2018 IEEE 59th Annual Symposium on Foundations of Computer Science

A Cryptographic Test of Quantumness and Certifiable Randomness from a Single Quantum Device

Zvika Brakerski*, Paul Christiano†, Urmila Mahadev‡, Umesh Vazirani‡ and Thomas Vidick§
*Weizmann Institute of Science, Israel
Email: zvika.brakerski@weizmann.ac.il
†OpenAI, USA
Work performed while at UC Berkeley
‡UC Berkeley, USA
Email: {mahadev, vazirani}@cs.berkeley.edu
§California Institute of Technology, USA
Email: vidick@cms.caltech.edu

***Abstract*—We give a protocol for producing certifiable randomness from a single untrusted quantum device that is polynomial-time bounded. The randomness is certified to be statistically close to uniform from the point of view of any computationally unbounded quantum adversary, that may share entanglement with the quantum device. The protocol relies on the existence of post-quantum secure trapdoor claw-free functions, and introduces a new primitive for constraining the power of an untrusted quantum device. We then show how to construct this primitive based on the hardness of the learning with errors (LWE) problem.**

The randomness protocol can also be used as the basis for an efficiently verifiable "quantum supremacy" proposal, thus answering an outstanding challenge in the field. experiments.

***Keywords*-randomness; quantum information; cryptography**

I. Introduction

In this paper we propose solutions to two basic tasks: how to generate *certifiably random* strings from a *single untrusted* quantum device (also referred to as a prover), and efficient verification of quantum supremacy. The setting we consider is one where the quantum device is polynomial-time bounded but untrusted, and the verifier is entirely classical and also polynomial-time bounded. The peculiarity of this setting is that it allows the verifier to leverage post-quantum cryptography, i.e. the existence of cryptographic primitives that can be implemented efficiently on a classical computer but that cannot be broken by any efficient quantum computer.

There has been considerable research into certifiable quantum random number generation [Col06], [PAM+10], [FGS11], [PM11], [VV11], [MS16], [BKG+18]. However, all prior works providing verifiable guarantees have focused on the setting where there are multiple quantum devices that share entanglement, and where the randomness certification relies crucially on the violation of a Bell inequality. By contrast, in the setting that we study there is a single polynomial-time bounded quantum device, and the guarantee we seek is that provided the device is unable to break the post-quantum cryptographic assumption during the execution of the protocol, then the output of the protocol must be statistically indistinguishable from a uniformly random sequence of bits, to any computationally *unbounded* adversary that may share prior entanglement with the computationally bounded device. This information-theoretic guarantee, the same guarantee as that offered in the aforementioned works [VV11], [MS16], is stronger than computational pseudo-randomness (that is easily achievable under standard cryptographic assumptions, since the verifier starts with a short random seed).

The specific cryptographic primitive we rely on is the existence of a post-quantum secure trapdoor claw-free (in short, TCF) family of functions $f : \{0,1\}^n \to \{0,1\}^m$, the post-quantum analog of a notion introduced by Goldwasser, Micali and Rivest in the context of digital signatures [GMR84]. A TCF is a 2-to-1 function f that satisfies the following properties: $f(x)$ is efficiently computable on a classical computer, and if $f(x) = y$, then there is a unique $x' \neq x$ such that $f(x') = y$. Moreover, with knowledge of a secret trapdoor it is possible to efficiently (classically) compute x and x' from y, but without the trapdoor there is no efficient quantum algorithm that can compute such a triple (x, x', y), for any y.

By contrast, a quantum algorithm can simultaneously hold an image y, as well as a superposition $\frac{1}{\sqrt{2}}(|x\rangle + |x'\rangle)$ over two pre-images of y, simply by evaluating f on a uniform superposition over all inputs and measuring the image y. At this point, measuring the quantum state in the standard basis will yield a random pre-image, x or x'. This is not any stronger than a classical device could do, by first sampling a random x and then computing $y = f(x)$. However, the quantum device can do something

[0]A full version of this extended abstact can be found at [BCM+18].

2575-8454/18/$31.00 ©2018 IEEE
DOI 10.1109/FOCS.2018.00038

different from directly measuring a pre-image. Instead, the device can perform Fourier sampling (Hadamard transform followed by a measurement), which yields a random n-bit string $d : d \cdot (x \oplus x') = 0$, thereby revealing some joint information about both pre-images of y.

The ability to perform *either* of these tasks, identifying a pre-image or sampling an equation, but not both, seems to be a unique quantum capability. But of course a classical verifier cannot access the inner workings of the quantum device: from her viewpoint the device is an untrusted black box that outputs y, and then either a pre-image x, or an equation d. However, assuming the verifier knows the secret trapdoor, she can efficiently compute both pre-images x and x', and verify that indeed $d \cdot (x \oplus x') = 0$. At a high level, the consideration of a trapdoor family restores some symmetry between the quantum prover (the untrusted quantum device) and the classical verifier, by providing a primitive which allows the quantum capabilities of the prover to play a useful role while at the same time giving the classical verifier a handle, namely the ability to compute both x and x', that the prover does not have access to.

One might now conjecture that for a generic TCF (e.g. modeled as a random oracle), if the output of any efficient quantum device passes this test with non-negligible advantage over $\frac{1}{2}$, then the pair y, d returned in the equation test must have high min-entropy. Such a strong statement would immediately yield a randomness certification protocol. The difficulty in showing such a statement is that both y and d may be adaptively and adversarially chosen. While we do not know how to prove this, we are still able to design and analyze a simple protocol, that can be based on any TCF family with suitable cryptographic properties, and succeeds in certifiable randomness expansion from a single untrusted quantum (polynomial-time) device.

We will refer to protocol described above as a single round test, and our certified randomness expansion protocol will consist of a large number of such rounds. To facilitate the analysis of the protocol, we elaborate on a specific TCF construction in [Mah17a] based on the learning with errors problem (LWE). Specifically, we achieve a construction of TCF that has additional security guarantees, including an "adaptive hardcore bit" condition that is explained below. This condition helps in going beyond the analysis of a single round of our randomness-generation protocol (this analysis can be performed based on a more generic requirement for the TCF) to guarantee randomness accumulation across multiple rounds, a more stringent requirement. For clarity, in this paper we refer to the specific (relaxed) kind of TCF that we rely on as an NTCF, or post-quantum secure noisy trapdoor claw-free family. We give a construction of an NTCF that rests on the hardness of the Learning With Errors (LWE) problem, introduced by Regev [Reg05], with slightly super-polynomial noise ratio, against quantum polynomial-time attacks with nonuniform quantum advice (where the state of the art classical and quantum complexity scale exponentially with the dimension). This construction is similar to the one used in [Mah17a], albeit with some changes in parameters that allow us to prove the adaptive hardcore bit statement. Roughly, it asserts that it is computationally intractable to sample from any distribution on (y, x, d, b) such that $f(x) = y$ and $\Pr(b = d \cdot (x + x')) \geq 1/2 + \varepsilon$ for non-negligible ε.

We now note that the adaptive hardcore bit statement already has an interesting consequence: passing the single round test constitutes a proof of quantumness of the device. This is because there cannot be any efficient classical algorithm that can reliably answer a random challenge, "pre-image" or "equation", since such an algorithm could be "rewound" to simultaneously answer both challenges, thus contradicting the adaptive hardcore bit property. This has implications for an important milestone in the experimental realization of quantum computers, namely "quantum supremacy": a proof that an (untrusted) quantum computing device performs some computational task that cannot be solved classically without impractical resources. While quantum supremacy could be achieved, in principle, by demonstrating quantum factoring, the latter requires quantum resources well beyond the capability of near term experiments. Instead current proposals for quantum supremacy are based on sampling problems (see e.g. [HM17] for a recent survey). The major challenge for these proposals is verifying that the quantum computer did indeed sample from the desired probability distribution, and indeed all existing proposals rely on exponential time classical algorithms for verification. By contrast, our single round test provides a proof of quantumness and can be carried out by a classical verifier in polynomial time. This proposal for quantum supremacy seems promising from a practical viewpoint — indeed, even using off-the-shelf bounds for LWE-based cryptography suggests that a protocol providing 50 bits of security could be implemented with a quantum device of around 2000 qubits (see e.g. [LP11]). It would be worth exploring whether there are clever implementations of this scheme that can lead to a quantum supremacy protocol in the $200 - 500$ qubit range. Our protocol is robust to a device that only successfully answers the verifier's challenges a sufficiently large, but constant, fraction of the time; it would be interesting to explore whether such a device could be implemented without resorting to general fault-tolerance techniques.

