Certifiable Quantum Dice

Or, True Random Number Generation Secure Against Quantum Adversaries

[Extended Abstract]

Umesh Vazirani*
Computer Science division
UC Berkeley, USA
vazirani@eecs.berkeley.edu

Thomas Vidick†
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, USA
vidick@csail.mit.edu

ABSTRACT

We introduce a protocol through which a pair of quantum mechanical devices may be used to generate n bits that are ε-close in statistical distance from n uniformly distributed bits, starting from a seed of $O(\log n \log 1/\varepsilon)$ uniform bits. The bits generated are certifiably random based only on a simple statistical test that can be performed by the user, and on the assumption that the devices do not communicate in the middle of each phase of the protocol. No other assumptions are placed on the devices' inner workings. A modified protocol uses a seed of $O(\log^3 n)$ uniformly random bits to generate n bits that are $\text{poly}^{-1}(n)$-indistinguishable from uniform even from the point of view of a quantum adversary who may have had prior access to the devices, and may be entangled with them.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Random number generation—*complexity measures, performance measures* ; F.1.2 [**Computation by Abstract Devices**]: Modes of Computation—*Probabilistic Computation*

General Terms

Theory

*Supported by NIST award No. 60NANB10D262, ARO Grant W911NF-09-1-0440 and NSF Grant CCF-0905626
†Supported by NSF Grant 0844626. Part of this work was completed while at UC Berkeley, supported by ARO Grant W911NF-09-1-0440 and NSF Grant CCF-0905626.

STOCŠ12, May 19–22, 2012, New York, New York, USA.

Keywords

Certifiable Randomness, Quantum Computing, Entanglement, Random Number Generator

1. INTRODUCTION

A source of independent random bits is a basic resource in many modern-day computational tasks, such as cryptography, game theoretic protocols, algorithms and physical simulations. However, constructing a physical source of randomness is an unexpectedly tricky task [1] — one that touches on fundamental questions about the nature of randomness. What makes this task particularly challenging is this: how can one even test whether one has succeeded? In other words, suppose someone was to claim that a given box outputs uniformly random bits; is there a practical test to verify that claim?

The root of the difficulty in carrying out such a test is the following: a uniform random generator must output every n-bit sequence with equal probability $1/2^n$, and there seems to be no basis on which to reject any particular output in favor of any other. On the face of it, testing the output of the box amounts to classifying a *single* n-bit string x (or a very small sample of the exponentially many n-bit strings) as being random or not random.

Starting in the mid-80's, computer scientists explored a different approach to the question of designing a uniform random number generator: they assumed that they already had access to a physical device that was guaranteed to output random strings, except that the randomness was of "low quality". They modeled such devices as adversarially controlled sources of randomness, starting with the semi-random source [23], and weak random sources [28]. This sequence of papers has culminated in sophisticated algorithms called randomness extractors that are guaranteed to output a sequence that is arbitrarily close to truly random bits from physical sources of low-quality randomness (see [24] for a survey). It was clear, in a classical World, that these results were the best one could hope for — since it was necessary to assume that randomness

[1]The quest for good hardware number generators goes as far back as the first commercially available computer, the Ferranti Mark I, and continues through Intel's recently announcement of the first *digital* such generator, as part of its new "Ivy bridge" microprocessor [26].

in some form was output by the device in the first place, the only progress could be in minimizing the assumptions placed on the quality of that randomness.

Quantum nonlocality and a test for randomness.

Unlike classical physics, where randomness is implicitly an assertion about our lack of knowledge or computational ability, quantum mechanics offers a source of intrinsic randomness, ensconced in the Born rule, one of the fundamental axioms of the theory. So in principle, it is very simple to design a quantum device that outputs a sequence of independent unbiased bits: simply pass a sequence of qubits in state $|0\rangle$ through a Hadamard gate and measure. This brings us back to the main question addressed in this paper: is it possible to certify that the output of a randomness generating device (based on quantum mechanics) is "really random" even though the user does not trust the experimental skills of the manufacturer, the calibration of the device, the manufacturer's motivations (particularly in cryptographic settings), or even the correctness of quantum mechanics? We describe the construction of a specific kind of quantum random number generator for which the answer to these questions is affirmative. This construction builds upon a proposal of Colbeck [5] and follow up work by Pironio et al. [20] that provides a link between randomness certification and quantum non-locality. Before we can describe the actual construction, we must introduce some basic ideas behind quantum non-locality.

