Certified Randomness from Quantum Supremacy

Scott Aaronson
The University of Texas at Austin
USA
aaronson@cs.utexas.edu

Shih-Han Hung
The University of Texas at Austin
USA
shung@cs.utexas.edu

ABSTRACT

We propose an application for near-term quantum devices: namely, generating cryptographically certified random bits, to use (for example) in proof-of-stake cryptocurrencies. Our protocol repurposes the existing "quantum supremacy" experiments, based on random circuit sampling, that Google and USTC have successfully carried out starting in 2019. We show that, whenever the outputs of these experiments pass the now-standard Linear Cross-Entropy Benchmark (LXEB), under plausible hardness assumptions they necessarily contain $\Omega(n)$ min-entropy, where n is the number of qubits. To achieve a net gain in randomness, we use a small random seed to produce pseudorandom challenge circuits. In response to the challenge circuits, the quantum computer generates output strings that, after verification, can then be fed into a randomness extractor to produce certified nearly-uniform bits—thereby "bootstrapping" from pseudorandomness to genuine randomness. We prove our protocol sound in two senses: (i) under a hardness assumption called Long List Quantum Supremacy Verification, which we justify in the random oracle model, and (ii) unconditionally in the random oracle model against an eavesdropper who could share arbitrary entanglement with the device. (Note that our protocol's output is unpredictable even to *a computationally unbounded adversary who can see the random oracle*.) Currently, the central drawback of our protocol is the exponential cost of verification, which in practice will limit its implementation to at most $n \sim 60$ qubits, a regime where attacks are expensive but not impossible. Modulo that drawback, our protocol appears to be the *only* practical application of quantum computing that both requires a QC and is physically realizable today.

CCS CONCEPTS

• **Theory of computation → Cryptographic protocols**.

KEYWORDS

certified randomness, random circuit sampling

ACM Reference Format:

Scott Aaronson and Shih-Han Hung. 2023. Certified Randomness from Quantum Supremacy. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23), June 20–23, 2023, Orlando, FL, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3564246.3585145

STOC '23, June 20–23, 2023, Orlando, FL, USA

ACM ISBN 978-1-4503-9913-5/23/06.
https://doi.org/10.1145/3564246.3585145

1 INTRODUCTION

After three decades of quantum computing theory and experiment, the world finally has noisy quantum devices, with $50 - 60$ qubits or ~ 100 photons, that solve special sampling problems in a way that's conjectured to outperform any existing classical computer. The devices include Google's 53-qubit "Sycamore" chip [9], USTC's "Jiuzhang" [48] and "Zu Chongzhi" [46], and most recently Xanadu's "Borealis" [30]. The sampling problems, which include Random Circuit Sampling [5] and BosonSampling [3], grew directly out of work in quantum complexity theory beginning around 2010.

To be clear, it's still debated in which senses current devices have achieved the milestone of "quantum supremacy"—a term coined by Preskill [37] in 2012, to refer to an orders-of-magnitude speedup over all known classical approaches for some well-defined (but not necessarily useful) computational task. On the one hand, since Google's original 2019 announcement [9], the sampling experiments have continued to improve, for example in number of qubits and circuit depth (for RCS) [46], and in number of photons and measurement fidelity (for BosonSampling) [30]. One expects further improvements. On the other hand, classical spoofing attacks against the experiments have *also* improved—with some attacks based on tensor-network contraction (e.g., [35]), and others taking advantage of noise in the devices (e.g., [10]). Notably, however, the attacks that fully replicate the Google device's observed performance, such as that of [35], still have inherently exponential scaling, and still seem to require an ExaFLOPS supercomputer to match or beat the Google device's running time of ~ 3 minutes. As a rough estimate, the Summit supercomputer uses 13 megawatts, while Google [9, Appendix H] estimated that the dilution refrigerator for its 53-qubit QC uses ~ 20 kilowatts. Thus, *despite* the QC's extreme need for refrigeration, it still wins by a factor of hundreds as measured by electricity cost.

For some, the recent quantum supremacy demonstrations were important mostly because they showcased many of the key ingredients of a future *fault-tolerant, scalable* quantum computer—and just as importantly, did not detect any correlated errors of the sort that would render fault-tolerant quantum computing impossible. For others, however, these experiments have done more: namely, they've inaugurated the era of "NISQ" or Noisy Intermediate Scale Quantum computation, another term coined by Preskill [38]. The hope of NISQ is that, even *before* fault-tolerance is achieved, noisy QCs with up to (say) 1000 qubits might already prove useful for certain practical problems, just like various analog computing devices were useful even before the invention of the transistor inaugurated the digital era.

Unfortunately, despite the billions that have by now been invested into NISQ hopes, the lack of any obvious "killer app" for NISQ devices has emerged as a defining fact of the field.

Perhaps NISQ devices will be useful for simulation of condensed-matter physics or even quantum chemistry. Alas, while there are exciting proposals for quantum simulations that would need only a few hundred qubits (e.g., [40]), these proposals invariably have the drawback of requiring thousands or millions of layers of gates. Unless it can be remedied, this would put them completely out of reach for NISQ devices. Or perhaps NISQ devices will yield speedups for optimization problems, via quantum annealing or QAOA [23]. Alas, despite years of intense theoretical and empirical work, researchers have struggled to show any clear advantage for quantum annealing or QAOA over classical computing, for any practical optimization problem—*let alone* an advantage that would be achievable on a NISQ device.

We should add that, in recent years, there have been striking *new* ideas for how to demonstrate quantum supremacy. These include interactive protocols that exploit trapdoor one-way functions [15, 27], as well as the spectacular result of Yamakawa and Zhandry [47], which gave an exponential quantum speedup for an NP search problem relative to a random oracle. Alas, pending some breakthrough, none of these ideas seem to be implementable on a NISQ device.

1.1 Our Contribution

This paper studies, to our knowledge for the first time, whether current sampling-based supremacy experiments might *themselves* have a useful application outside of physics.[1] We focus on the generation of *cryptographically certified random bits*.

Needless to say, it is easy to use a quantum computer—or for that matter, even a Geiger counter next to some radioactive material—to generate as many random bits as we like: bits that quantum mechanics itself predicts will be fundamentally unpredictable. The problem is, how do we convince a skeptic over the Internet, with no access to our hardware, that the bits were indeed random, and not secretly backdoored? This is not just a theoretical worry: for example, as a byproduct of the Edward Snowden revelations in 2013, the world learned that a NIST pseudorandomness standard known as Dual_EC_DRBG was indeed backdoored by the US National Security Agency.[2]

Certified randomness has become a significant practical problem—particularly with the rise of *proof-of-stake cryptocurrencies*, which notably include Ethereum[3] (market cap at time of writing: $163 billion), following its migration on September 15, 2022. In proof-of-stake systems, lotteries are continually run to decide which currency holder gets to add the next block to the blockchain. There is no trusted authority to manage these lotteries, yet the entire system rests on the assumption that they are conducted honestly and without bias. Other applications of certified randomness include non-interactive zero-knowledge proofs, and financial and electoral audits.

One approach to the certified-randomness problem uses blockchains themselves as a source of random bits—with the argument being that anyone who could predict the bits could exploit their predictability to get rich [14]. Other approaches look to the social or natural worlds for a source of publicly verifiable entropy: for example, perhaps one could use the least significant digits of the Dow Jones Industrial Average, or the patterns of granules that form on the surface of the Sun.

More relevant for us, since 2009, an exciting line of work has shown how to use measurements on entangled particles as a source of physically certified randomness [19, 20, 33, 34, 36, 44]. The idea is that, if the measurement outcomes are observed to violate the Bell/CHSH inequality, then by that very fact, the outcomes *cannot* have been secretly deterministic, unless there was secret communication between the "Alice" and "Bob" detectors. Furthermore, depending on the experimental setup, this communication might need to have occurred faster than light. Thus, the outcomes must contain genuine entropy, which can be fed into a randomness extractor to purify it into nearly-uniform random bits. The technical part is that, in a Bell/CHSH experiment, the measurement bases must *themselves* be unpredictable—and thus, they need to be chosen judiciously if we want an overall net gain in randomness. This is the problem that the line of works [19, 20, 33, 34, 36, 44] has now almost completely solved.

Bell/CHSH-based certified randomness protocols have already been experimentally demonstrated [13], and are even in consideration for practical deployment in the NIST Randomness Beacon [28], which generates 512 random bits per minute.

The central drawback of these protocols is that a user, downloading allegedly random bits from the Internet, has no obvious way to verify that the "Alice" and "Bob" detectors were out of communication—the key assumption needed for security. Indeed, in some Bell/CHSH experiments, "Alice" and "Bob" are mere feet away! But even if they weren't, how would this be proved?

