microcrypt: Separating OWSGs and quantum money from QEFID. Cryptology ePrint Archive, Paper 2024/1567, 2024.

[BQSY24] John Bostanci, Luowen Qian, Nicholas Spooner, and Henry Yuen. An efficient quantum parallel repetition theorem and applications. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC 2024)*. ACM, 2024.

[CFSY23] Nai-Hui Chia, Honghao Fu, Fang Song, and Penghui Yao. A Cryptographic Perspective on the Verifiability of Quantum Advantage. 10 2023.

[CGGH24] Bruno P. Cavalar, Eli Goldin, Matthew Gray, and Peter Hall. A meta-complexity characterization of quantum cryptography, 2024.

[CIKK16] Marco Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. Learning algorithms from natural proofs. In *Proceedings of the 31st Computational Complexity Conference (CCC 2016)*, volume 50 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:24, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[CLS24] Nai-Hui Chia, Daniel Liang, and Fang Song. Quantum State Learning Implies Circuit Lower Bounds. 5 2024.

[DHJB24] Alexander M Dalzell, Nicholas Hunter-Jones, and Fernando GSL Brandão. Random quantum circuits transform local noise into global white noise. *Communications in Mathematical Physics*, 405(3):78, 2024.

[DLSS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity, 2014.

[DNS+22] Abhinav Deshpande, Pradeep Niroula, Oles Shtanko, Alexey V. Gorshkov, Bill Fefferman, and Michael J. Gullans. Tight bounds on the convergence of noisy random circuits to the uniform distribution. *PRX Quantum*, 3(4), December 2022.

[FGG+24] Bill Fefferman, Soumik Ghosh, Michael Gullans, Kohdai Kuroiwa, and Kunal Sharma. Effect of nonunital noise on random-circuit sampling. *PRX Quantum*, 5(3), July 2024.

[FGZ24] Bill Fefferman, Soumik Ghosh, and Wei Zhan. Anti-concentration for the unitary haar measure and applications to random quantum circuits, 2024.

[FIP23] NIST FIPS203. Module-Lattice-based Key-Encapsulation Mechanism Standard. *Federal Information Processing Standards Publication*, 2023.

[Gol01] Oded Goldreich. *Foundations of Cryptography: Volume 2, Basic Applications*, volume 2. Cambridge University Press, 2001.

[HH24] Taiga Hiroka and Min-Hsiu Hsieh. Computational complexity of learning efficiently generatable pure states. *arXiv preprint arXiv:2410.04373*, 2024.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[HKP20] 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.

[HLB+23] Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean. Learning shallow quantum circuits. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, pages 1343–1351. ACM, 2023.

[HM24] Taiga Hiroka and Tomoyuki Morimae. Quantum cryptography and meta-complexity, 2024.

[HMY23] Minki Hhan, Tomoyuki Morimae, and Takashi Yamakawa. From the hardness of detecting superpositions to cryptography: Quantum public key encryption and commitments. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 639–667. Springer, 2023.

[HPS24] Hsin-Yuan Huang, John Preskill, and Mehdi Soleimanifar. Certifying almost all quantum states with few single-qubit measurements, 2024.

[IK10] Russell Impagliazzo and Valentine Kabanets. Constructive Proofs of Concentration Bounds. *Electron. Colloquium Comput. Complex.*, TR10-072, 2010.

[IL90] R. Impagliazzo and L.A. Levin. No better ways to generate hard NP instances than picking uniformly at random. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, pages 812–821 vol.2, 1990.

[Imp95] Russell Impagliazzo. A personal view of average-case complexity. In *Proceedings of the 10th Annual Structure in Complexity Theory Conference*, pages 134–147. IEEE Computer Society, 1995.

[JLS+18] Zhengfeng Ji, Yi-Kai Liu, Fang Song, John Watrous, and Henry Yuen. Pseudorandom quantum states. In *Advances in Cryptology – CRYPTO 2018*, volume 10993 of *Lecture Notes in Computer Science*, pages 126–152. Springer, 2018.

[KQST23] William Kretschmer, Luowen Qian, Makrand Sinha, and Avishay Tal. Quantum cryptography in algorithmica. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC 2023)*, pages 1589–1602. ACM, 2023.

[Kre21] William Kretschmer. Quantum pseudorandomness and classical complexity. In *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)*, pages 2:1–2:20. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021.

[KS06] Adam R. Klivans and Alexander A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 553–562. IEEE, 2006.

[KT24a] Dakshita Khurana and Kabir Tomer. Commitments from quantum one-wayness. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 968–978, 2024.