Using the single round test as a building block for certified randomness expansion requires a deeper analysis. Let us focus on a single bit of information that the device outputs in the test: whether the pre-image is x or x' and whether the equation d satisfies $d \cdot (x + x') = 0$ or not. We are able to characterize the quantum state and quantum measurements for this task that are derived from the measurements of any efficient device that passes the single round test. Specifically,

we show that the state and measurements of the device are close to the following: the device starts with a qubit initialized to $|0\rangle$, which it measures in the standard basis when the challenge is $C = 0$ and in the Hadamard basis when the challenge is $C = 1$. Note that the fact that an efficient quantum device cannot break the cryptographic assumption has thus been translated into a characterization of the state and actions of the quantum device, which further implies that the output of the device in this single round must contain almost a bit of true (information theoretic) randomness.

Achieving certifiable randomness expansion requires an analysis of a protocol with a long sequence of single round tests, which we describe in the next section. We note that while assuming that LWE is hard for polynomial-time adversaries already implies polynomial randomness expansion, we can stretch the parameters further to achieve exponential randomness expansion by assuming that the problem remains hard even against quantum sub-exponential time adversaries, which is still consistent with the current state of the art.

Recently Aaronson [Aar18] discovered a different connection between quantum supremacy proposals and randomness generation Aaronson's scheme has the advantage of being as easy to implement as the supremacy proposal. Unfortunately, it also faces the same limitation in terms of verifiability as existing supremacy proposals, with the verification time scaling exponentially with the number of generated bits.

The idea of using a TCF as a basic primitive in interactions between an efficient quantum prover and a classical verifier has been further developed in recent work by Mahadev [Mah17b], giving the first construction of a quantum fully homomorphic encryption scheme with classical keys. In further follow-up work, Mahadev [Mah18] shows a remarkable use of a NTCF family with adaptive hardcore bit. Namely, that the NTCF can be use to certify that a prover measures a qubit in a prescribed basis (standard or Hadamard). This allows to achieve single prover *verifiability* for quantum computations using a purely classical verifier (but relying on computational assumptions).

Independently of this work, a construction of trapdoor one-way functions with second pre-image resistance based on LWE was recently introduced in [CCKW18], where it is used to achieve delegated computation in the weaker honest-but-curious model for the adversary (i.e. without soundness against provers not following the protocol). The family of functions considered in [CCKW18] is not sufficient for our purposes, as it lacks the essential adaptive hardcore bit property.

We believe that the technique of constraining the power of a quantum device using NTCFs, and with the adaptive hardcore bit put forth in this work promises to be a powerful tool for the field of untrusted quantum devices.

Overview of results and proof techniques

We provide a high level description of the main components of our protocol and the proof of security.

A single round of the protocol: test of quantumness: At the computational heart of our protocol is the single round test. In each round, the device is first required to output an element y in the range of NTCF f. The classical verifier then issues one of two challenges, $C = 0$ ("pre-image") or $C = 1$ ("equation"). The device responds by providing either a nontrivial $d : d \cdot (x \oplus x') = 0$ [1] or a valid pre-image for y, respectively. Both conditions can be efficiently checked by the verifier, the equation check requires the verifier to use the trapdoor for f. An honest quantum device can accomplish this by creating a uniform superposition on all x, computing f in superposition, and then measuring $y = f(x)$. This projects the x-register into the superposition $(1/\sqrt{2})|x\rangle + (1/\sqrt{2})|x'\rangle$, where x, x' are the preimages of y. Now measuring the x-register in the standard basis yields a random preimage x or x', and measuring it in the Hadamard basis yields a random equation $d : d \cdot (x \oplus x') = 0$.

The fact that passing the single round test constitutes a proof of quantumness of the device relies on the adaptive hardcore bit property, which we sketch here and explain in greater detail later. Roughly, it asserts that it is computationally intractable to sample from any distribution on (y, x, d, b) such that $f(x) = y$ and $\Pr(b = d \cdot (x + x')) \geq 1/2 + \varepsilon$ for non-negligible ε. This property implies that any efficient device that can reliably answer a random challenge (equation or pre-image) must be quantum. This is because any classical algorithm could be "rewound" to simultaneously answer both challenges, which is impossible by the adaptive hard core predicate property.

At the heart of certifiable randomness generation lies the claim that if an efficient device passes the single round test, then with high probability the state of the device and its measurements, up to some coarse-graining, are close to the following: it starts with a qubit initialized to $|0\rangle$, which it measures in the standard basis when the challenge is $C = 0$ and in the Hadamard basis when the challenge is $C = 1$. This characterization of the quantum state of the device and its measurements is where the computational bound on the device translates into true randomness rather than pseudorandomness.

To see how the proof proceeds, notice that in a single round the device must make one of two measurements: either a "pre-image" measurement, or an "equation" measurement. The "pre-image" measurement can be treated as a projection into one of two orthogonal subspaces corresponding to the two pre-images x, x' for the element y that the device has

[1] The exact condition for the equation is slightly more complicated, due to the formal definition of a NTCF; see Section IV for a complete description of the protocol.

returned to the verifier. The "equation" measurement can similarly be coarse-grained into a projection on one of two orthogonal subspaces, "valid" or "invalid", i.e. the subspace that corresponds to all measurement outcomes d such that $d \cdot (x \oplus x') = 0$, or the subspace associated with outcomes d such that $d \cdot (x \oplus x') = 1$.

Applying Jordan's lemma, it is possible to decompose the device's Hilbert space into a direct sum of one- or two-dimensional subspaces, such that within each two-dimensional subspace the "pre-image" and "equation" measurements each correspond to an orthonormal basis, such that the the two bases make a certain angle with each other. We argue that almost all angles must be very close to $\pi/4$. Indeed, whenever the angles are *not* near-maximally unbiased, it is possible to show that by considering the effect of performing the measurements in sequence, one can devise an "attack" on the NTCF of a kind that contradicts the adaptive hardcore bit property of the NTCF — informally, the attack can simultaneously produce a valid pre-image and a valid equation, with non-negligible advantage.

Outline of randomness generation protocol: The correctness of our protocol relies on the claim that in a single round, if an efficient quantum algorithm has the ability to generate a valid equation with probability sufficiently close to 1, then, if instead it is asked for a pre-image, this pre-image must be close to uniformly distributed over the two possibilities. Therefore, our protocol repeatedly asks for pre-images (to generate randomness), while inserting a few randomly located "equation" challenges, to test the device. Each time an "equation" challenge has been answered, we refresh the pseudorandom keys used for the NTCF. This is required to avoid a simple "attack" by the device, which would repeatedly use the same y, preimage x, and guessed equation d — succeeding in the protocol with probability $\frac{1}{2}$ without generating any randomness.

Let's call the sequence of rounds with a particular set of pseudorandom keys an epoch. Intuitively, we would like to claim that if the device passes all the equation challenges, then for most epochs and for most rounds within that epoch, the state of the device and its measurements must be (close to) as characterized above: it starts with a qubit initialized to $|0\rangle$, which it measures in the standard basis when the challenge is $C = 0$ and in the Hadamard basis when the challenge is $C = 1$. To show this we would like to claim that if the device passes all the equation challenges, for most such challenges it must produce a valid equation with probability close to 1. Since each equation challenge occurs at a random round in the epoch, it should follow from the adaptive hardcore bit property that the sequence of bits that the verifier extracts from the device's answers to "pre-image" challenges during that epoch must look statistically random. We give a martingale based argument to formalize this intuition, though the quantum setting makes it considerably more challenging than it appears at first sight.

There is however a bigger challenge to analyzing the protocol — we must show that the sequence that the verifier extracts from the device's answers to "pre-image" challenges must look statistically random even to an infinitely powerful quantum adversary, who may share an arbitrary entangled state with the quantum device. If we could assert that each round of the protocol is played with a qubit exactly in state $|0\rangle$, and measured in Hadamard basis on challenge $C = 1$, then this would preclude any entanglement with the quantum adversary, leading to an easy proof that the extracted sequence looks random to the adversary. Unfortunately the characterization of device's qubits leaves plenty of room for entanglement with the adversary — showing that such entanglement cannot leak too much information about the device's measurements was the major challenge in previous work on certified randomness through Bell inequality violations [VV11], [MS14], [AFDF$^+$18]. Our cryptographic setting presents a new difficulty, which is that in contrast to the two-player game Bell inequality violation scenarios, in our setting it is not *impossible* for a deterministic device to succeed in the test: it is merely *computationally hard* to do so. This prevents us from directly applying the results in [AFDF$^+$18], [MS14], and requires us to suitably modify their framework.