The central insight of this paper is that sampling-based quantum supremacy experiments provide an entirely different route to certified randomness—a route that requires only a single quantum device, while also being practical today. In our protocol, a classical verifier uses a small random seed to generate n-qubit challenge circuits $C_1, C_2, \ldots$ pseudorandomly. The verifier then submits these C_i's one at a time, presumably over the standard Internet, to a quantum computer server. For each C_i, the server needs to respond quickly—say, in less than one second—with independent samples $s_1, \ldots, s_k$ from C_i's output distribution: that is, the distribution over $\{0,1\}^n$ obtained by running C_i on the initial state $|0^n\rangle$ and then measuring in the computational basis.

The verifier, at its leisure, can then calculate the so-called *Linear Cross-Entropy Benchmark*,

$$\text{LXEB} := \sum_{j=1}^{k} |\langle s_j | C_i | 0^n \rangle|^2,$$

for at least some of the challenge circuits C_i. *If* the LXEB scores are sufficiently large, our analysis shows that the verifier can then be

[1] Some earlier work explored whether BosonSampling might be useful for (e.g.) calculating molecular vibronic spectra [26] or graph similarity detection [41], but those hopes were killed by efficient classical simulations.
[2] https://en.wikipedia.org/wiki/Dual_EC_DRBG
[3] https://en.wikipedia.org/wiki/Ethereum

confident, under plausible computational assumptions, that there must be $\Omega(n)$ bits of genuine min-entropy in the returned samples.

In other words: even a quantum computer should need $\exp(n)$ time to generate samples that pass the LXEB test and yet are secretly deterministic or nearly-deterministic functions of C_i. For a typical circuit C_i, an honest sample from the output distribution will contain $n - O(\log n)$ bits of min-entropy. A dishonest quantum computer could *somewhat* reduce the entropy of the returned samples—for example, by generating many samples and then returning only those that start with 0 bits. But doing better, by finding (e.g.) the lexicographically first samples that pass the LXEB test, or the samples that maximize the LXEB score, should be exponentially hard even quantumly, requiring amplitude amplification or the like (while a subexponential classical algorithm wouldn't stand a chance). The purpose of our security reductions, which we will explain in detail in Section 2, is just to formalize these simple intuitions.

Assuming the returned samples (or enough of them) pass the LXEB test, the last step of our protocol is to feed them into a classical *seeded randomness extractor*, to produce output bits that are exponentially close in total variation distance to uniformly random.

Stepping back, many people have pointed out the close analogy between

(1) the Bell/CHSH experiments, which ruled out local hidden-variable theories (and which have now been recognized with the Nobel Prize in Physics), and
(2) sampling-based quantum supremacy experiments, which seek to rule out "classical polynomial-time hidden-variable theories."

This paper shows that the analogy goes even further. In both cases, the original purpose of the experiment was just to demonstrate the reality of some quantum phenomenon, and rule out any classical explanation—but we then get certified randomness as a "free byproduct" of the demonstration. In both cases, the entire setup hinges on a numerical inequality—one that any classical explanation must satisfy, that quantum mechanics predicts can be violated by a large amount, and that realistic experiments can violate albeit by less than the maximum that quantum mechanics predicts. In both cases, *any* violation of the inequality turns out to suffice for the certified randomness application.

We note, lastly, that our protocol *inherently requires* the use of a quantum computer. This can be seen as follows: consider any server that's simulable in classical probabilistic polynomial-time. Then by definition, there can be no efficient way to distinguish that server from a simulation *whose randomness has been replaced by the output of a pseudorandom generator*. Indeed, if the pseudorandom generator has an m-bit seed, then the best distinguishing algorithm would be expected to take $\exp(m)$ time—which means that even given the $\sim 2^n$ time that we allow for verification, the verifier *still* cannot distinguish an honest server from one with only m bits of true entropy, for any $m \gg n$.

How does our actual quantum protocol evade the above impossibility argument? Simply by a fact used again and again in quantum complexity theory: namely, that there is no notion of "pulling the randomness" (or quantumness) out of a quantum algorithm, for example to replace it with pseudorandomness, analogous to what is possible with classical randomized algorithms. One could also say: our protocol's security analysis will depend on a computational assumption, that the problem of "Long List Quantum Supremacy Verification" is hard for the complexity class QCAM, whose classical analogue is simply false. The reasons for this, in turn, are closely related to one of the elemental differences between classical and quantum computation, that PostBPP (BPP with postselected outputs) is contained in the polynomial hierarchy and can be simulated using approximate counting, whereas PostBQP = PP can express #P-complete problems.

1.2 Practical Considerations

Our certified randomness protocol could be demonstrated on existing devices, with $n = 60$ qubits or some other number in the "quantum supremacy regime." However, there are practical and even conceptual issues to be sorted out before deploying the protocol for proof-of-stake cryptocurrency or any other critical application.

Verification cost. The central drawback of our protocol, as it stands, is that to check the server's outputs, the classical verifier needs to calculate a Linear Cross-Entropy score, and this is expected to take $\sim 2^n$ time—similar to the time needed for classical spoofing. This drawback is directly inherited from Random Circuit Sampling and all other current approaches to NISQ quantum supremacy itself.

Because of the verification cost, n, the number of qubits, must be chosen small enough that 2^n is still within range of the most powerful classical supercomputers available. If so, however, the issue is obvious: 2^n would *also* be within range of a sufficiently dedicated classical spoofer, who could then predict and control the allegedly random bits.

Nevertheless, we claim that not all hope is lost. The crucial observation is that spoofing, to be effective, needs to be *continual*: for example, if the challenge circuits are submitted every second, then the spoofer needs to run nearly every second as well. Even if a real quantum computer were used (say) every *other* second, the outputs would contain a lot of genuine min-entropy, which would suffice for a secure protocol. The spoofing also needs to be *fast*—as fast as the QC itself.

One might object that, since most classical algorithms to simulate quantum circuits are highly parallelizable, spoofing our protocol within some exacting time limit is "merely" a matter of spending enough money on classical computing hardware. When (say) $n = 60$, though, we estimate that the expenditure, to do $\exp(60)$ operations per second, would run into billions of dollars, outside the means of all but corporations and nation-states.

Verification, by contrast, only needs to be *occasional*. Using a tiny amount of seed randomness, the verifier can choose $O(1)$ random rounds of the protocol and spot-check only those. Then a malicious server that spoofed even (say) 10% of the rounds would be caught with overwhelming probability. Verification can also be done *at leisure*: so long as the verifier is satisfied to catch the spoofer after the fact, the verifier could spend hours or days where the spoofer needed to take less than a second. Indeed, to keep the server honest, arguably the verification need not even be done: it's enough to threaten credibly that it *might* be done!

Having said that, of course it would be preferable if the verification could be done in $n^{O(1)}$ time, in some way that retained the protocol's "NISQiness."

In our view, whether it's possible to achieve sampling-based quantum supremacy, *on a NISQ device and with efficient classical verification*, has become one of the most urgent open problems in quantum computing theory, even independently of this work. Our work further underscores the problem's importance, by showing how a solution could turn secure, practical certified randomness into the first real application of quantum computers.

Interactivity. A second practical issue with our protocol is the need for the verifier continually to generate new challenge circuits that are unpredictable to the quantum computing server. One could reasonably ask: if the verifier has that ability, then why does it even *need* the quantum computer to generate random bits?

The short answer is that our protocol offers an "upgrade" in the level of unpredictability: the challenge circuits only need to be *pseudorandom* (for reasons to be explained later, against a QSZK adversary). So in particular, the verifier can generate all the circuits deterministically from a single initial random seed. The protocol's output, by contrast, is guaranteed to be *genuinely* random.

Indeed, our protocol offers an appealing "forward secrecy" property. Namely, even if we imagine that the verifier's pseudorandom generator will be broken in the future, so long as the server can't break the PRG *at the time the protocol is run*, the server is forced to generate truly random bits. Such bits will of course remain unpredictable, conditioned on anything that doesn't depend on themselves, regardless of any future advances in cryptanalysis.

Who verifies the verifier? Still, there remains a difficulty: the verifier checks the QC's outputs, but who checks the verifier? If the verifier just wants random bits for its own private use, then there is no problem: the verifier could use our protocol, for example, to check random bits output by a QC that was bought from an untrusted manufacturer. But consider an application like proof-of-stake cryptocurrency, where the certified random bits need to be shared with the world. Does the world designate some organization to play the role of the verifier? If so, then why couldn't that organization be corrupted or infiltrated, as surely as the organization that owns the quantum computer—bringing us back where we started?