[KT24b] Dakshita Khurana and Kabir Tomer. Founding Quantum Cryptography on Quantum Advantage, or, Towards Cryptography from $\#P$-Hardness. *arXiv preprint arXiv:2409.15248*, 2024.

[KV94] Michael Kearns and Leslie Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM (JACM)*, 41(1):67–95, 1994.

[KW00] Alexei Kitaev and John Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 608–617, 2000.

[Lam79] Leslie Lamport. Constructing digital signatures from a one-way function. Technical Report CSL-98, SRI International Computer Science Laboratory, 1979.

[LL24] Zeph Landau and Yunchao Liu. Learning quantum states prepared by shallow circuits in polynomial time. *arXiv preprint arXiv:2410.23618*, 2024.

[LMN93] Nati Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the Association for Computing Machinery*, 40(3):607–620, 1993.

[LMW24] Alex Lombardi, Fermi Ma, and John Wright. A one-query lower bound for unitary synthesis and breaking quantum cryptography. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 979–990. ACM, 2024.

[LPR10] Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. In Henri Gilbert, editor, *Advances in Cryptology – EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 1–23. Springer, 2010.

[MAG+24] Antonio Anna Mele, Armando Angrisani, Soumik Ghosh, Sumeet Khatri, Jens Eisert, Daniel Stilck França, and Yihui Quek. Noise-induced shallow circuits and absence of barren plateaus, 2024.

[MSY24] Tomoyuki Morimae, Yuki Shirakawa, and Takashi Yamakawa. Cryptographic characterization of quantum advantage, 2024.

[MVM+23] A. Morvan, B. Villalonga, X. Mi, et al. Phase transition in random circuit sampling. *Nature*, 616:70–76, 2023.

[MY22a] Tomoyuki Morimae and Takashi Yamakawa. One-wayness in quantum cryptography. *arXiv preprint arXiv:2210.03394*, 2022.

[MY22b] Tomoyuki Morimae and Takashi Yamakawa. Quantum commitments and signatures without one-way functions. In *Annual International Cryptology Conference*, pages 269–295. Springer, 2022.

[Nat24] National Institute of Standards and Technology. FIPS 204: Module-Lattice-Based Digital Signature Standard, 2024.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *Journal of Computer and System Sciences*, 49(2):149–167, 1994.

[NZ24] Barak Nehoran and Mark Zhandry. A computational separation between quantum no-cloning and no-telegraphing. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2024.

[OJF23] Changhun Oh, Liang Jiang, and Bill Fefferman. On classical simulation algorithms for noisy boson sampling, 2023.

[OS16] Igor C. Oliveira and Rahul Santhanam. Conspiracies between learning algorithms, circuit lower bounds and pseudorandomness, 2016.

[PQS24] Alexander Poremba, Yihui Quek, and Peter Shor. The learning stabilizers with noise problem, 2024.

[QRZ24] Luowen Qian, Justin Raizes, and Mark Zhandry. Hard quantum extrapolations in quantum cryptography. *arXiv preprint arXiv:2409.16516*, 2024.

[RABFF+23] C. Ryan-Anderson, C. H. Baldwin, M. Foss-Feig, D. Hayes, K. Mayer, E. Nielsen, D. Regaldo, S. Ryan, J. Sedlacek, R. T. Sutherland, E. Tirrito, C. Volin, T. Walker, K. White, J. Wootton, and K. Wright. Demonstration of logical qubits and repeated error correction with better-than-physical error rates. *Nature*, 618:500–505, 2023.

[Reg24] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography, 2024.

[S+23] Mark Saffman et al. Quantum computation with programmable neutral-atom arrays. *Nature Physics*, 2023.

[Uhl76] A. Uhlmann. The "transition probability" in the state space of a $*$-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.

[W+22] Christian Weedbrook et al. Quantum advantage with gaussian boson sampling in photonic quantum computers. *Nature Physics*, 2022. Preprint.

[Wer98] Reinhard F Werner. Optimal cloning of pure states. *Physical Review A*, 58(3):1827, 1998.

[Yan22] Jun Yan. General properties of quantum bit commitments. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 628–657. Springer, 2022.

[ZCY+21] Qi Zhao, Hao Chen, Xiao Yuan, et al. Quantum computational advantage via 62-qubit superconducting processor. *Physical Review Letters*, 127(18):180501, 2021.

[ZL+21] Han-Sen Zhong, Yuan Li, et al. Phase-programmable gaussian boson sampling using stimulated squeezed light. *Physical Review Letters*, 127(18):180502, 2021.