In terms of efficiency, for the specific LWE-based NTCF that we construct, our protocol can use as few as $\mathrm{poly}\log(N)$ bits of randomness to generate $O(N)$ bits that are statistically within negligible distance from uniform. However, this requires assuming that the underlying LWE assumption is hard even for sub-exponential size quantum circuits, with polynomial-size quantum advice (which is consistent with current knowledge). The more conservative assumption that our variant of LWE is only hard for polynomial size quantum circuits requires $O(N^\epsilon)$ bits of randomness for generating the NTCF, for any constant $\epsilon > 0$. The following is an informal description; see Theorem 8 for a more formal statement.

Theorem 1 (Informal)**.** *Let $\mathcal{F}$ be a NTCF family and λ a security parameter. Let $N = \Omega(\lambda^2)$ and assume the quantum hardness of solving lattice problems of dimension λ in time $\mathrm{poly}(N)$. There is an N-round protocol for the interaction between a classical polynomial-time verifier and a quantum polynomial-time device such that the protocol can be executed using $\mathrm{poly}(\log(N), \lambda)$ bits of randomness, and for any efficient device and side information E correlated with the device's initial state,*

$$H_\infty^{N\delta}(O|CE)_{\overline{\rho}} \geq (\xi - o(1))N \,.$$

Here ξ is a positive constant, δ is a negligible function of λ, and $\overline{\rho}$ is the final state of the classical output register O, the classical register C containing the verifier's messages to the device, and the side information E, restricted to transcripts that are accepted by the verifier in the protocol.

Sketch of the security analysis: We first describe the protocol in slightly more detail (see Section IV for a formal description). The verifier first uses $\mathrm{poly}(\log(N), \lambda)$ bits of randomness to select a pair of functions $\{f_{k,b}\}_{b\in\{0,1\}}$ from an NTCF family, and sends the public function key k to the quantum device. This pair of functions can be interpreted as a single 2-to-1 function $f_k : (b, x) \mapsto f_{k,b}(x)$. The verifier keeps private the trapdoor information that allows to invert f_k. The protocol then proceeds for N rounds. In each round the device first outputs a point y in the common range of $f_{k,0}$ and $f_{k,1}$. After having received y, the verifier issues one of two challenges: 0 or 1 — pre-image or equation. If the challenge is "pre-image", then the device must output an x such that $f(x) = y$. If the challenge is "equation" then the device must output a nontrivial binary vector d such that $d \cdot (x_0 + x_1) = 0$, where x_0 and x_1 are the unique pre-images of y under $f_{k,0}$ and $f_{k,1}$ respectively. Since the verifier has the secret key, she can efficiently compute x_0 and x_1 from y, and therefore check the correctness of the device's response to each challenge. The verifier chooses $\mathrm{poly}\log(N)$ rounds in which to issue the challenge 1, or "equation", at random. Selecting these rounds requires only $\mathrm{poly}\log(N)$ random bits. At the end of each such round, the verifier samples a new pair of functions from the NTCF family, and communicates the new public key to the device. On each of the remaining $N - \mathrm{poly}\log(N)$ rounds the verifier records a random bit according to whether the device returns the pre-image x_0, or x_1 (e.g. recording 0 for the lexicographically smaller pre-image). At the end of the protocol the verifier uses a strong quantum-proof randomness extractor to extract $\Omega(N)$ bits of randomness from the recorded string (this requires at most an additional $\mathrm{poly}\log(N)$ random bits of seed).

To guarantee that the extractor produces bits that are statistically close to uniform, we would like to prove that the $N - \mathrm{poly}\log(N)$ random bits recorded by the verifier must have $\Omega(N)$ bits of (smoothed) min-entropy,[2] even conditioned on the side information available to an infinitely powerful quantum adversary, who may share an arbitrary entangled state with the quantum device. The analysis proceeds as follows. First we assume without loss of generality that the entire protocol is run coherently, i.e. we may assume that the initial state of the quantum device (holding quantum register D) and the adversary (holding quantum register E) is a pure state $|\phi\rangle_{\mathsf{DE}}$, since the adversary may as well start with a purification of their joint state. We may also assume that the verifier starts with a cat state on $\mathrm{poly}\log(N)$ qubits, and uses one of the registers of the state, C, to provide the random bits used to select the test rounds and to issue the challenges in those rounds. (This is for the sake of analysis only, the actual verifier is of course completely classical.) We can similarly arrange that the state remains pure throughout the protocol by using the principle of deferred measurement. Our goal is to show a lower bound on the smooth min-entropy of the output register O in which the verifier has recorded the device's outputs, conditioned on the state E of the adversary, and on the register C of the cat state (conditioning on the latter represents the fact that the verifier's choice of challenges may be leaked to the adversary, and we would like security even in this scenario). Intuitively, this amounts to bounding the information accessible to the most powerful adversary quantum mechanics allows, conditioned on the joint state of the verifier and device.

In order to bound the entropy of the final state we need to show that the entropy "accumulates" at each round of the protocol. A general framework to establish entropy accumulation in quantum protocols such as the one considered here was introduced in [AFDF+18]. At a high level, the approach consists in reducing the goal of a min-entropy bound to a bound on the appropriate notion of $(1+\varepsilon)$ quantum conditional Rényi entropy, and then arguing that, under suitable conditions on the process that generates the outcomes recorded in the protocol, entropy accumulates sequentially throughout the protocol.

In a little more detail, the first step on getting a handle on the smooth min-entropy is to use the quantum asymptotic equipartion property (QAEP) [TCR09] to relate it to the $(1+\varepsilon)$ conditional Rényi entropy, for suitably small ε. The second step uses a duality relation for the conditional Rényi entropy to relate the $(1+\varepsilon)$ conditional Rényi entropy of the output register O, conditioned on the adversary side information in R and the register C of the cat state, to a quantity analogous to the $(1-\varepsilon')$ conditional Rényi entropy of the output register, conditioned on the register E for the device, and a purifying copy of the register C of the cat state. The latter quantity, a suitable conditional entropy of the output register conditioned on the challenge register and the state of the device, is the quantity that we ultimately aim to bound. Note what these transformations have achieved for us: it is now sufficient to consider as side information only "known" quantities in the protocol, the verifier's choice of challenges and the device's state; the information held by the adversary plays no other role than that of a purifying register.

Our cryptographic setting presents a new difficulty for this framework, which is that in contrast to the scenarii based on two-player games considered in [AFDF+18] (see also [MS14]), in our setting it is not *impossible* for a deterministic device to succeed in the test: it is merely *computationally hard* to do so. This prevents us from directly applying the results in [AFDF+18], [MS14], which seem to crucially rely on the fact that the process that generates the randomness does so irrespective of the quantum state in which it is initialized (as long as the output of the process satisfies the test's success criterion). This require-

[2]We refer to Section II for definitions of entropic quantities.

ment comes about from conditioning that is performed in order to show that entropy accumulates; in our setting, conditioning is more delicate as it can in principle induce non-computationally efficient states for the device. s in reducing the goal of a min-entropy bound to a bound on a suitable $(1+\varepsilon)$ conditional Rényi entropy, and then arguing that, under suitable conditions on the process that generates the outcomes recorded in the protocol, entropy accumulates sequentially throughout the protocol.

Recall that we argued that for a single round of the protocol, we can decompose the device's Hilbert space into a direct sum of one- or two-dimensional subspaces, such that within most two-dimensional subspace the "pre-image" and "equation" measurements correspond to orthonormal bases that make an angle close to $\pi/4$ with each other. Showing that the Rényi entropy accumulates in each round requires a device in which *all* angles are close to $\pi/4$, not "almost all". To accommodate for this we "split" the state of the device into its component on the good subspace, where the angles are unbiased, and the bad subspace, where the measurements may be aligned. The fact that the distinction between good and bad subspace is not measured in the protocol, but is only a distinction made for the analysis, requires us to apply a fairly delicate martingale based argument that takes into account possible interference effects and bounds those "branches" where the state has gone through the bad subspace an improbably large number of times. Whenever the state lies in the good susbpace, we can appeal to an uncertainty principle from [MS14] to show that the device's measurement increases the conditional Rényi entropy of the output register by a small additive constant. Pursuing this approach across all N rounds, we obtain a linear lower bound on the conditional Rényi entropy of the output register, conditioned on the state of the device. As argued above this in turn translates into a linear lower bound on the smooth conditional min-entropy of the output, conditioned on the state of the adversary and the verifier's choice of challenges. It only remains to apply a quantum-proof randomness extractor to the output, using a poly-logarithmic number of additional bits of randomness, to obtain the final result.