Classical cryptography suggests a variety of potential solutions to this dilemma. For example, perhaps a dozen or more classical verifiers each generate their own pseudorandom sequences, and those sequences are then XORed together to produce a single sequence which is used to generate the challenge circuits to send to the quantum computing server. If even one verifier wants the sequence to be unpredictable to the server, then it will be, provided that no verifier can see any other verifier's sequence before committing to its own.

Again one could ask: if we trust such a XOR protocol, then why not just use its outputs directly, and skip the quantum computer? Again our answer appeals to the "randomness bootstrap": provided we agree that the XOR'ed sequence is unpredictable *in practice, for now*, the quantum computer's output will then be *fundamentally* unpredictable. Our protocol thus provides an upgrade in the level of unpredictability.

1.3 Related Work

We are not the first to propose using a quantum computer to generate certified random bits, which are secure under some computational hardness assumption. Brakerski, Christiano, Mahadev, Vazirani, and Vidick [15] gave an elegant scheme based on the assumed hardness of the Learning With Errors (LWE) problem. In subsequent work, Mahadev, Vazirani, and Vidick [31] showed that the Brakerski et al. protocol generates $\Omega(n)$ random bits per round, which matches our protocol.

The central advantage of the Brakerski et al. protocol over ours is that its outputs can be verified in classical polynomial time. On the other hand, unlike ours, their protocol seems difficult or impossible to implement on a NISQ device, because it requires evaluating complicated cryptographic functions on superpositions of inputs. In addition, their protocol requires the quantum computer to maintain a coherent superposition state *while it interacts with the verifier*, presumably over the Internet. This is currently feasible only with certain hardware platforms, such as trapped ions, and not for example with superconducting qubits (whose coherence times are measured in microseconds).

More recently, as a byproduct of their breakthrough on an exponential quantum speedup for NP search relative to a random oracle, Yamakawa and Zhandry [47] gave a different interactive protocol to certify $\Omega(\log n)$ random bits, in the random oracle model and also assuming the so-called *Aaronson-Ambainis conjecture* [2]. We do not know whether the Yamakawa-Zhandry protocol remains secure against an entangling adversary, nor whether it accumulates entropy across multiple rounds. In any case, theirs is again a protocol that evaluates complicated functions on superpositions of inputs, meaning there is little or no hope of running it on a NISQ device.

In contrast to these earlier works, here we pursue the "minimalist approach" to generating certified randomness using a quantum computer: we eschew all cryptography done in superposition, and just examine the output distributions of random or pseudorandom quantum circuits. By taking this route, we give up on efficient verification, but we gain feasibility on current hardware, as well as a conceptual unification of certified randomness with sampling-based quantum supremacy itself.

Recently, building on the unpublished announcements by one of us (SA) of the research now reported in this paper, Bassirian, Bouland, Fefferman, Gunn, and Tal [11] took some first steps toward analyzing the use of sampling-based quantum supremacy experiments for certified randomness. Their first result says that, relative to a random oracle, any efficient quantum algorithm for Fourier Sampling must generate $\Omega(n)$ bits of min-entropy as a byproduct of its operation. Their second result says that Long List Quantum Supremacy Verification (LLQSV), the problem that underlies our hardness reduction, is neither in BQP nor in PH relative to a random oracle. To prove non-containment in PH, they build on the breakthrough oracle separation between BQP and PH due to Raz and Tal [39].

These results are of course closely related to ours, but they fall short of a soundness analysis for our certified randomness protocol. We go further than [11] in at least four respects:

(1) We prove that a plausible hardness assumption about LLQSV implies the generation of certified random bits. This reduction does not depend on a random oracle.
(2) We give black-box evidence for that hardness assumption. (The result of [11], that black-box LLQSV is not in PH, is interesting and new, but neither necessary nor sufficient for us. As we'll explain, we need non-containment in the class QCAM/qpoly.)
(3) We prove the accumulation of entropy across multiple rounds.
(4) In the black-box setting, we prove security against an entangled adversary.

Our proof techniques are also independent of those in [11].

Lastly, let us mention that Brandão and Peralta [17] have already reported numerical calculations to find appropriate parameter settings for the protocol described in this paper.

1.4 This Paper's History

One of us (SA) conceived the certified randomness protocol, as well as basic elements of its soundness analysis (e.g., the LLQSV $\notin$ QCAM/qpoly hardness assumption), in February 2018. SA then gave various public talks about the proposal (e.g., [1]), albeit only sketching the analysis. Those talks influenced subsequent work on quantum supremacy: for example, the Google group cited them as motivation in its 2019 paper announcing its 53-qubit Sycamore experiment [9].

Alas, the soundness analysis ended up being too involved for SA to complete alone. That and other factors caused a more than four-year delay in writing up this paper. Here, we not only complete the analysis that SA announced in 2018: we also prove security, in the random oracle model, *against an adversary who could be arbitrarily entangled with the QC*. This goes beyond what SA claimed in 2018, and indeed addresses one of the central open problems raised at that time.

1.5 Future Directions

Many important problems remain:

- As mentioned before, perhaps the biggest problem is to design a sampling-based quantum supremacy experiment that both runs on a NISQ device *and* admits efficient classical verification. If such an experiment were developed, then based on our results here, we predict that it could be readily repurposed to get a secure, efficiently-verifiable certified randomness scheme that runs on existing devices.
- Short of that, it would also be interesting to adapt our randomness protocol from Random Circuit Sampling (RCS) to other known quantum supremacy proposals, such as BosonSampling [3] and IQP [18]. With BosonSampling, the problem is that we currently lack a crisp, quantitative conjecture about the best that a polynomial-time classical algorithm can do to spoof tests such as the Linear Cross-Entropy Benchmark (LXEB). From 2013 work of Aaronson and Arkhipov [4], we know that, in contrast to what we conjecture for RCS, efficient classical algorithms can get *some* depth-independent, $\Omega(1)$ LXEB advantage for BosonSampling, but how much? Answering this question seems like a prerequisite to designing a certified randomness protocol, as it would set the lower bound on how well a BosonSampling experiment has to do before it can be used for such a protocol.
- Of course, it would be great to know more about the truth or falsehood of the central hardness conjectures on which we base our protocol's security—e.g., that "Long List Quantum Supremacy Verification" (LLQSV) lacks a QCAM/qpoly protocol. It would be also great to prove our protocol's security under weaker assumptions. Can we at least remove the exponentially long list of circuits, and use a hardness assumption involving a single circuit?
- In the setting with an entangled adversary, we can currently prove security *only* in the random oracle model. Can we state a plausible hardness assumption that suffices for that setting?
- Under some plausible hardness assumption, can we tighten the lower bound on the amount of min-entropy generated per sample—even up to the maximum of $n - O(\log n)$?[4]
- Likewise, can we show that more and more min-entropy continues to be generated, even if we sample with the same circuit C over and over? Clearly there is a limit here: once enough time has elapsed that a spoofer could have explicitly calculated C's entire output distribution, and perhaps even stored it in a giant lookup table, C is no longer secure and needs to be replaced by a new circuit. But can we at least go up to that limit? To whatever extent we can, our protocol would become much more efficient in practice—especially once we factor in (e.g.) the time needed to calibrate a superconducting QC on a new circuit C.
- We know, both from the Haar-random approximation and from extensive numerical evidence, that an ideal, honest QC *does* succeed at the Linear Cross-Entropy Benchmark with overwhelming probability, given a random quantum circuit C as input. And in some sense, since this fact is never needed in our security analysis, empirical evidence suffices for it! All the same, it is strange that a rigorous proof of the fact is still lacking, at least for "natural" quantum circuit ensembles. We prove the fact in the random oracle model, but can we prove it outright? Recent advances showing that random quantum circuits yield t-designs [16, 25] take us part of the way, but an additional idea seems needed.

2 TECHNICAL OVERVIEW

In this section, we give an overview of our technical contribution. For the detailed analysis and discussions, we refer the readers to the full version of our paper [7].

[4] $n - O(\log n)$ is the maximum because the quantum computer could always (say) generate $n^{O(1)}$ samples from the correct distribution until it finds one whose first $O(\log n)$ bits are all 0's.

2.1 Our Basic Result (without Entangled Adversary)

Throughout this paper, we let C be a quantum circuit acting on n qubits, and we let $N = 2^n$ be the Hilbert space dimension. Let P_C be the probability distribution defined by $p_C(z) = |\langle z|C|0^n\rangle|^2$. It is well-known that when $C \sim \text{Haar}(N)$, the Haar measure over $N \times N$ unitary matrices, we have

$$\mathop{\mathbb{E}}_{C} \mathop{\mathbb{E}}_{z\sim P_C} [p_C(z)] = \frac{2}{N+1}. \tag{1}$$

In 2019, Google [9] announced an experiment to show quantum advantage on the following task, called Linear Cross-Entropy Benchmarking (LXEB).