Non-locality is one of the most interesting features of quantum mechanics, and was famously explored in the work of Einstein, Podolski and Rosen [10], and later in the work of Bell [2, 3]. We focus here on a concrete realization of an experiment inspired by Bell's work, that is best phrased as a game, the CHSH game (illustrated in Figure 1), named after its inventors Clauser, Horne, Shimony and Holt [4]. In this game two *non-communicating* parties, represented by spatially separated boxes A, B, are given inputs $x, y \in \{0,1\}$ respectively. Their task is to produce outputs $a, b \in \{0,1\}$ such that the *CHSH condition* $a \oplus b = x \wedge y$ holds. Let p_{CHSH} be the probability that a certain pair of boxes produces outputs satisfying this condition, when the inputs x, y are chosen uniformly at random.

While classical players can achieve a success probability at most $p_{\text{CHSH}} \leq 3/4$ in this game, there is a simple quantum strategy that succeeds with $p_{\text{CHSH}} = \cos^2 \pi/8 \approx 0.85$. Hence we may define a *quantum regime* corresponding to success probability $3/4 < p_{\text{CHSH}} \leq \cos^2 \pi/8 \approx 0.85$. For any value in that range there is a simple quantum-mechanical pair of boxes, still obeying the condition of no communication, which achieves that success probability.

These well-known facts have a striking consequence: any boxes producing correlations that fall in the quantum regime *must be randomized*! Indeed, deterministic boxes are inherently classical, so that their success probability must fall in the classical regime $p_{\text{CHSH}} \leq 3/4$. Hence a simple *statistical test* guaranteeing the presence of randomness, under a single assumption on the process that produced the bits: that it obeys the no-communication condition.[2] This powerful observation was first made in

[2]By this we mean that the probability distribution $p(a, b|x, y)$ describing the distribution of the boxes' outputs, as a function of their inputs, should be *no-signaling*: the marginal distribution of either box's outputs should be independent from inputs to the other.

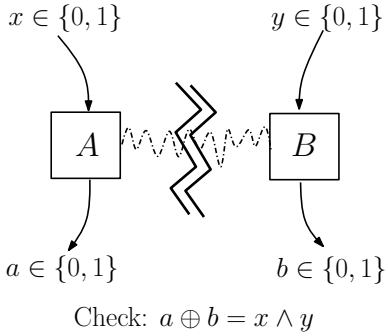

Figure 1: The CHSH game. Any pair of boxes A, B is characterized by a distribution $p(a, b|x, y)$ which is required to be *no-signaling*: the marginal distribution of b is independent of x, and that of a is independent of y.

Colbeck's Ph.D. thesis [5] (see also [6] for an expanded version). The idea was then developed in [20], in which the authors showed that Colbeck's idea could be used to device a procedure that expands an initial seed of $\sqrt{n}$ bits into n bits that are guaranteed to contain a linear amount of min-entropy.[3] (An extractor could then be applied to produce linearly many near-uniform bits.) Pironio et al. even reported an experimental realization of their scheme, demonstrating the generation of 42 new random numbers, in addition to the randomness used to execute the protocol.

Our results.

Let n be an integer, and $\varepsilon > 0$ a parameter such that ε is at least an inverse polynomial in n. We introduce a very simple protocol which uses a random seed of length $O(\log^3 n)$ to generate an n-bit string that is ε-close to uniformly random in statistical distance. This exponentially improves upon the quadratic expansion of [20]. Moreover, the procedure comes with a test which guarantees that the bits produced are ε-close to being indistinguishable from uniform bits even from the point of view of an arbitrary quantum adversary. Establishing such a strong security guarantee was an important open question left open in previous works.