Our NTCF Family: Our goal is to construct a family of pairs of injective functions f_0, f_1 with the same image such that it is hard to find a collision x_0, x_1 with $f_0(x_0) = f_1(x_1)$, but so that given a suitable trapdoor it is possible to recover, for any y, values x_0, x_1 such that $f_0(x_0) = f_1(x_1) = y$. We do this by relying on the hardness of the Learning with Errors (LWE) problem [Reg05]. LWE states that given a public uniformly random matrix $\mathbf{A} \in \mathbb{Z}_q^{m\times n}$ for $m \gg n$, it is intractable to distinguish between $\mathbf{u} = \mathbf{As} + \mathbf{e} \pmod q$ and a uniform vector, for a uniform vector $\mathbf{s}$ and small discrete Gaussian vector $\mathbf{e}$ (all arithmetic from here on is performed modulo q; we use $\oplus$ to denote binary XOR). Inspired by [Mah17a], our function pair will be characterized by $(\mathbf{A}, \mathbf{u} = \mathbf{As} + \mathbf{e})$, but for a *binary* vector $\mathbf{s}$.[3] The trapdoor for our function is a lattice trapdoor for $\mathbf{A}$ that allows to recover $\mathbf{s}, \mathbf{e}$ given a vector of the form $\mathbf{As} + \mathbf{e}$ (it is possible to generate $\mathbf{A}$ together with a trapdoor such that $\mathbf{A}$ is indistinguishable from uniform, as originally shown by Ajtai [Ajt99]).

The structure of the LWE problem motivates us to consider functions f_0, f_1 that range over probability distributions. Specifically, we define the distribution $f_b(\mathbf{x})$ as $f_b(\mathbf{x}) = \mathbf{Ax} + b \cdot (\mathbf{As}) + \mathbf{e}'$ where $\mathbf{e}'$ is a discrete Gaussian random variable with a sufficiently wide Gaussian parameter. Observe that the functions have overlapping images in the following sense: $f_0(\mathbf{x}) = f_1(\mathbf{x} - \mathbf{s})$. Moreover, since $\mathbf{As} + \mathbf{e}'$ is statistically indistinguishable from $\mathbf{u} + \mathbf{e}'$, we can efficiently sample from the distribution $f_b(\mathbf{x})$ up to negligible statistical distance. This probabilistic notion complicates the definition and use of the family, but the principles are similar to the deterministic version.[4]

Finally, we need to show the adaptive hardcore bit property. Formally, letting $\mathbf{x}_0, \mathbf{x}_1 \in \mathbb{Z}_q^n$ be a collision, and letting $z_0, z_1 \in \{0,1\}^{n \log q}$ be their binary representations, respectively, we need to show that it is intractable to come up with a pair (b, z_b), for some $b \in \{0,1\}$, together with a nontrivial vector d and with the value $d \cdot (z_0 \oplus z_1)$, with probability noticeably better than $\frac{1}{2}$. "Nontrivial" here means belonging to a well defined and efficiently recognizable set D with density ≈ 1 in $\{0,1\}^{n \log q}$ (e.g. the zero vector is obviously excluded). Assume for the sake of this overview that we get a tuple (z_0, d, c). We first notice that since $\mathbf{x}_0, \mathbf{x}_1$ is a collision, then $\mathbf{x}_1 = \mathbf{x}_0 - \mathbf{s} \pmod q$. We now use the fact that $\mathbf{s}$ is a binary vector to show, using simple arithmetic, that $z_0 \oplus z_1$ can be expressed as a linear function of the bits of $\mathbf{s}$, so that $d \cdot (z_0 \oplus z_1) = \hat{d} \cdot \mathbf{s} \pmod 2$, for some $\hat{d} \in \{0,1\}^n$. (the description of this transformation will effect our choice of the set D). We thus need to show that it is intractable, given the instance $(\mathbf{A}, \mathbf{u} = \mathbf{As} + \mathbf{e})$, to come up with $\hat{d}, \hat{d} \cdot \mathbf{s} \pmod 2$. To prove this we use the lossiness technique used in [GKPV10] and show that this is equivalent to coming up with $\hat{d}, \hat{d} \cdot \mathbf{s} \pmod 2$ given $\mathbf{B}, \mathbf{Bs}$ where $\mathbf{B} \in \mathbb{Z}_q^{k\times n}$ is now a highly shrinking function, even for binary inputs, i.e. $k \log q \ll n$. This seems like an easy task since the adversary now doesn't have the complete information about $\mathbf{s}$ so it shouldn't be able to compute $\hat{d} \cdot \mathbf{s} \pmod 2$ for any reasonable $\hat{d}$, except $\hat{d}$ might depend on $\mathbf{B}$ itself (recall that $\hat{d}$ is chosen adversarially). We prove via Fourier analysis that if $\mathbf{B}$ is sufficiently shrinking, then there is no $\hat{d}$ that can take advantage of the dependence on

[3] It is known that LWE is hard even with binary secrets. We do not use this property explicitly but rather employ the respective techniques in our proof.

[4] Another possible variant is to define $f_b(\mathbf{x}) = \lfloor \mathbf{Ax} + b \cdot (\mathbf{As}) \rceil$ where $\lfloor \cdot \rceil$ is a rounding function that truncates "many" of the least significant bits of its operand. However, we remain with the Gaussian variant which is easier to analyze.

B, which completes the proof.

Organization: We start with some notation and preliminaries in Section II. Section III contains the definition of a noisy trapdoor claw-free family (NTCF). Our construction for such a family is given in Section **??** (with Appendix **??** containing relevant preliminaries on the learning with errors problem). The randomness generation protocol is described in Section IV. In Section V we introduce our formalism for modeling the actions of an arbitrary prover, or device, in the protocol. In Section **??** we analyze a single round of the protocol, and in Section **??** we show that randomness accumulates across multiple rounds.

Due to space constraints all proofs have been omitted from this extended abstract. We refer to the full version [BCM+18] for complete statements and proofs.

II. Preliminaries

A. Notation

$\mathbb{Z}$ is the set of integers, and $\mathbb{N}$ the set of natural numbers. For any $q \in \mathbb{N}$ such that $q \geq 2$ we let $\mathbb{Z}_q$ denote the ring of integers modulo q. We generally identify an element $x \in \mathbb{Z}_q$ with its unique representative $[x]_q \in (-\frac{q}{2}, \frac{q}{2}] \cap \mathbb{Z}$. For $x \in \mathbb{Z}_q$ we define $|x| = |[x]_q|$. When considering an $s \in \{0,1\}^n$ we sometimes also think of s as an element of $\mathbb{Z}_q^n$, in which case we write it as $\mathbf{s}$.

We use the terminology of polynomially bounded and negligible functions. A function $n : \mathbb{N} \to \mathbb{R}_+$ is *polynomially bounded* if there exists a polynomial p such that $n(\lambda) \leq p(\lambda)$ for all $\lambda \in \mathbb{N}$. A function $n : \mathbb{N} \to \mathbb{R}_+$ is *negligible* if for every polynomial p, $p(\lambda)n(\lambda) \to_{\lambda\to\infty} 0$. We write $\mathrm{negl}(\lambda)$ to denote an arbitrary negligible function of λ. For two parameters κ, λ we write $\kappa \ll \lambda$ to express the constraint that κ should be "sufficiently smaller than" λ, meaning that there exists a small universal constant $c > 0$ such that $\kappa \leq c\lambda$.

$\mathcal{H}$ always denotes a finite-dimensional Hilbert space. We use indices $\mathcal{H}_\mathsf{A}$, $\mathcal{H}_\mathsf{B}$, etc., to refer to distinct spaces. $\mathrm{Pos}(\mathcal{H})$ is the set of positive semidefinite operators on $\mathcal{H}$, and $\mathrm{D}(\mathcal{H})$ the set of density matrices, i.e. the positive semidefinite operators with trace 1. For an operator X on $\mathcal{H}$, we use $\|X\|$ to denote the operator norm (largest singular value) of X, and $\|X\|_{tr} = \frac{1}{2}\|X\|_1 = \frac{1}{2}\mathrm{Tr}\sqrt{XX^\dagger}$ for the trace norm.