Problem 1 (Linear Cross-Entropy Benchmarking $\text{LXEB}_{b,k}(\mathcal{D})$)**.** *Let $\mathcal{D}$ be a probability distribution over quantum circuits on n qubits. Then the $\text{LXEB}_{b,k}(\mathcal{D})$ problem is as follows: given C drawn from $\mathcal{D}$, output samples $z_1, \ldots, z_k \in \{0,1\}^n$ such that*

$$\frac{1}{k}\sum_{i=1}^{k} p_C(z_i) \geq \frac{b}{N}. \tag{2}$$

We sometimes omit the argument $\mathcal{D}$.

Intuitively, we expect that a polynomial-time classical algorithm should be unable to solve $\text{LXEB}_{b,k}$ for any $b = 1+\Omega(1)$. By contrast, if an ideal quantum computer simply runs C over and over on the initial state $|0^n\rangle$, measures in the computational basis, and returns the results, then approximating C by a random unitary, by (1) we expect the QC to solve $\text{LXEB}_{b,k}$ with $1 - 1/\exp(k)$ success probability for any constant $b < 2$. Meanwhile, a noisy QC could be expected to solve $\text{LXEB}_{b,k}$ for some b greater than 1 but less than 2—and indeed that's exactly what's observed empirically, with (for example) Google's 2019 experiment achieving $b \sim 1.002$.

In this paper, our aim is to show, not merely a quantum advantage over classical in solving $\text{LXEB}_{b,k}$, but a quantum sampling advantage over *any efficient algorithm—quantum or classical—that returns the same s_i's a large fraction of the time when given the same circuit C.*

To do this, we'll use a new and admittedly nonstandard hardness assumption, but one that strikes us as extremely plausible. Our assumption concerns the following problem:

Problem 2 (Long List Quantum Supremacy Verification $\text{LLQSV}(\mathcal{U})$)**.** *We are given oracle access to $M = O(2^{3n})$ quantum circuits $C_1, \ldots, C_M$, each on n qubits, which are promised to be drawn independently from the distribution $\mathcal{U}$. We're also given oracle access to M strings $s_1, \ldots, s_M \in \{0,1\}^n$. Then the task is to distinguish the following two cases:*

1. ***No-Case**: Each s_i is sampled uniformly and uniformly from $\{0,1\}^n$.*
2. ***Yes-Case**: Each s_i is sampled from p_{C_i}, the output distribution of C_i.*

Our hardness assumption, which we call the Long List Hardness Assumption ($\text{LLHA}_B(\mathcal{U})$), now says the following, for some parameter $B < n$:

$$\text{LLQSV}(\mathcal{U}) \notin \text{QCAMTIME}(2^B)/\text{q}(2^B n^{O(1)}). \tag{3}$$

Here QCAM, or Quantum Classical Arthur Merlin, is the class of problems that admit an AM protocols with classical communication and a quantum verifier. $\text{QCAMTIME}(T)$ is the generalization of QCAM where the verifier can use running time T (the communication is still restricted to be polynomial). $\text{QCAMTIME}(T)/\text{q}(A)$ is the same, but where the verifier now receives A bits of quantum advice that depend only on n.

Our first main result is then the following.

Theorem 2.1 (Single-round analysis, no side information, informal). *Let $\mathcal{U}$ be a distribution over n-qubit quantum circuits, and suppose $\text{LLHA}_B(\mathcal{U})$ holds. Also, let $\mathcal{A}$ be a polynomial-time quantum algorithm that solves $\text{LXEB}_{b,k}(\mathcal{U})$ with probability at least q. Then $\mathcal{A}$'s output, $s_1, \ldots, s_k$, satisfies*

$$\Pr_{C'\sim\mathcal{U}}\left[H_{\min}(s_1 \ldots s_k|C = C') \geq B/2\right] \geq \left(\frac{bq-1}{b-1} - o(1)\right), \tag{4}$$

where $H_{\min}(\{p_i\}) := \min_i \log_2 \frac{1}{p_i}$ is the min-entropy.

To illustrate, suppose we set $B := 0.49n$—the best *upper* bound that we know, $B < n/2$, follows from Grover's algorithm. Suppose also that $b = 1.002$, as in Google's experiment [9], and that k is chosen large enough so that $q \geq 0.9990$ by a large deviation inequality. Then Theorem 2.1 is telling us that $\mathcal{A}$'s output must contain at least $(0.12 - o(1))n$ random bits.

Interestingly, while Theorem 2.1 is stated in terms of min-entropy, and while our eventual multi-round result will *also* be stated in terms of min-entropy, as an intermediate step it's convenient to switch to Shannon entropy, as this is what entropy accumulation theorems use. Of course, since $H_{\min}(\mathcal{D}) \leq H(\mathcal{D})$ for every distribution $\mathcal{D}$, Theorem 2.1 immediately implies the same lower bound on Shannon entropy. Indeed, since Shannon entropy behaves linearly with respect to expectation, Theorem 2.1 implies that

$$H(s_1, \ldots, s_k|C) \geq \frac{B}{2} \cdot \left(\frac{bq-1}{b-1} - o(1)\right). \tag{5}$$

While $\text{LLHA}_B(\mathcal{U})$ is admittedly a strong assumption, our next result justifies it by proving that it holds in the random oracle model:

Theorem 2.2 (Hardness of $\text{LLQSV}(\mathcal{U})$, informal). *Given a random oracle $\boldsymbol{O}$, let $\mathcal{U}$ be the uniform distribution over $M = 2^{O(n)}$ quantum circuits $C_1, \ldots, C_M$, which Fourier-sample disjoint Boolean functions $f_1, \ldots, f_M : \{0,1\}^n \to \{-1,+1\}$ respectively defined by A. Then $\text{LLHA}_B(\mathcal{U})$ holds relative to $\boldsymbol{O}$ for $B = \Omega(n)$.*

Here we outline the proof. First we give a reduction for LLQSV from another problem called Boolean Function Bias Detection (BFBD). In the latter problem, the algorithm is given access to M functions sampled from either a distribution $\mathcal{D}$ or the uniform distribution. The distribution $\mathcal{D}$ can be described with the following process: First sample a integer $r \in \{0, 1, \ldots, N\}$ with probability $N(1 - 2r/N)^2 \cdot \binom{N}{r} 2^{-N}$, sample a random subset $R \subseteq \{0,1\}^n$ of size r, and finally set $f(x) = -1$ if and only if $x \in R$. Since both distributions are concentrated around $r = N/2$, a simple hybrid argument leads to a basic lower bound for BQP. We then extend the hardness result to interactive proof systems, specifically, the class QIP[2] of two-message quantum interactive proofs, using a

similar argument: Recall that the prover's goal is always to convince the verifier that the given function is sampled from $\mathcal{D}$. From the observation stated above, we can modify a function $f \sim \mathcal{D}$ on a small number of points to yield a random function. Thus a prover which convinces the verifier to accept $f \sim \mathcal{D}$ would also convince the verifier to accept a random function. Then by the inclusion QCAM $\subseteq$ QIP[2], LLQSV is hard for QCAM.

To prove the desired hardness for the non-uniform class QCAM/qpoly (or generalizations of it to use more queries and more advice), we observe that by Aaronson and Drucker's exchange theorem [6], QCAM/qpoly $\subseteq$ QMA/poly, so it suffices to show hardness against the latter class. We then change the oracle model (and not the problem itself) as follows: The oracle O contains N sections, each indexed by an n-bit string x. The non-uniform protocol given oracle access to O, input x, and all the samples (also indexed by x), is challenged to determine whether the sample s_x is sampled from O_x or uniform (see Problem 2).

Clearly any solver for LLQSV can determine whether s_x is sampled from O_x for every $x \in \{0,1\}^n$. By replacing the classical advice with a random guess, we show that any QMA/poly verifier solving the problem would imply a $Nn^{O(1)}$-query QMA verifier solving N problems with probability $2^{-\mathrm{poly}(n)}$. Then we appeal to the *strong direct product theorem* by Sherstov [42], who showed that even to achieve success probability $2^{-\Omega(N)}$, computing N independent problems requires $\Omega(Nd)$ queries, where d is the query lower bound of a single problem obtained using the polynomial method. Finally we derive a contradiction by showing that a single problem has an exponential lower bound.

Building upon the above single-round analysis, the next step is to showk *accumulation* of entropy across multiple rounds. We give a simple m-round entropy accumulation process using $\mathrm{LXEB}_{1+\delta,k}$ as the verification for $m = n^{O(1)}$[5] and constant $0 < \delta < 1$. In each round, for $\gamma = O(\log n/m)$, the verifier sends the same circuit as in the previous round with probability $1-\gamma$, or samples a fresh random circuit with probability γ. We define an *epoch* to be an interval of consecutive rounds where the same circuit is sent.[6] With overwhelming probability, there are at most $O(\log n)$ epochs. The verifier chooses k random samples for each circuit, and checks if the verifier passes $\mathrm{LXEB}_{1+\delta,k}$ for 99% of the given circuits. Applying an Entropy Accumulation Theorem (EAT, explained in Section 2.4), we prove the following statement.