The protocol to achieve this prescribes the interaction of a trusted user with an *untrusted physical device* which we assume is made of two separate boxes, $\mathcal{A}$ and $\mathcal{B}$. The boxes may have been tampered with by the adversary, who could for instance have decided to initialize them in an entangled quantum state that extends into her own laboratory.

The protocol consists of $m = \text{poly}(n)$ phases. Each phase lasts for $k = O(\log n + \log 1/\varepsilon)$ rounds, during each of which the user inputs a single bit to each box and collects a single bit as output. This sequence of mk interactions is non-adaptive: the user can generate the

[3]The paper [20] contained an error that was later fixed in work of Fehr et. al. [11].

$2mk$ input bits in advance from his $O(\text{poly}\log(n, 1/\varepsilon))$-bit random seed before the interaction. It is of critical importance that he reveals the input bits of phase $i+1$ to the boxes only after the completion of phase i. In each phase the user also performs a very simple statistical test (which simply checks that a large fraction of rounds satisfy a CHSH-like condition). If the test is passed in all phases, then the output of box $\mathcal{B}$, say, is efficiently (and classically) post-processed to produce the final output, an-n bit string. If the test is failed in any phase, then the user outputs a special "fail" symbol.

We show that, however the adversary may have prepared the physical boxes, provided the bits they produce in the protocol are accepted in the user's test, the maximum success probability with which the adversary can guess the n bits output at the end of the protocol is exponentially small in n. This condition is stronger than the mere fact that the bits produced contain a linear amount of min-entropy: it implies that no outside adversary can have gained any information about them, and in particular they may be securely used in subsequent post-quantum cryptographic primitives. This conclusion is guaranteed to hold provided the following conditions are met:

1. The user's private random bits are uniformly random.
2. The simple statistical test is passed in all m phases of the protocol.
3. Boxes $\mathcal{A}$ and $\mathcal{B}$ are arbitrary, but their functioning must admit a description consistent with quantum mechanics. In addition, *there must not be any communication between $\mathcal{A}$ and $\mathcal{B}$* throughout the duration of any given phase of the protocol. Formally, this last requirement states the following: for each phase i, the marginal distribution of outputs produced by $\mathcal{A}$ (resp. $\mathcal{B}$) during phase i is independent of all inputs to $\mathcal{B}$ (resp. $\mathcal{A}$) in phase i.

As we shall see, it is possible to implement using very simple quantum mechanics a pair of boxes $\mathcal{A}$ and $\mathcal{B}$ that pass the statistical tests in all m phases with high probability. However, the key point established by our results is that by using such a device as prescribed by the protocol, its outputs can be trusted based only on a belief in the correctness of quantum mechanics (and even that assumption will be relaxed in our second result) and in the fact that there is no communication between the devices — it is not required to, say, believe that the device's manufacturer is trustworthy, experimentally skilled, or that the device is always well calibrated.

We state informally our main result, referring to Section 5 for a precise statement.

THEOREM 1 (INFORMAL). *Let n be an integer, and $\varepsilon = n^{-\alpha}$ for some $\alpha > 0$. Let $(\mathcal{A}, \mathcal{B})$ be an arbitrary pair of boxes, and assume that the physical behavior of $\mathcal{A}$ and $\mathcal{B}$ can be described by two isolated but possibly entangled quantum-mechanical systems (and in particular there is no possibility for communication in-between the boxes). Suppose Protocol B, as described in Figure 3, is executed with $\mathcal{A}$ and $\mathcal{B}$. This execution requires the use of $\tilde{O}(\log^2 n \log 1/\varepsilon)$ random bits, and results in an output string B of $\text{poly}(n)$ bits. Let* CHSH *be the event that the boxes' outputs pass the test performed by the user at the end of an execution of Protocol B, as described in Figure 3, and suppose that* $\Pr\big(\text{CHSH}\big) \geq \varepsilon$.