[ZLK+24] Haimeng Zhao, Laura Lewis, Ishaan Kannan, Yihui Quek, Hsin-Yuan Huang, and Matthias C Caro. Learning quantum states and unitaries of bounded gate complexity. *PRX Quantum*, 5(4):040306, 2024.

[ZWD+20] Han-Sen Zhong, Hui Wang, Yi-Han Deng, et al. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, 2020.

A A computational Chernoff bound for one-way state generators

In this section we prove a *computational Chernoff bound* for one-way state generators. Roughly speaking, this states that if it is hard to efficiently invert a OWSG G with probability more than γ, then it is hard to efficiently invert noticeably more than γ fraction of instances of the repeated OWSG G^t with non-negligible probability. The reason that this is called a "Chernoff bound" is because it is analogous to proving that the probability a sum of independent, bounded random variables deviates far from its mean is exponentially small. The reason this bound is "computational" is because it only applies to efficient algorithms. This is also commonly known as a *threshold direct product theorem* in complexity theory and cryptography [IK10].

This is a strengthening of the OWSG hardness amplification results of Morimae and Yamakawa [MY22a] which states that it is not possible to efficiently invert *all* t instances of the repeated OWSG G^t with non-negligible probability. Morimae and Yamakawa raised the question of whether a thresholded version of their result can be proven; we answer this affirmatively.

Let $G = (\mathrm{Gen}, \mathrm{Ver})$ be an OWSG, and let $|\psi_k\rangle$ denote the output of G on input k (we omit mention of the security parameter n for convenience). We defined the corresponding threshold repetition, denoted by $G^{t,k}$, in Protocol 5.2. The following lemma bounds its security error.

Lemma A.1 (Computational Chernoff bound for OWSGs)**.** *Let $\xi(n)$ denote an inverse polynomial, i.e., $\xi(n) = 1/p(n)$ for some polynomial $p(n)$. Let $G = (\mathrm{Gen}, \mathrm{Ver})$ be a OWSG with security error $\gamma(n)$. Then for all sufficiently large polynomials $t(n)$, the threshold OWSG $G^{t,k}$ has negligible security error, where $k(n) = (\gamma(n) + \xi(n))t(n)$.*

Proof. Throughout this proof we omit the dependence on n and simply write $t = t(n), k = k(n), \gamma = \gamma(n), \xi = \xi(n)$, etc. Furthermore for notational convenience we write

$$|\psi_{k_1,\dots,k_t}\rangle := |\psi_{k_1}\rangle \otimes \cdots \otimes |\psi_{k_t}\rangle$$

for a tuple of keys $(k_1, \dots, k_t)$.

We prove this via contradiction. Suppose there was a polynomial $q = q(n)$ and a QPT algorithm A such that for infinitely many n,

$$\mathbb{P}\left[\mathrm{Ver}^{t,k}\Big(A(|\psi_{k_1,\dots,k_t}\rangle^{\otimes q}), |\psi_{k_1,\dots,k_t}\rangle\Big) : \begin{array}{c}(k_1,\dots,k_t) \leftarrow (\{0,1\}^r)^t \\ |\psi_{k_1,\dots,k_t}\rangle \leftarrow \mathrm{Gen}^t(k_1,\dots,k_t)\end{array}\right] \geq \frac{1}{q}$$

We now construct a QPT algorithm B that inverts the original OWSG G with probability noticeably greater than γ, which is a contradiction. We closely follow the analysis of the so-called Threshold Direct Product Theorem (which is another name for a computational Chernoff bound) by Impagliazzo and Kabanets [IK10].

This algorithm B is constructed in two stages. We first construct an algorithm A' that tries to solve $t/2$ copies of G, but it either outputs $\perp$, or with high probability inverts $\sim \gamma$ fraction of instances. Then, in the second stage we design the algorithm B for a single instance of G that uses the zero error algorithm A' as a subroutine.

First stage. We first construct the "zero error" algorithm A'. The reason it is called "zero error" is because the algorithm either outputs $\perp$ or outputs (with high probability) a tuple of keys that

passes verification in at least $\sim \gamma$ fraction of coordinates. Furthermore, there is a non-negligible probability of outputting a tuple of keys.

Protocol A.2. The "zero error" algorithm A'

Input: $(4q \ln 8q) \cdot (q+1)$ copies of $|\psi_{k_1}\rangle \otimes \cdots \otimes |\psi_{k_{t/2}}\rangle$.