Theorem 2.3 (Entropy accumulation, no side information, informal). *For $\beta \in [0,1]$, if $\mathrm{LLHA}_{\beta n}$ holds, then for integer $k = \Omega(n^2)$ and $m = \Omega(\log n)$, there exists an m-round entropy accumulation protocol taking $k \cdot m$ samples such that conditioned on the event Ω of not aborting,*

$$H_{\min}(Z|C)_{\rho|\Omega} \geq n\left(\left(0.99 - \frac{0.01}{\delta}\right)\frac{\beta}{2}m - O(\sqrt{m})\right) \tag{6}$$

for every device solving $\mathrm{LXEB}_{1+\delta,k}$, where ρ is the output state, Z is the responses received from the device, and C is the circuit in each round.

[5] Potentially m can be exponentially large, but we do not pursue this here.

[6] Brakerski, Christiano, Mahadev, Vazirani and Vidick [15] used the same concept of "epochs" to analyze their certified randomness protocol.

Input: security parameter n,
a distribution $\mathcal{D}$ over circuits on n qubits,
the threshold constant $b \in [1,2]$,
the number of samples $k = O(n^2)$ per iteration,
the number of rounds m, and
the fraction $\gamma = O((\log n)/m)$ of circuit updates.

Protocol:

(1) For $i = 1, \ldots, m$, run the following steps:
   (a) The verifier samples $T_i \sim \mathrm{Bernoulli}(\gamma)$. If $T_{i-1} = 1$ (when $i > 1$) or $i = 1$, the device samples $C_i \sim \mathcal{D}$. Otherwise, the device sets $C_i = C_{i-1}$. The verifier sends C_i to the device (and keeps T_i secret).
   (b) The device returns k samples $d_i = (z_1, \ldots, z_k)$.
   (c) If $T_i = 1$, the verifier sets
$$W_i = \delta\left[\frac{1}{k}\sum_{i=1}^{k} p_C(z_i) \geq \frac{b}{N} \wedge E_i\right]. \tag{7}$$
   where $E_i = 0$ if there exist distinct $\ell, \ell' \in \{j : C_j = C_i\}$ such that the samples $d_\ell = (z_{\ell 1}, \ldots, z_{\ell k})$ and $d_{\ell'} = (z_{\ell' 1}, \ldots, z_{\ell' k})$ are not all distinct. (This check is used to prevent the device repeats responses for any two rounds using the same challenge circuit.) If $T_i = 0$, the verifier sets $W_i = \perp$.
(2) Let $t = |\{i : T_i = 1\}|$ be the number of test rounds. The verifier computes
$$W = \sum_{i:T_i=1} W_i. \tag{8}$$
   If $W \geq 0.99t$, then the verifier accepts and outputs $(d_1, \ldots, d_m)$ to the quantum-proof randomness extractor.

Figure 1: The entropy accumulation protocol based on LLHA.

To summarize the results presented in this section, we give the description of the protocol in Figure 1.

2.2 Entangled Adversary and Ideal Measurements

In the previous section, we assumed that an attacker, Eve, trying to predict the quantum computer's outputs had no preshared entanglement with the quantum computer. Now we relax that assumption.

To build intuition, we start with the special case where the quantum computer performs an ideal measurement—i.e., it "just" applies C to n qubits, followed by a measurement in the standard basis. The "only" problem is that the qubits might not start in the desired initial state $|0^n\rangle$, but rather in some arbitrary state entangled with Eve's qubits.

We define the following idealized score, called b-XHOG($\mathcal{U}$).

Problem 3 (b-XHOG($\mathcal{D}$) [29]). *For a distribution $\mathcal{D}$ over quantum circuits on n qubits, an algorithm $\mathcal{A}$ given access to $C \sim \mathcal{D}$ is said to*

solve b-XHOG if it outputs a sample z such that

$$\mathop{\mathbb{E}}_{C\sim\mathcal{D}}\left[\mathop{\mathbb{E}}_{z\sim\mathcal{A}^C}[p_C(z)]\right] \geq \frac{b}{N}. \tag{9}$$

For an algorithm $\mathcal{A}$ which is given access to C and outputs z, we define the "XHOG score" of $\mathcal{A}$ to be the value $\mathbb{E}_C[\mathbb{E}_{z\sim\mathcal{A}^C}[p_C(z)]]$. This score was first considered by Kretschmer for showing a Tsirelson bound for random circuit sampling in the oracle model [29]. Recall that with the CHSH game, a violation of the classical bound 3/4 implies certified randomness. Interestingly, our result may be interpreted as certified randomness from a violation of the classical XHOG score.

We will proceed with our analysis with b-XHOG first, and show a von Neumann entropy lower bound $\Omega(\delta n)$ when the device solves $(1+\delta)$-XHOG. First, the problem itself is linear in the device's output distribution: that is, for two devices $\mathcal{A}$ with score $s_{\mathcal{A}}$ and $\mathcal{B}$ with score $s_{\mathcal{B}}$, a third device that runs $\mathcal{A}$ with probability p and $\mathcal{B}$ with probability $(1-p)$ has score $p\cdot s_{\mathcal{A}} + (1-p)\cdot s_{\mathcal{B}}$. The linearity condition coincides with the score calculation from violation of Bell's inequality.

More concretely, recall that to establish certified randomness from a violation of Bell's inequality, two devices $\mathcal{A}$ and $\mathcal{B}$ are asked to play, say, the CHSH game: the verifier sends two questions x, y to the devices and collecting the responses a, b. The verifier sets the score to 1 if $x\wedge y = a\oplus b$ and 0 otherwise, and the expectation of the score is defined as

$$\omega = \mathop{\mathbb{E}}_{x,y\sim\{0,1\}}\left[\mathop{\mathbb{E}}_{a,b\sim\mathcal{A}^x\otimes\mathcal{B}^y(\rho)}[\delta[x\wedge y = a\oplus b]]\right]. \tag{10}$$

It is well-known that the best achievable expectation is $\omega = \cos^2(\pi/8)$. For certified randomness, it was further shown that when the expectation $\omega \geq \cos^2(\pi/8+\varepsilon)$, the output of $\mathcal{A}$ has von Neumann entropy lower-bounded by $1-h(\sin 4\varepsilon)\approx 1-O(\varepsilon)$, where $h(x) := -x\log x - (1-x)\log(1-x)$ is the binary entropy function [8]. Like the XHOG score, here the score ω can be exactly computed only by taking infinitely many samples from the same devices. With a finite number of samples, we can only approximate the score.

Proving a conditional min-entropy lower bound from sample statistics, in an m-round sequential process, amounts to the problem of entropy accumulation. An Entropy Accumulation Theorem (EAT) for certified randomness is usually stated as follows: In an m-round sequential process, the verifier randomly selects $O(\log m)$ rounds to get an approximation of the score. If the approximation is sufficiently close to $\cos^2(\pi/8)$, the number of extractable random bits is at least $\Omega(m)$ times the von Neumann entropy lower bound established in a single-round analysis.

Without loss of generality, let the entanglement shared between the device and Eve be a pure state $|\psi\rangle$. For every state ρ_{ZE} classical on Z, we show that the conditional von Neumann entropy $H(Z|CE)_\rho \geq H(Z|C)_\rho - \chi(Z:CE)_\rho$, where χ is the Holevo quantity. To see why they they must use weak entanglement, we can write $|\psi\rangle := \sum_x \alpha_x|\psi_x\rangle|\phi_x\rangle$ for orthonormal bases $\{|\psi_x\rangle\}$ and $\{|\phi_x\rangle\}$ in the Schmidt decomposition. We show that the device solving b-XHOG for $b\geq 1+\delta$, the amplitude α_x must concentrate at a single compoment, say x^*, such that $|\alpha_{x^*}|^2 \geq \delta$. By the observation that the Holevo quantity equals the entanglement entropy, we establish an upper bound $O((1-\delta)n)$ on the Holevo quantity.

In this simplified setting, we have already seen that if the device has a large XHOG score, then they must use weak entanglement to pass the verification. However, the entire analysis relies on the assumption that the device must perform the ideal measurement.

2.3 A Fully General Device

Next, we consider the setting in which the adversary may share arbitrary entanglement with Eve. We give an unconditional proof for certified randomness in the random oracle model.