B. Entropies

For $p \in [0,1]$ we write $H(p) = -p\log p - (1-p)\log(1-p)$ for the binary Shannon entropy. We measure randomness using Rényi conditional entropies. For a positive semidefinite matrix $\sigma \in \mathrm{Pos}(\mathcal{H})$ and $\varepsilon \geq 0$, let

$$\langle \sigma \rangle_{1+\varepsilon} = \mathrm{Tr}\big(\sigma^{1+\varepsilon}\big) .$$

In addition, for positive semidefinite $\sigma, \rho \in \mathrm{Pos}(\mathcal{H})$ such that the support of ρ is included in the support of σ, and $\varepsilon \geq 0$, let

$$\tilde{Q}_{1+\varepsilon}(\rho\|\sigma) = \langle \sigma^{-\frac{\varepsilon}{2(1+\varepsilon)}} \rho \sigma^{-\frac{\varepsilon}{2(1+\varepsilon)}} \rangle_{1+\varepsilon} . \tag{1}$$

Quantum analogues of the conditional Rényi entropies can be defined as follows.

Definition 2. Let $\rho_{\mathsf{AB}} \in \mathrm{Pos}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B})$ be positive semidefinite. Given $\varepsilon > 0$, the $(1+\varepsilon)$ *Rényi entropy* of A conditioned on B is defined as

$$H_{1+\varepsilon}(A|B)_\rho = \sup_{\sigma \in \mathrm{D}(\mathcal{H}_\mathsf{B})} H_{1+\varepsilon}(A|B)_{\rho|\sigma} ,$$

where for any $\sigma_\mathsf{B} \in \mathrm{D}(\mathcal{H}_\mathsf{B})$,

$$H_{1+\varepsilon}(A|B)_{\rho|\sigma} = -\frac{1}{\varepsilon}\log \tilde{Q}_{1+\varepsilon}(\rho\|\sigma) .$$

.

Rényi entropies are used in the proofs because they have better "chain-rule-like" properties than the min-entropy, which is the most appropriate measure for randomness quantification.

Definition 3. Let $\rho_{\mathsf{AB}} \in \mathrm{Pos}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B})$ be positive semidefinite. Given a density matrix the *min-entropy* of A conditioned on B is defined as

$$H_\infty(A|B)_\rho = \sup_{\sigma \in \mathrm{D}(\mathcal{H}_\mathsf{B})} H_\infty(A|B)_{\rho|\sigma} ,$$

where for any $\sigma_\mathsf{B} \in \mathrm{D}(\mathcal{H}_\mathsf{B})$,

$$H_\infty(A|B)_{\rho|\sigma} = \max\left\{\lambda \geq 0 \,|\, 2^{-\lambda}\,\mathrm{Id}_A \otimes \sigma_B \geq \rho_{AB}\right\} .$$

It is often convenient to consider the *smooth* min-entropy, which is obtained by maximizing the min-entropy over all positive semidefinite operators matrices in an ε-neighborhood of ρ_{AB}. The definition of neighborhood depends on a choice of metric; the canonical choice is the "purified distance". Since this choice will not matter for us we defer to [Tom15] for a precise definition.

Definition 4. Let $\varepsilon \geq 0$ and $\rho_{\mathsf{AB}} \in \mathrm{Pos}(\mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B})$ positive semidefinite. The *ε-smooth min-entropy* of A conditioned on B is defined as

$$H_\infty^\varepsilon(A|B)_\rho = \sup_{\sigma_{\mathsf{AB}} \in \mathcal{B}(\rho_{\mathsf{AB}}, \varepsilon)} H_\infty(A|B)_\sigma ,$$

where $\mathcal{B}(\rho_{\mathsf{AB}}, \varepsilon)$ is the ball of radius ε around ρ_{AB}, taken with respect to the purified distance.

The following theorem relates the min-entropy to the the Rényi entropies introduced earlier. The theorem expresses the fact that, up to a small amount of "smoothing" (the parameter δ in the theorem), all these entropies are of similar order.

Theorem 5 (Theorem 4.1 [MS14])**.** *Let $\rho_{XE} \in \mathrm{Pos}(\mathcal{H}_X \otimes \mathcal{H}_E)$ be positive semidefinite of the form $\rho_{XE} = \sum_{x\in\mathcal{X}} |x\rangle\langle x| \otimes \rho_E^x$, where $\mathcal{X}$ is a finite alphabet. Let $\sigma_E \in$*

$\mathrm{D}(\mathcal{H}_E)$ *be an arbitrary density matrix. Then for any* $\delta > 0$ *and* $0 < \varepsilon \leq 1$,

$$H_\infty^\delta(X|E)_\rho \geq -\frac{1}{\varepsilon}\log\Big(\sum_x \tilde{Q}_{1+\varepsilon}(\rho_E^x \| \sigma_E)\Big) - \frac{1+2\log(1/\delta)}{\varepsilon}\,.$$

III. Trapdoor claw-free hash functions

Let λ be a security parameter, and $\mathcal{X}$ and $\mathcal{Y}$ be finite sets (depending on λ). For our purposes an ideal family of functions $\mathcal{F}$ would have the following properties. For each public key k, there are two functions $\{f_{k,b} : \mathcal{X} \to \mathcal{Y}\}_{b\in\{0,1\}}$ that are both injective and have the same range (equivalently, $(b,x) \mapsto f_{k,b}(x)$ is 2-to-1), and are invertible given a suitable trapdoor t_k (i.e. t_k can be used to compute x given b and $y = f_{k,b}(x)$). Furthermore, the pair of functions should be claw-free: it must be hard for an attacker to find two pre-images $x_0, x_1 \in \mathcal{X}$ such that $f_{k,0}(x_0) = f_{k,1}(x_1)$. Finally, the functions should satisfy an adaptive hardcore bit property, which is a stronger form of the claw-free property: assuming for convenience that $\mathcal{X} = \{0,1\}^w$, we would like that it is computationally infeasible to simultaneously generate an $(b, x_b) \in \{0,1\} \times \mathcal{X}$ and a $d \in \{0,1\}^w \setminus \{0^w\}$ such that with non-negligible advantage over $\frac{1}{2}$ the equation $d \cdot (x_0 \oplus x_1) = 0$, where x_{1-b} is defined as the unique element such that $f_{k,1-b}(x_{1-b}) = f_{k,b}(x_b)$, holds.

Unfortunately, we do not know how to construct a function family that exactly satisfies all these requirements under standard cryptographic assumptions. Instead, we construct a family that satisfies slightly relaxed requirements, that we will show still suffice for our purposes, based on the hardness of the learning with errors (LWE) problem. The requirements are relaxed as follows. First, the range of the functions is no longer a set $\mathcal{Y}$; instead, it is $\mathcal{D}_\mathcal{Y}$, the set of probability densities over $\mathcal{Y}$. That is, each function returns a density, rather than a point. The trapdoor injective pair property is then described in terms of the support of the output densities: these supports should either be identical, for a colliding pair, or be disjoint, in all other cases.

The consideration of functions that return densities gives rise to an additional requirement of efficiency: there should exist a quantum polynomial-time procedure that efficiently prepares a superposition over the range of the function, i.e. for any key k and $b \in \{0,1\}$, the procedure can prepare the state

$$\frac{1}{\sqrt{\mathcal{X}}} \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} \sqrt{f_{k,b}(x)(y)}\,|x\rangle|y\rangle\,. \tag{2}$$

In our instantiation based on LWE, it is not possible to prepare (2) perfectly, but it is possible to create a superposition with coefficients $\sqrt{f'_{k,b}(x)}$, such that the resulting state is within negligible trace distance of (2). The density $f'_{k,b}(x)$ is required to satisfy two properties used in our protocol. First, it must be easy to check, without the trapdoor, if an $y \in \mathcal{Y}$ lies in the support of $f'_{k,b}(x)$. Second, the inversion algorithm should operate correctly on all y in the support of $f'_{k,b}(x)$.

We slightly modify the adaptive hardcore bit requirement as well. Since the set $\mathcal{X}$ may not be a subset of binary strings, we first assume the existence of an injective, efficiently invertible map $J : \mathcal{X} \to \{0,1\}^w$. Next, we only require the adaptive hardcore bit property to hold for a subset of all nonzero strings, instead of the set $\{0,1\}^w \setminus \{0^w\}$. Finally, membership in the appropriate set should be efficiently checkable, given access to the trapdoor.

A formal definition follows.