1. Sample keys $k_{t/2+1}, \ldots, k_t \in \{0,1\}^{r(n)}$.
2. Run the generation algorithm Gen of the OWSG G on the keys to obtain q copies of the output states $|\psi_{k_{t/2+1}}\rangle \otimes \cdots \otimes |\psi_{k_t}\rangle$.
3. Sample a random permutation π on $[t]$.
4. Run the algorithm A on input $\bigotimes_{i\in[t]} |\psi_{k_{\pi(i)}}\rangle^{\otimes q}$ to obtain a tuple of candidate keys $(k'_{\pi(1)}, \ldots, k'_{\pi(t)})$. In other words, the copies of $|\psi_{k_i}\rangle$ states are permuted according to π before being passed into A, and the output of A is unpermuted according to π.
5. Run the threshold verification procedure $\mathrm{Ver}^{t/2,\gamma t/2}$ on input $((k'_{t/2+1}, \ldots, k'_t), |\psi_{k_{t/2+1},\cdots,k_t}\rangle)$. If it accepts, then output $(k'_1, \ldots, k'_{t/2})$ and halt. Otherwise, repeat steps 1 – 4 for at most $4q \ln 8q$ times. If none of the repetitions are successful, output $\perp$.

First, it is easy to verify that A' runs in polynomial time. Next we argue that, conditioned on not outputting $\perp$, the algorithm A' successfully inverts at least a γ' fraction of G instances with high probability.

Claim A.3. *Let $\gamma' = \gamma + \xi/2$. Consider the following probabilistic process. Sample keys $k_1, \ldots, k_{t/2}$, and generate $(4q \ln 8q) \cdot (q+1)$ copies of the corresponding states $|\psi_{k_i}\rangle$. Run A' on those copies, and obtain either $\perp$ or a tuple $(k'_1, \ldots, k'_{t/2})$. The following hold, where the probabilities are over the randomness of the aforementioned process.*

1. $\mathbb{P}\left[\mathrm{Ver}^{t/2,\gamma' t/2}((k'_1, \ldots, k'_{t/2}), |\psi_{k_1,\ldots,k_{t/2}}\rangle) \text{ accepts} \mid A' \text{ does not output } \perp\right] \geq 1 - \xi/4$.
2. $\mathbb{P}\left[A' \text{ does not output } \perp\right] \geq \frac{1}{8q}$.

Proof. This is essentially proved in [IK10, Lemma 5.6]; although they assumed that the Ver algorithm is classical, it can be checked that the proof goes through essentially unchanged in our setting where Ver is quantum and receives a quantum input. □

Second stage. For the second stage we construct an algorithm B that tries to solve a single instance of G by embedding it into the threshold repetition $G^{t,k}$. The algorithm B uses A' as a subroutine.

Protocol A.4. The algorithm B to solve a single instance of G

Input: $(128q \ln \frac{120}{\xi}) \cdot (4q \ln 8q) \cdot (q+1)$ copies of $|\psi_k\rangle$.

1. Sample a random index $i^* \in [t/2]$.
2. Sample random keys k_j for $j \in [t/2] \setminus \{i^*\}$ and generate $(4q \ln 8q) \cdot (q+1)$ copies of the corresponding states $|\psi_{k_j}\rangle$.
3. Run the algorithm A' on input $(4q \ln 8q) \cdot (q+1)$ copies of $\left((\bigotimes_{j<i^*} |\psi_{k_j}\rangle\right) \otimes |\psi_k\rangle \otimes \left((\bigotimes_{j>i^*} |\psi_{k_j}\rangle\right)$.
4. If A' outputs $\perp$, then try steps $1-3$ again for at most $128q \ln \frac{120}{\xi}$ times. If none of the repetitions succeed, then output $\perp$. Otherwise, if A' outputs $(k'_1, \ldots, k'_{t/2})$, then output k'_{i^*}.

Again, it should be clear from construction that B runs in polynomial time. The following argues that B successfully inverts the OWSG G, using a polynomial number of copies of the output of G.

Claim A.5. *Let* $s = (128q \ln \frac{120}{\xi}) \cdot (4q \ln 8q) \cdot (q+1)$. *Then*

$$\mathbb{P}\left[\mathrm{Ver}(B(|\psi_k\rangle^{\otimes s}), |\psi_k\rangle) \text{ accepts} : \begin{array}{c} k \leftarrow \{0,1\}^{r(n)} \\ |\psi_k\rangle \leftarrow \mathrm{Gen}(k) \end{array}\right] \geq \gamma + \xi/20 \ .$$

Proof. This is essentially proved in [IK10, Lemma 5.8]. Again, they prove their result for classical algorithms, but the proof goes through in our quantum setting. □

Putting everything together, we get that B inverts the OWSG G with probability at least $\gamma + \xi/20$, which contradicts the security guarantee of G. Therefore the threshold repetition of $G^{t,k}$ has negligible security error.

□