Instead of setting a binary-valued score for each question-answer pair as in Section 2.1, for question C and response z, the score is set to $p_C(z)$. Then, in a single round analysis, we first show that if the device passes $(1+\delta)$-XHOG score, then the von Neumann entropy of the joint state is at least $\Omega(n)$.

Theorem 2.4 (Single-round analysis, informal). *Every device $\mathcal{A}$ that on input the first system of a bipartite state ρ_{DE} and given oracle access to a Haar random C, makes $T \leq 2^{n/7}$ queries and solves $(1+\delta)$-XHOG must output a state $2^{-\Omega(n)}$-close to a state ψ_{ZE} classical on system Z such that*

$$H(Z|CE)_\psi \geq 0.99\delta n - O(\log T). \tag{11}$$

Furthermore, there is a single-query device that solves $(2-2^{-n})$-XHOG.

To prove Theorem 2.4, the key observation is that one can approximate, to diamond distance $2^{-\Omega(n)}$, any device $\mathcal{A}$ making T queries to a Haar random C by another device $\mathcal{F}$ which does not make any queries, but is given k samples $z_1,\ldots,z_k \sim P_C$ for $k = T^2\cdot 2^{O(n)}$. For each $\mathcal{A}$, we call the associated device $\mathcal{F}$ the simplified device. With probability $1-O(k^2/N)$, these samples does not contain any collision. In this event (no collision occurring), $\mathcal{F}$ solves $(1+\delta)$-XHOG implies that $\mathcal{F}$ outputs $z\in S = \{z_1,\ldots,z_k\}$ with probability at least $\delta - o(1)$. Intuitively, this robustly certifies that $\mathcal{A}$'s output must be ε-close to a strategy where the output is prepared by sampling from C for k times and choosing one of the samples.

From this point of view, a simplified device solving $(1+\delta)$-XHOG is equivalent to winning the following game with probability at least $\delta - o(1)$: Given k independent samples S from P_C, outputs a string $z\in S$. Though the game looks quite trivial, it yields a sharp lower bound of the von Neumann min-entropy. By the no-communication theorem, Eve, even if she learns P_C, has no information about the samples given to $\mathcal{F}$, but Eve can potentially control the output distribution when the device sees a particular set of samples. Since with high probability over C, P_C has min-entropy $n - O(\log n)$, we show that averaging over any distribution supported on these samples, the resulting distribution has von Neumann entropy at least $0.99\delta n - O(\log T)$.

Our lower bound in Theorem 2.4 is close to optimal. Consider a device which samples from P_C with probability δ and outputs a uniform string obtained by performing a standard basis measurement on EPR pairs shared with Eve with probability $1-\delta$. In the former event, the device solves b-XHOG for $b\approx 2$, whereas in the latter,

the output is a uniformly random string, which solves 1-XHOG. Thus, by linearity, the device solves $(1+\delta)$-XHOG. The output joint classical-quantum state from the device is a probabilistic mixture of the two states. Moreover, with overwhelming probability over choices of C, the Shannon entropy of P_C is $n - O(\log n)$. Then by the concavity of von Neumann entropy, the output has conditional von Neumann entropy $\delta n - o(n)$.

To accumulate the entropy, we give a sequential process which is very similar to the one introduced in Section 2.1: The verifier samples $t = O(\log n)$ different circuits, and asks for at least k samples for each circuit. Upon receiving the samples, the verifier chooses k random samples for each circuit, and checks if the device passes $\mathrm{LXEB}_{1+\delta,k}$ for a constant δ for 99% of the circuits. We show that the accumulation process certifies $\Omega(\delta mn)$ bits.

Theorem 2.5 (Entropy accumulation from a general device, informal). *For integer $k = \Omega(n^2)$, $m = \Omega(k \log n)$, there exists an entropy accumulation protocol taking m samples such that conditioned on the event Ω of non-aborting,*

$$H_{\min}(Z|CE)_{\rho|\Omega} \geq n\Big(0.99\delta m - O(\sqrt{m})\Big) \tag{12}$$

for devices solving $LXEB_{1+\delta,k}$, where ρ is the output state, Z is the samples from the device, C is the circuits and E is the information held by Eve.

More concretely, taking $\delta = 0.1$, this bound is $n \cdot \big(0.099m - O(\sqrt{m})\big)$ by taking m samples from the device. We note that this bound is seemingly weaker than the bound in Theorem 2.3, but the number of samples is m (instead of km as in Theorem 2.3). The minimal sample complexities in the protocols are no different—both are $\Omega(n^2 \log n)$—for a perfect device to pass the verification with overwhelming probability. For technical reasons, in the latter protocol, k samples for each verification are randomly chosen (from all samples sent by the device corresponding to the same challenge circuit) and received sequentially. In contrast, in the former protocol, in each round the device is asked to send k samples, and the verifier checks one round for each circuit.

The above analysis lead to an entropy accumulation protocol, described in Figure 2.

We also note that Theorem 2.3 and Theorem 2.5 are incomparable results. In particular, the security analysis for Theorem 2.5 heavily relies on the model in which the device is given access to the circuit and the distribution (the Haar measure) over circuits. In contrast, the security analysis based on LLHA may still hold when the device is given access to a description of circuits sampled from other distributions. We leave it as an open question whether there exists a hardness assumption under which linear cross-entropy benchmarking certifies min-entropy against an entangling adversary in the plain model.

2.4 Entropy Accumulation

The proof of Theorem 2.5 is based on the entropy accumulation theorem (EAT) by Dupuis, Fawzi and Renner [22], with modifications explained as follows. Let f be an affine function, called the min-tradeoff function, such that in a single-round analysis, one can show that $H(Z|E)_\rho \geq f(q)$ for distribution $q = (p, 1-p)$ and any

---

Input: Security parameter n,
the number of rounds m,
the score parameter $\delta \in [0, 1]$,
the fraction of circuit updates $\gamma = O((\log n)/m)$, and
the fraction of test rounds $\eta = O((n^2 \log n)/m)$.

Protocol:

(1) For $i = 1, \ldots, m$, run the following steps:
   (a) The verifier samples $T_i \sim \mathrm{Bernoulli}(\gamma)$, and $F_i \sim \mathrm{Bernoulli}(\eta)$. If $T_{i-1} = 1$ or $i = 1$, the device chooses a fresh circuit $C_i \sim \mathrm{Haar}(N)$. Otherwise, if $T_{i-1} = 0$ and $i > 1$, then the device sets $C_i = C_{i-1}$ to be the circuit used in the previous round. The verifier sends C_i to the device.
   (b) The prover returns a sample z_i.
(2) Let the number of epoches, i.e., the set of consecutive rounds i such that the same circuit $C_i = C$ is used, be t. For each epoch E_j, let $t_j = |\{i \in E_j : F_i = 1\}|$ denote the number of test rounds in this epoch. The verifier rejects if there exists a pair of collisions that correspond to the same circuit; otherwise it computes

$$s_j = \frac{1}{t_j} \sum_{i \in E_j : T_i = 1} p_{C_i}(z_i). \tag{13}$$

If $\frac{1}{t}\sum_{j=1}^{t} \delta[s_j \geq (1+\delta)/N] \geq 0.99$, then the verifier accepts and outputs $(z_1, \ldots, z_m)$ to the quantum-proof randomness extractor.

Figure 2: The entropy accumulation protocol.

state ρ whose acceptance probability is p. In an m-round sequential process, the verifier checks γm rounds (called test rounds) by computing the decision bits from the samples, and computes an approximate distribution $\tilde{q} = (\tilde{p}, 1-\tilde{p})$. The min-entropy round across the m rounds is then $m \cdot f(\tilde{q}) - O(\sqrt{m})$. Thus an EAT reduces a multi-round analysis to a lower bound on the single-round von Neumann entropy.

Since we adopt the b-XHOG score for a bound of the von Neumann entropy in a single round analysis, the score obtained from the test rounds is no longer computed from binary random variables. Thus we define a new min-tradeoff function f' which maps the *score* to a lower bound of the von Neumann entropy. Then we show that if an approximation of the score, defined as the average of $p_{C_i}(z_i)$ is more than s, then the accumulated entropy is at least $m \cdot f'(s) - O(\sqrt{m})$.

The entropy accumulation procedure allows for *spot checking*, that is, in the m-round process, instead of computing $p_{C_i}(z_i)$ for every round $i \in [m]$, the verifier only computes $p_{C_i}(z_i)$ for a subset of indices i of size $O(n^2 \log n)$. In more details, the verifier changes the circuits for $O(\log n)$ times, and in each epoch the verifier computes the average of $k = O(n^2)$ samples. The number of test rounds is set for the device that takes i.i.d. samples from p_C on each circuit C to pass the verification with overwhelming probability. By Hoeffding's inequality, a device that samples from p_C outputs k

samples whose average score is concentrated above $2 - O(1)$ for a typical C with overwhelming probability. If the verifier passes $\mathrm{LXEB}_{b,k}$ for the epoches of a sufficiently large fraction $\Omega(1)$, the average is above $1 + \Omega(1)$ with overwhelming probability, and by the entropy accumulation theorem, the conditional min-entropy is $\Omega(nm)$.