Definition 6 (NTCF family)**.** Let λ be a security parameter. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets. Let $\mathcal{K}_\mathcal{F}$ be a finite set of keys. A family of functions

$$\mathcal{F} = \{f_{k,b} : \mathcal{X} \to \mathcal{D}_\mathcal{Y}\}_{k\in\mathcal{K}_\mathcal{F}, b\in\{0,1\}}$$

is called a **noisy trapdoor claw free (NTCF) family** if the following conditions hold:

1) **Efficient Function Generation.** There exists an efficient probabilistic algorithm $\mathrm{GEN}_\mathcal{F}$ which generates a key $k \in \mathcal{K}_\mathcal{F}$ together with a trapdoor t_k:
$$(k, t_k) \leftarrow \mathrm{GEN}_\mathcal{F}(1^\lambda)\,.$$
2) **Trapdoor Injective Pair.** For all keys $k \in \mathcal{K}_\mathcal{F}$ the following conditions hold.
   a) *Trapdoor*: For all $b \in \{0,1\}$ and $x \neq x' \in \mathcal{X}$, $\mathrm{SUPP}(f_{k,b}(x)) \cap \mathrm{SUPP}(f_{k,b}(x')) = \emptyset$. Moreover, there exists an efficient deterministic algorithm $\mathrm{INV}_\mathcal{F}$ such that for all $b \in \{0,1\}$, $x \in \mathcal{X}$ and $y \in \mathrm{SUPP}(f_{k,b}(x))$, $\mathrm{INV}_\mathcal{F}(t_k, b, y) = x$.
   b) *Injective pair*: There exists a perfect matching $\mathcal{R}_k \subseteq \mathcal{X} \times \mathcal{X}$ such that $f_{k,0}(x_0) = f_{k,1}(x_1)$ if and only if $(x_0, x_1) \in \mathcal{R}_k$.
3) **Efficient Range Superposition.** For all keys $k \in \mathcal{K}_\mathcal{F}$ and $b \in \{0,1\}$ there exists a function $f'_{k,b} : \mathcal{X} \mapsto \mathcal{D}_\mathcal{Y}$ such that
   a) For all $(x_0, x_1) \in \mathcal{R}_k$ and $y \in \mathrm{SUPP}(f'_{k,b}(x_b))$, $\mathrm{INV}_\mathcal{F}(t_k, b, y) = x_b$ and $\mathrm{INV}_\mathcal{F}(t_k, b \oplus 1, y) = x_{b\oplus 1}$.
   b) There exists an efficient deterministic procedure $\mathrm{CHK}_\mathcal{F}$ that, on input k, $b \in \{0,1\}$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, returns 1 if $y \in \mathrm{SUPP}(f'_{k,b}(x))$ and 0 otherwise. Note that $\mathrm{CHK}_\mathcal{F}$ is not provided the trapdoor t_k.
   c) For every k and $b \in \{0,1\}$,
   $$\mathrm{E}_{x\leftarrow_U\mathcal{X}}\big[\,H^2(f_{k,b}(x),\, f'_{k,b}(x))\,\big] \leq \mu(\lambda)\,,$$
   for some negligible function $\mu(\cdot)$. Here H^2 is the Hellinger distance; see (**??**). Moreover, there exists an efficient procedure $\mathrm{SAMP}_\mathcal{F}$ that on

input k and $b \in \{0,1\}$ prepares the state

$$\frac{1}{\sqrt{|\mathcal{X}|}} \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} \sqrt{(f'_{k,b}(x))(y)}|x\rangle|y\rangle \ . \quad (3)$$

4) **Adaptive Hardcore Bit.** For all keys $k \in \mathcal{K}_\mathcal{F}$ the following conditions hold, for some integer w that is a polynomially bounded function of λ.
   a) For all $b \in \{0,1\}$ and $x \in \mathcal{X}$, there exists a set $G_{k,b,x} \subseteq \{0,1\}^w$ such that $\Pr_{d\leftarrow_U\{0,1\}^w}[d \notin G_{k,b,x}]$ is negligible, and moreover there exists an efficient algorithm that checks for membership in $G_{k,b,x}$ given k, b, x and the trapdoor t_k.
   b) There is an efficiently computable injection $J : \mathcal{X} \to \{0,1\}^w$, such that J can be inverted efficiently on its range, and such that the following holds. If

$$\begin{aligned} H_k &= \big\{(b, x_b, d, d\cdot(J(x_0)\oplus J(x_1))) \,\big|\, \\ &\qquad b\in\{0,1\},\ (x_0,x_1)\in\mathcal{R}_k, \\ &\qquad d \in G_{k,0,x_0}\cap G_{k,1,x_1}\big\}\ ,^{5} \\ \overline{H}_k &= \{(b, x_b, d, c) \,|\, (b, x, d, c\oplus 1) \in H_k\}\ , \end{aligned}$$

   then for any quantum polynomial-time procedure $\mathcal{A}$ there exists a negligible function $\mu(\cdot)$ such that

$$\Big|\Pr_{(k,t_k)\leftarrow \mathrm{GEN}_\mathcal{F}(1^\lambda)}[\mathcal{A}(k)\in H_k] - \Pr_{(k,t_k)\leftarrow \mathrm{GEN}_\mathcal{F}(1^\lambda)}[\mathcal{A}(k)\in \overline{H}_k]\Big| \le \mu(\lambda)\ . \quad (4)$$

IV. Protocol description

We introduce two protocols. The first we call the *(general) randomness expansion protocol*, or Protocol 1. This is our main randomness expansion protocol. It is introduced in Section IV-A, and summarized in Figure 1. The protocol describes the interaction between a *verifier* and *prover*. Ultimately, we aim to obtain the guarantee that any computationally bounded prover that is accepted with non-negligible probability by the verifier in the protocol must generate transcripts that contain information-theoretic randomness.

The second protocol is called the *simplified protocol*, or Protocol 2. It is introduced in Section IV-B, and summarized in Figure 2. This protocol abstracts some of the main features Protocol 1, and will be used as a tool in the analysis (it is not meant to be executed literally).

A. The randomness expansion protocol

Our randomness expansion protocol, Protocol 1, is described in Figure 1. The protocol is parametrized by a choice of security parameter λ. All other parameters are assumed to be specified as a function of λ: the number of rounds N, the error tolerance parameter $\gamma \ge 0$, and the testing parameter $q \in (0,1]$. For intuition, γ can be thought of as a small constant, q as a parameter that scales as $\mathrm{poly}(\lambda)/N$ (for example, $q = \lambda/N$), and N as a function that may grow super-polynomially, or even exponentially, with λ. In addition, the protocol depends on an NTCF family $\mathcal{F}$ (see Definition 6) that is known to both the verifier and the prover.

At the start of the protocol, the verifier executes $(k, t_k) \leftarrow \mathrm{GEN}_\mathcal{F}(1^\lambda)$ to obtain the public key k and trapdoor t_k for a pair of functions $\{f_{k,b} : \mathcal{X} \to \mathcal{D}_\mathcal{Y}\}_{b\in\{0,1\}}$ from the NTCF family. The verifier sends the key k to the prover and keeps the associated trapdoor private.

In each of the N rounds of the protocol, the prover is first required to provide a value $y \in \mathcal{Y}$. For each $b \in \{0,1\}$, the verifier uses the trapdoor to compute $\hat{x}_b \leftarrow \mathrm{INV}_\mathcal{F}(t_k, b, y)$. (If the inversion procedure fails, the verifier requests another sample from the prover.) For convenience, introduce a set

$$\hat{G}_y = G_{k,0,x_0} \cap G_{k,1,x_1}\ . \quad (5)$$

The verifier then chooses a round type $G \in \{0,1\}$ according to a biased distribution: either a *test round*, $G = 0$, chosen with probability $\Pr(G = 0) = q$, or a *generation round*, $G = 1$, chosen with the remaining probability $\Pr(G = 1) = 1 - q$. The former type of round is less frequent, as the parameter q will eventually be set to a very small value, that goes to 0 with the number of rounds of the protocol. The prover is not told the round type.

Depending on the round type, the verifier chooses a challenge $C \in \{0,1\}$ that she sends to the prover. In the case of a test round the challenge is chosen uniformly at random; in the case of a generation round the challenge is always $C = 1$. In case $C = 0$ the prover is asked to return a pair $(m, d) \in \{0,1\} \times \{0,1\}^w$. The pair is called valid if $m = d \cdot (J(\hat{x}_0) \oplus J(\hat{x}_1))$ and $d \in \hat{G}_y$. If $d \in \hat{G}_y$, the verifier sets a decision bit $W = 1$ if the answer is valid, and $W = 0$ if not. If $d \notin \hat{G}_y$, the verifier sets the decision bit $W \in \{0,1\}$ uniformly at random.[6] In case $C = 1$, the prover should return a pair $(b, x) \in \{0,1\} \times \mathcal{X}$. The pair is called valid if $f_{k,b}(x) = y$. The verifier sets a decision bit $W = 1$ in case the pair is valid, and $W = 0$ otherwise. The set of valid pairs on challenge $C = c \in \{0,1\}$ is denoted $V_{y,c}$.