2.5 Pseudorandomness and Statistical Zero Knowledge

The protocols for certified randomness rely on perfect randomness for generating the challenge circuit. However, by a counting argument, the challenge space is doubly exponentially large, and it requires exponentially many random bits to compute a truly random circuit. To produce a net gain in randomness, we must rely an efficiently computable function which uses polynomially many random bits and generates pseudorandomness with security level sufficient for our purpose.

However, the standard notion of pseudorandom functions (PRFs) against quantum polynomial-time adversaries does not seem to be sufficient, since it only guarantees the output of the device is *pseudorandom*! Thus for certified randomness, we require a stronger pseudorandom function, when a truly random function is replaced with which, the output remains *statistically indistinguishable* from the uniform distribution.

To provide such a security guarantee, we construct a pseudorandom function indistinguishable from a truly random function for any QSZK protocols. To see why such a security level is sufficient, we recall facts about the class QSZK which consists of promise problems that admit a quantum statistical zero-knowledge protocol. A QSZK protocol is one that consists of a proof system, i.e., a quantum polynomial-time verifier and an unbounded prover, and an efficient quantum simulator which simulates the interaction of the proof system without access to a witness.

Watrous showed that QSZK has a natural complete problem called the *quantum state distinguishability problem* (QSD) [45]. In this problem, the instance is a tuple of two efficiently computable quantum circuits Q_0, Q_1. For $\alpha \in (0, 1]$, the verifier is challenged to determine the trace distance $\|\rho_0 - \rho_1\|_{\mathrm{tr}}$ is at least α, or at most α^2, where ρ_b is a marginal state obtained by computing Q_b for $b \in \{0, 1\}$. It is known for this class, there is an amplification procedure, and therefore the gap can be made exponentially close to 1 [45]. More recently, Menda and Watrous showed that relative to a random oracle, $\mathsf{UP} \not\subset \mathsf{QSZK}$ [32]. Ben-David and Kothari defined a query measure on statistical zero-knowledge proof, and showed that the positive-weighted adversary method can only prove suboptimal lower bounds for certain problems [12].

For certified randomness, we define the QSZK-distinguishability between two distributions over functions, and a similar definition can be extended to distributions over unitaries.

Definition 2.6 (QSZK-distinguishability, informal)**.** *Two distributions $\mathcal{D}_0, \mathcal{D}_1$ over functions are said to be* QSZK*-distinguishable if there exist a pair of algorithms $\mathcal{A}, \mathcal{B}$ such that the averaged trace distance between $\mathcal{A}^F$ and $\mathcal{B}^F$'s output states has non-neglgigible difference between $F \sim \mathcal{D}_0$ and $F \sim \mathcal{D}_1$. The distributions are said to be* QSZK*-indistinguishable if no such algorithms exist.*

A QSZK-secure pseudorandom function is defined as one QSZK-indistinguishable from a random function. We propose an assumption, called pseudorandom function assumption (PRFA), that there exists a QSZK-secure pseudorandom function. We justify the assumptions are valid by giving a construction for algorithms given oracle access to the function.

Theorem 2.7 (Pseudorandom functions, informal). *There exists a* QSZK*-secure pseudorandom function with key length $O(n)$ relative to a random oracle.*

Similarly, we say a pseudorandom unitary is QSZK-secure if it is QSZK-indistinguishable from a random unitary. We propose a similar assumption, called pseudorandom unitary assumption (PRUA), that there exists a QSZK-secure pseudorandom unitary, and prove the existence relative to an oracle with key length $O(n)$.

Under these assumptions, when replacing a random circuit with a pseudorandom one, the output remains statistical indistinguishable from a uniform distribution, conditioned on Eve's side information. To see why, recall that De, Portmann, Vidick, and Renner [21] showed that Trivesan's randomness extractor [43] is quantum-proof. That is, the output from the randomness extractor together with Eve's side information is a quantum state ρ statistically indistinguishable from $\sigma \otimes \rho_E$, where σ is a maximally mixed state and ρ_E is the marginal state held by Eve. If the device given a pseudorandom circuit outputs a quantum state that changes the distance by a non-negligible amount from $\sigma \otimes \rho_E$, then such a device implies a QSZK protocol that distinguishes a pseudorandom circuit from a random one.

To see there is a net gain in randomness, the protocol samples $O(\log n)$ pseudorandom circuits, each of which takes $O(n)$ random bits for the keys of the pseudorandom function, and finally it produces $\Omega(mn)$ random bits. For $m = \mathrm{poly}(n)$, we have a polynomial expansion.

While we do not know whether a weaker assumption can work for certified randomness, the security level seems necessary against an entangling adversary. Indeed, if there is no quantum side information, then all we need is to use a pseudorandom circuit against adversaries solving the *statistical difference from uniform* problem. In the purely classical setting, the problem is known to be complete for NISZK, a subclass of SZK consisting of problems that admits a non-interactive statistical zero-knowledge protocol [24]. However, in the presence of quantum side information, an unbounded Eve can prepare any ρ_E, and security against QSZK seems necessary.

ACKNOWLEDGMENTS

We thank Fernando Brandão, Alex Halderman, William Kretschmer, John Martinis, Carl Miller, Ron Peled, René Peralta, Or Sattath, Thomas Vidick, and David Zuckerman for helpful discussions. SA is supported by a Vannevar Bush Fellowship from the US Department of Defense, the Berkeley NSF-QLCI CIQC Center, a Simons Investigator Award, and the Simons "It from Qubit" collaboration. SHH is supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Quantum Systems Accelerator.

REFERENCES

[1] Scott Aaronson. 2018. Certified randomness from quantum supremacy. *Talk at Google Quantum Symposium, Venice Beach, CA, May 16, 2018* (2018). https://www.scottaaronson.com/talks/certrand-huji.ppt.

[2] Scott Aaronson and Andris Ambainis. 2014. The Need for Structure in Quantum Speedups. *Theory of Computing* 10, 1 (2014), 133–166. https://doi.org/10.4086/toc.2014.v010a006

[3] Scott Aaronson and Alex Arkhipov. 2011. The computational complexity of linear optics. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 333–342. https://doi.org/10.1145/1993636.1993682

[4] Scott Aaronson and Alex Arkhipov. 2014. Bosonsampling is far from uniform. *Quantum Information & Computation* 14, 15-16 (2014), 1383–1423.

[5] Scott Aaronson and Lijie Chen. 2017. Complexity-Theoretic Foundations of Quantum Supremacy Experiments. In *32nd Computational Complexity Conference (CCC 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. https://doi.org/10.4230/LIPIcs.CCC.2017.22

[6] Scott Aaronson and Andrew Drucker. 2010. A full characterization of quantum advice. In *Proceedings of the forty-second ACM symposium on Theory of computing*. 131–140. https://doi.org/10.1145/1806689.1806710

[7] Scott Aaronson and Shih-Han Hung. 2023. Certified randomness from quantum supremacy. *arXiv preprint arXiv:2303.01625* (2023). https://doi.org/10.48550/arXiv.2303.01625

[8] Rotem Arnon-Friedman, Frédéric Dupuis, Omar Fawzi, Renato Renner, and Thomas Vidick. 2018. Practical device-independent quantum cryptography via entropy accumulation. *Nature communications* 9, 1 (2018), 1–11. https://doi.org/10.1038/s41467-017-02307-4

[9] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandão, David A Buell, et al. 2019. Quantum supremacy using a programmable superconducting processor. *Nature* 574, 7779 (2019), 505–510. https://doi.org/10.1038/s41586-019-1666-5

[10] Boaz Barak, Chi-Ning Chou, and Xun Gao. 2021. Spoofing Linear Cross-Entropy Benchmarking in Shallow Quantum Circuits. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.ITCS.2021.30

[11] Roozbeh Bassirian, Adam Bouland, Bill Fefferman, Sam Gunn, and Avishay Tal. 2021. On Certified Randomness from Quantum Advantage Experiments. *arXiv preprint arXiv:2111.14846* (2021). https://doi.org/10.48550/arXiv.2111.14846

[12] Shalev Ben-David and Robin Kothari. 2019. Quantum Distinguishing Complexity, Zero-Error Algorithms, and Statistical Zero Knowledge. In *14th Conference on the Theory of Quantum Computation, Communication and Cryptography*, Vol. 21. 24. https://doi.org/10.4230/LIPIcs.TQC.2019.2

[13] Peter Bierhorst, Emanuel Knill, Scott Glancy, Yanbao Zhang, Alan Mink, Stephen Jordan, Andrea Rommal, Yi-Kai Liu, Bradley Christensen, Sae Woo Nam, et al. 2018. Experimentally generated randomness certified by the impossibility of superluminal signals. *Nature* 556, 7700 (2018), 223–226. https://doi.org/10.1038/s41586-018-0019-0

[14] Joseph Bonneau, Jeremy Clark, and Steven Goldfeder. 2015. On Bitcoin as a public randomness source. Cryptology ePrint Archive, Paper 2015/1015. https://eprint.iacr.org/2015/1015

[15] Zvika Brakerski, Paul Christiano, Urmila Mahadev, Umesh Vazirani, and Thomas Vidick. 2018. A cryptographic test of quantumness and certifiable randomness from a single quantum device. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 320–331. https://doi.org/10.1145/3441309

[16] Fernando GSL Brandão, Aram W. Harrow, and Michał Horodecki. 2016. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics* 346, 2 (2016), 397–434. https://doi.org/10.1007/s00220-016-2706-8

[17] Luís TAN Brandão and René Peralta. 2020. Notes on Interrogating Random Quantum Circuits. (2020). https://doi.org/10.13140/RG.2.2.24562.94400 NIST White Paper.