After each test round the verifier samples a fresh $(k, t_k) \leftarrow \mathrm{GEN}_\mathcal{F}(1^\lambda)$ and communicates the new public key k to the prover.

At the end of the protocol, the verifier computes the fraction of test rounds in which the decision bit has been set to 1. If this fraction is smaller than $(1 - \gamma)$, the verifier aborts. Otherwise, the verifier returns the concatenation of the bits b obtained from the prover in generation rounds. (These bits

[5] Note that although both x_0 and x_1 are referred to to define the set H_k, only one of them, x_b, is explicitly specified in any 4-tuple that lies in H_k.

[6] This choice if made for technical reasons that have to do with the definition of the adaptive hardcore bit property; see Section **??** and the proof of Proposition **??** for details.

are recorded in the verifier's output string $O_1 \cdots O_N$, such that $O_i = 0$ whenever the round is a test round.)

Let λ be a security parameter. Let N be a polynomially bounded function of λ, and $\gamma, q > 0$ functions of λ. Let $\mathcal{F}$ be an NTCF family.

At the start of the protocol, the verifier communicates N to the device. In addition, the verifier samples an initial key $(k, t_k) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda)$, sends k to the prover and keeps the trapdoor information t_k private.

1) For $i = 1, \ldots, N$:
   a) The prover returns a $y \in \mathcal{Y}$ to the verifier. For $b \in \{0,1\}$ the verifier uses the trapdoor to compute $\hat{x}_b \leftarrow \text{INV}_{\mathcal{F}}(t_k, b, y)$.
   b) The verifier selects a round type $G_i \in \{0,1\}$ according to a Bernoulli distribution with parameter q: $\Pr(G_i = 0) = q$ and $\Pr(G_i = 1) = 1 - q$. In case $G_i = 0$ (*test round*), she chooses a challenge $C_i \in \{0,1\}$ uniformly at random. In case $G_i = 1$ (*generation round*), she sets $C_i = 1$. The verifier keeps G_i private, and sends C_i to the prover.
      i) In case $C_i = 0$ the prover returns $(m, d) \in \{0,1\} \times \{0,1\}^w$. If $d \notin \hat{G}_y$, the set defined in (5), the verifier sets W to a uniformly random bit. Otherwise, the verifier sets $W = 1$ if $d \cdot (J(\hat{x}_0) \oplus J(\hat{x}_1)) = m$ and $W = 0$ if not.
      ii) In case $C_i = 1$ the prover returns $(b, x) \in \{0,1\} \times \mathcal{X}$. The verifier sets W as the value returned by $\text{CHK}_{\mathcal{F}}(k, b, x, y)$.
   c) In case $G_i = 1$, the verifier sets $O_i = b$. In case $G_i = 0$, she sets $W_i = W$.
   d) In case $G_i = 0$, the verifier samples a new key $(k, t_k) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda)$. She sends k to the prover and keeps the trapdoor information t_k private. This key will be used until the next test round, included.
2) If $\sum_{i:G_i=0} W_i < (1 - \gamma) qN$, the verifier aborts. Otherwise, she returns the string O obtained by concatenating the bits O_i for all $i \in \{1, \ldots, N\}$ such that $G_i = 1$.

Figure 1. The randomness expansion protocol, Protocol 1. See Definition 6 for notation associated with the NTCF family $\mathcal{F}$.

B. The simplified protocol

For purposes of analysis only we introduce a simplified variant of Protocol 1, which is specified in Figure 2. We call it the *simplified protocol*, or Protocol 2. The protocol is very similar to the randomness expansion protocol described in Figure 1, except that the prover's answers and the verifier's checks are simplified, and in test rounds there is an additional challenge bit $T \in \{0,1\}$. This new challenge asks the prover to perform a projective measurement on its private space that indicates whether the state lies in a "good subspace" (indicated by an outcome $K = 0$) or in the complementary "bad subspace" (outcome $K = 1$). The "good" and "bad" subspace represent portions of space where the device's other two measurements, M and Π are anti-aligned and aligned respectively; see the definition of a simplified device in Section **??** for details.

Let λ be a security parameter. Let N be a polynomially bounded function of λ, and $\gamma, \eta, \kappa, q > 0$ functions of λ.

1) For $i = 1, \ldots, N$:
   a) The verifier selects a round type $G_i \in \{0,1\}$ according to a Bernoulli distribution with parameter q: $\Pr(G_i = 0) = q$ and $\Pr(G_i = 1) = 1 - q$. In case $G_i = 0$ (*test round*), she chooses $C_i \in \{0,1\}$ uniformly at random and $T_i \in \{0,1\}$ such that $\Pr(T_i = 0) = 1 - \kappa$ and $\Pr(T_i = 1) = \kappa$. In case $G_i = 1$ (*generation round*), she sets $C_i = 1$ and $T_i = 0$. The verifier keeps G_i private, and sends (C_i, T_i) to the prover.
      i) In case $C_i = 0$ the prover returns $u \in \{0,1\}$. If $T_i = 1$ the prover in addition reports $k \in \{0,1\}$.[7] If $T_i = 0$ the verifier sets $W_i = u$. If $T_i = 1$ the verifier sets $W_i = u(1 - k)$.
      ii) In case $C_i = 1$ the prover returns $v \in \{0,1,2\}$. The verifier sets $O_i = v$ and $W_i = 1_{v \in \{0,1\}}$.
2) If $\sum_{i:G_i=0 \wedge T_i=1} W_i < (1 - \frac{\gamma}{\kappa} - \eta)\kappa q N$, the verifier rejects the interaction. Otherwise, she returns the string O obtained by concatenating the bits O_i for all $i \in \{1, \ldots, N\}$ such that $G_i = 1$.

Figure 2. The simplified protocol, Protocol 2.

For the case of a challenge $C = 0$, in Protocol 1 the prover returns an equation (d, m). In the simplified protocol the prover returns a single bit $u \in \{0,1\}$ that is meant to directly indicate the verifier's decision (i.e. the bit W). If moreover $T = 1$ the prover is required to reply with an additional bit $k \in \{0,1\}$. In this case, the verifier makes the decision to accept, i.e. sets $W = 1$, if and only if $u = 1$ and $k = 0$. For the case of a challenge $C = 1$, in Protocol 1 the prover returns a pair (b, x). In the simplified protocol the prover returns a value $v \in \{0,1,2\}$ that is such that $v = b$ in case (b, x) is valid, i.e. $(b, x) \in V_{y,1}$, and $v = 2$ otherwise.

Note that this "honest" behavior for the prover is not necessarily efficient. Moreover, it is easy for a "malicious" prover to succeed in Protocol 2, e.g. by always returning

$u = 1$ (valid equation), $k = 0$ (good subspace) and $v \in \{0,1\}$ (valid pre-image). Our analysis will not consider arbitrary provers in Protocol 2, but instead provers whose measurements satisfy certain constraints that arise from the analysis of Protocol 1. For such provers, it will be impossible to succeed in the simplified protocol without generating randomness. Further details are given in Section **??**.

C. Completeness

We describe the intended behavior for the prover in protocol 1. Fix an NTCF family $\mathcal{F}$ and a key $k \in \mathcal{K}_{\mathcal{F}}$. In each round, the "honest" prover performs the following actions.

1) The prover executes the efficient procedure $\mathrm{SAMP}_{\mathcal{F}}$ in superposition to obtain the state $|\psi^{(1)}\rangle$ defined as
$$\frac{1}{\sqrt{|\mathcal{X}|}} \sum_{x\in\mathcal{X}, y\in\mathcal{Y}, b\in\{0,1\}} \sqrt{(f'_{k,b}(x))(y)}|b,x\rangle|y\rangle \ .$$
2) The prover measures the last register to obtain an $y \in \mathcal{Y}$. Using item 2. from the definition of an NTCF, the prover's re-normalized post-measurement state is
$$|\psi^{(2)}\rangle = \frac{1}{\sqrt{2}}\big(|0,x_0\rangle + |1,x_1\rangle\big)|y\rangle \ .$$
   a) In case $C_i = 0$, the prover evaluates the function J on the second register, containing x_b, and then applies a Hadamard transform to all $w+1$ qubits in the first two registers. Tracing out the register that contains y, this yields the state
$$\begin{aligned}|\psi^{(3)}\rangle &= 2^{-\frac{w+2}{2}} \sum_{d,b,m} (-1)^{d\cdot J(x_b)\oplus mb}|m\rangle|d\rangle \\ &= (-1)^{J(x_0)}2^{-\frac{w}{2}} \\ &\quad \sum_{d\in\{0,1\}^w} |d\cdot(J(x_0)\oplus J(x_1))\rangle|d\rangle \ .\end{aligned}$$
   The prover measures both registers to obtain an (m,d) that it sends back to the verifier.
   b) In case $C_i = 1$, the prover measures the first two registers of $|\psi^{(2)}\rangle$ in the computational basis, and returns the outcome (b,x_b) to the verifier.

Lemma 7. *For any λ and $k \leftarrow \mathrm{GEN}_{\mathcal{F}}(1^\lambda)$, the strategy for the honest prover (on input k) in one round of the protocol can be implemented in time polynomial in λ and is accepted with probability negligibly close to* 1.

Proof: Both efficiency and correctness of the prover follow from the definition of an NTCF (Definition 6). The prover fails only if he obtains an outcome $d \notin \hat{G}_y$, which by item 4(a) in the definition happens with negligible probability. ∎

[7]The bit k should not be confused with the public key k for the NTCF that is used in Protocol 1. In Protocol 2, there is no NTCF, and no key.

V. Security analysis

Due to lack of space we refer the reader to the full version [BCM+18] for a complete analysis of the randomness generation properties of Protocol 1. Here we very briefly sketch the main steps of the proof.

We model the behavior of an arbitrary prover in the randomness expansion protocol (Protocol 1 in Figure 1). We do this by introducing a simplified model for the prover as a *device* that implements the prover's actions.

The first step in the proof is to relate the randomness generated by an arbitrary device D in Protocol 1 to the randomness generated by a specific device, that we call the "simplified device" associated to D, in the simplified protocol, Protocol 2 described in 2.

The advantage of the simplified device is that it allows us to abstract the complex structured of Protocol 1, imposed by the use of the NTCF family, to a very simple kind of protocol that interweaves "test" and "generation" rounds, where each round essentially results in a single-bit outcome collected from the prover.

The last step consists in showing that the simplified device derived from D generates randomness in Protocol 2. For this the important point is that, under the assumption that D does not break our post-quantum hardness assumption, we are able to guarantee that the measurements that the simplified device performs in the test and generation rounds are (essentially) mutually unbiased. Thus, if the device consistently reports the same outcome ("win") in test rounds, then it must necessarily generate randomness in the generation rounds. At a more technical level, we adapt the framework for randomness generation from [MS14], based on the use of $(1+\varepsilon)$ Rényi entropies, to our setting, where the guarantees on unbiasedness of the device's measurements are somewhat weaker than in their setting.

For completeness we end the extended abstract by stating formally our main result.

Theorem 8. *Let $\mathcal{F}$ be an NTCF family and λ a security parameter. Let N be a polynomially bounded function of λ such that $N = \Omega(\lambda^2)$. Set $q = \lambda/N$. Then there is a negligible (as a function of λ) δ such that for any small enough $\gamma > 0$, any efficient prover, and side information E correlated with the prover's initial state,*
$$H_\infty^{N\delta}(O|CE)_{\overline{\rho}} \geq (\xi - o(1))N \ ,$$
where $\overline{\rho}$ is the final state of the output, challenge, and adversary registers, restricted to transcripts that are accepted by the verifier in the protocol, ξ is a positive constant,[8] and $o(1)$ is a function that goes to 0 as $\lambda \to \infty$.

Assume that an execution of $\mathrm{GEN}(1^\lambda)$ requires $O(\lambda^r)$ bits of randomness, for some constant r. (For example, for the

[8]The constant ξ is at least some positive universal constant of order $1/10$, for all small enough γ.

case of our construction of a NTCF family based on LWE, we have $r = 2$.) Then an execution of the protocol using the parameters in Theorem 8 requires only $\mathrm{poly}(\lambda, \log N)$ bits of randomness for the verifier to generate the key k and select the challenges. Taking N to be slightly sub-exponential in λ, e.g. $N = 2^{\sqrt{\lambda}}$, yields sub-exponential randomness expansion.

REFERENCES

[Aar18] Scott Aaronson. Certified randomness from quantum supremacy. Personal communication, 2018.

[AFDF+18] Rotem Arnon-Friedman, Frédéric Dupuis, Omar Fawzi, Renato Renner, and Thomas Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nature communications*, 9(1):459, 2018.

[Ajt99] Miklós Ajtai. Generating hard instances of the short basis problem. In *International Colloquium on Automata, Languages, and Programming*, pages 1–9. Springer, 1999.

[BCM+18] Zvika Brakerski, Paul Christiano, Urmila Mahadev, Umesh Vazirani, and Thomas Vidick. Certifiable randomness from a single quantum device. *arXiv preprint arXiv:1804.00640*, 2018.

[BKG+18] Peter Bierhorst, Emanuel Knill, Scott Glancy, Yanbao Zhang, Alan Mink, Stephen Jordan, Andrea Rommal, Yi-Kai Liu, Bradley Christensen, Sae Woo Nam, et al. Experimentally generated randomness certified by the impossibility of superluminal signals. *arXiv preprint arXiv:1803.06219*, 2018.

[CCKW18] Alexandru Cojocaru, Léo Colisson, Elham Kashefi, and Petros Wallden. Delegated pseudo-secret random qubit generator. *arXiv preprint arXiv:1802.08759*, 2018.

[Col06] R. Colbeck. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, Trinity College, University of Cambridge, November 2006.

[FGS11] S. Fehr, R. Gelles, and C. Schaffner. Security and composability of randomness expansion from Bell inequalities. Technical report arXiv:1111.6052, 2011.

[GKPV10] Shafi Goldwasser, Yael Tauman Kalai, Chris Peikert, and Vinod Vaikuntanathan. Robustness of the learning with errors assumption. In Andrew Chi-Chih Yao, editor, *ICS*, pages 230–240. Tsinghua University Press, 2010.

[GMR84] Shafi Goldwasser, Silvio Micali, and Ronald L. Rivest. A "paradoxical'"solution to the signature problem (abstract). In G. R. Blakley and David Chaum, editors, *Advances in Cryptology, Proceedings of CRYPTO '84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings*, volume 196 of *Lecture Notes in Computer Science*, page 467. Springer, 1984.

[HM17] Aram W Harrow and Ashley Montanaro. Quantum computational supremacy. *Nature*, 549(7671):203, 2017.

[LP11] Richard Lindner and Chris Peikert. Better key sizes (and attacks) for lwe-based encryption. In *Cryptographers Track at the RSA Conference*, pages 319–339. Springer, 2011.

[Mah17a] Urmila Mahadev. Classical homomorphic encryption for quantum circuits. *Arxiv preprint arXiv:1708.02130v1*, 2017.

[Mah17b] Urmila Mahadev. Classical homomorphic encryption for quantum circuits. *arXiv preprint arXiv:1708.02130*, 2017.

[Mah18] Urmila Mahadev. Classical verification of quantum computations. Manuscript, 2018.

[MS14] Carl A Miller and Yaoyun Shi. Universal security for randomness expansion from the spot-checking protocol. *arXiv preprint arXiv:1411.6608*, 2014.

[MS16] Carl A Miller and Yaoyun Shi. Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. *Journal of the ACM (JACM)*, 63(4):33, 2016.

[PAM+10] S. Pironio, A. Acin, S. Massar, A. Boyer De La Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and et al. Random numbers certified by Bell's theorem. *Nature*, 464(7291), 2010.

[PM11] S. Pironio and S. Massar. Security of practical private randomness generation. Technical report arXiv:1111.6056, 2011.

[Reg05] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In Harold N. Gabow and Ronald Fagin, editors, *STOC*, pages 84–93. ACM, 2005.

[TCR09] Marco Tomamichel, Roger Colbeck, and Renato Renner. A fully quantum asymptotic equipartition property. *IEEE Transactions on Information Theory*, 55(12):5840–5847, 2009.

[Tom15] Marco Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*, volume 5. Springer, 2015.

[VV11] Umesh Vazirani and Thomas Vidick. Certifiable quantum dice: or, true random number generation secure against quantum adversaries. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 61–76. ACM, 2011. Also available as arXiv:1111.6054.