[18] Michael J Bremner, Richard Jozsa, and Dan J Shepherd. 2011. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 467, 2126 (2011), 459–472. https://doi.org/10.1098/rspa.2010.0301

[19] Roger Colbeck. 2009. Quantum And Relativistic Protocols For Secure Multi-Party Computation. *Ph. D. Thesis* (2009).

[20] Matthew Coudron and Henry Yuen. 2014. Infinite randomness expansion with a constant number of devices. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. 427–436. https://doi.org/10.1145/2591796.2591873

[21] Anindya De, Christopher Portmann, Thomas Vidick, and Renato Renner. 2012. Trevisan's extractor in the presence of quantum side information. *SIAM J. Comput.* 41, 4 (2012), 915–940. https://doi.org/10.1137/100813683

[22] Frederic Dupuis, Omar Fawzi, and Renato Renner. 2020. Entropy accumulation. *Communications in Mathematical Physics* 379, 3 (2020), 867–913. https://doi.org/10.1007/s00220-020-03839-5

[23] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. 2014. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028* (2014). https://doi.org/10.48550/arXiv.1411.4028

[24] Oded Goldreich, Amit Sahai, and Salil Vadhan. 1999. Can statistical zero knowledge be made non-interactive? or on the relationship of SZK and NISZK. In *Annual International Cryptology Conference*. Springer, 467–484. https://doi.org/10.1007/3-540-48405-1_30

[25] Jonas Haferkamp. 2022. Random quantum circuits are approximate unitary t-designs in depth $O(nt^{5+o(1)})$. *Quantum* 6 (2022), 795. https://doi.org/10.22331/q-2022-09-08-795

[26] Joonsuk Huh, Gian Giacomo Guerreschi, Borja Peropadre, Jarrod R McClean, and Alán Aspuru-Guzik. 2015. Boson sampling for molecular vibronic spectra. *Nature Photonics* 9, 9 (2015), 615–620. https://doi.org/10.1038/nphoton.2015.153

[27] Gregory D Kahanamoku-Meyer, Soonwon Choi, Umesh V Vazirani, and Norman Y Yao. 2022. Classically verifiable quantum advantage from a computational Bell test. *Nature Physics* 18, 8 (2022), 918–924. https://doi.org/10.1038/s41567-022-01643-7

[28] John Kelsey, Luís TAN Brandão, Rene Peralta, and Harold Booth. 2019. *A reference for randomness beacons: Format and protocol version 2*. Technical Report. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8213-draft

[29] William Kretschmer. 2021. The quantum supremacy Tsirelson inequality. *Quantum* 5 (2021), 560. https://doi.org/10.22331/q-2021-10-07-560

[30] Lars S Madsen, Fabian Laudenbach, Mohsen Falamarzi Askarani, Fabien Rortais, Trevor Vincent, Jacob FF Bulmer, Filippo M Miatto, Leonhard Neuhaus, Lukas G Helt, Matthew J Collins, et al. 2022. Quantum computational advantage with a programmable photonic processor. *Nature* 606, 7912 (2022), 75–81. https://doi.org/10.1038/s41586-022-04725-x

[31] Urmila Mahadev, Umesh Vazirani, and Thomas Vidick. 2022. Efficient Certifiable Randomness from a Single Quantum Device. *arXiv preprint arXiv:2204.11353* (2022). https://doi.org/10.48550/arXiv.2204.11353

[32] Sanketh Menda and John Watrous. 2018. Oracle separations for quantum statistical zero-knowledge. *arXiv preprint arXiv:1801.08967* (2018). https://doi.org/10.48550/arXiv.1801.08967

[33] Carl A Miller and Yaoyun Shi. 2016. Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. *Journal of the ACM (JACM)* 63, 4 (2016), 1–63. https://doi.org/10.1145/2885493

[34] Carl A Miller and Yaoyun Shi. 2017. Universal security for randomness expansion from the spot-checking protocol. *SIAM J. Comput.* 46, 4 (2017), 1304–1335. https://doi.org/10.1137/15M1044333

[35] Feng Pan, Keyang Chen, and Pan Zhang. 2022. Solving the sampling problem of the sycamore quantum circuits. *Physical Review Letters* 129, 9 (2022), 090502. https://doi.org/10.1103/PhysRevLett.129.090502

[36] Stefano Pironio, Antonio Acín, Serge Massar, A Boyer de La Giroday, Dzmitry N Matsukevich, Peter Maunz, Steven Olmschenk, David Hayes, Le Luo, T Andrew Manning, et al. 2010. Random numbers certified by Bell's theorem. *Nature* 464, 7291 (2010), 1021–1024. https://doi.org/10.1038/nature09008

[37] John Preskill. 2012. Quantum computing and the entanglement frontier. *arXiv preprint arXiv:1203.5813* (2012). https://doi.org/10.48550/arXiv.1203.5813

[38] John Preskill. 2018. Quantum computing in the NISQ era and beyond. *Quantum* 2 (2018), 79. https://doi.org/10.22331/q-2018-08-06-79

[39] Ran Raz and Avishay Tal. 2022. Oracle separation of BQP and PH. *ACM Journal of the ACM (JACM)* 69, 4 (2022), 1–21. https://doi.org/10.1145/3313276.3316315

[40] Markus Reiher, Nathan Wiebe, Krysta M Svore, Dave Wecker, and Matthias Troyer. 2017. Elucidating reaction mechanisms on quantum computers. *Proceedings of the national academy of sciences* 114, 29 (2017), 7555–7560. https://doi.org/10.1073/pnas.1619152114

[41] Maria Schuld, Kamil Brádler, Robert Israel, Daiqin Su, and Brajesh Gupt. 2020. Measuring the similarity of graphs with a Gaussian boson sampler. *Physical Review A* 101, 3 (2020), 032314. https://doi.org/10.1103/PhysRevA.101.032314

[42] Alexander A Sherstov. 2011. Strong direct product theorems for quantum communication and query complexity. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 41–50. https://doi.org/10.1145/1993636.1993643

[43] Luca Trevisan. 2001. Extractors and pseudorandom generators. *J. ACM* 48, 4 (2001), 860–879. https://doi.org/10.1145/502090.502099

[44] Umesh Vazirani and Thomas Vidick. 2012. Certifiable quantum dice: or, true random number generation secure against quantum adversaries. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 61–76. https://doi.org/10.1145/2213977.2213984

[45] John Watrous. 2002. Limits on the power of quantum statistical zero-knowledge. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.* IEEE, 459–468. https://doi.org/10.1109/SFCS.2002.1181970

[46] Yulin Wu, Wan-Su Bao, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, et al. 2021. Strong quantum computational advantage using a superconducting quantum processor. *Physical review letters* 127, 18 (2021), 180501. https://doi.org/10.1103/PhysRevLett.127.180501

[47] Takashi Yamakawa and Mark Zhandry. 2022. Verifiable Quantum Advantage without Structure. *arXiv preprint arXiv:2204.02063* (2022). https://doi.org/10.

1109/FOCS54457.2022.00014

[48] Han-Sen Zhong, Hui Wang, Yu-Hao Deng, Ming-Cheng Chen, Li-Chao Peng, Yi-Han Luo, Jian Qin, Dian Wu, Xing Ding, Yi Hu, et al. 2020. Quantum computational advantage using photons. *Science* 370, 6523 (2020), 1460–1463. https://doi.org/10.1126/science.abe8770

Received 2022-11-07; accepted 2023-02-06