Then no adversary can guess B with success probability greater than 2^{-n}. More precisely, if $\mathbf{E}$ denotes an arbitrary quantum system, possibly entangled with $\mathcal{A}$ and $\mathcal{B}$ (but such that no communication occurs between $\mathcal{A}, \mathcal{B}$ and $\mathbf{E}$ during the execution of the protocol), then

$$H_\infty^\varepsilon(B|\mathbf{E}) \geq n,$$

where H_∞^ε is the smooth *quantum conditional min entropy.*

We note that the assumption $\Pr\big(\text{CHSH}\big) \geq \varepsilon$ is necessary, as there is always an unavoidable chance that the boxes successfully guess their whole inputs, and deterministically produce matching outputs. While the theorem as stated only guarantees that the bits output by the device have large (smooth) min-entropy, one can obtain bits that are (close to) uniformly random by applying an extractor. In order to preserve the security against quantum adversaries, one should use a construction that is also secure against quantum adversaries. Such constructions exist: for instance an efficient classical extractor due to Trevisan [27], which only requires an additional $O(\log^2 n)$ bits of seed, has been shown secure even against quantum adversaries [25, 8].

In case one is not concerned with the possibility of quantum adversaries, but solely with the production of high-entropy bits, we introduce a simplified protocol that generates n bits that are ε-close to being uniformly random (after application of an efficient extractor, such as the one in [13]), starting from a seed of only $O(\log n \log 1/\varepsilon)$ uniformly random bits. The protocol consists of $m = O(n)$ phases, each lasting for $k = O(\log n + \log 1/\varepsilon)$ rounds. Moreover, the output sequence of n bits is ε-close to uniformly random provided there is no communication between $\mathcal{A}$ and $\mathcal{B}$ in the middle of any phase (a condition enforced, say, by the speed of light limit imposed by special relativity). For the conclusion to hold, it is unnecessary to assume that the boxes are described by quantum mechanics.[4]

We note that a dependence of the initial seed on $\log(1/\varepsilon)$ is clearly necessary to guarantee that the output is ε-close to uniform in statistical distance. This is because a malicious device could attempt to guess the user's private random bits, and behave accordingly. If the user only uses $\log 1/\varepsilon$ random bits, the device's guess will be successful with probability ε, and in that case it can deterministically satisfy all the user's requirements (since they are known in advance). One can also argue (perhaps a little less emphatically) that $\log n$ bits may be necessary, since at best the device acts like a weak random source, and a random seed of $\log n$ bits is necessary to extract that randomness.

[4]One might say that the randomness is "Einstein certifiable", in the sense that the tests should convince even a quantum skeptic — viz, Einstein's famous quote from his 1926 letter to Max Born expressing his unhappiness with quantum mechanics as "God does not play dice with the Universe".

THEOREM 2 (INFORMAL). *Let $\varepsilon > 0$ be given, and n an integer. Let $(\mathcal{A}, \mathcal{B})$ be an arbitrary pair of non-communicating boxes. Suppose Protocol A, described in Figure 2, is executed with $\mathcal{A}$ and $\mathcal{B}$. This execution requires the use of $\tilde{O}(\log n \log(1/\varepsilon))$ random bits, and results in an output string B of $O(n)$ bits. Let* CHSH *be the event that the boxes' outputs are accepted in the final test described in the protocol, and suppose that* $\Pr\left(\text{CHSH}\right) \geq \varepsilon$. *Then*

$$H_\infty^\varepsilon(B|\text{CHSH}) \geq n.$$

In the simplified setting of Theorem 2 we are able to explicitly work through the constants involved in our bounds, and obtain actual estimates for the amount of randomness produced. For example, if one sets an error tolerance parameter $\varepsilon = 10^{-5}$, 15Kb of seed are necessary to produce roughly 15Kb of min-entropy: a shorter seed will only produce less entropy than it contained. Once that threshold is passed, however, the exponential expansion kicks in very quickly, and 30Kb of seed are already sufficient to generate one Terabyte of min-entropy!

Techniques.

The proofs of both our results proceed by contradiction. Suppose given a pair of boxes $(\mathcal{A}, \mathcal{B})$ that violate either theorem's conclusions: when subjected to an interaction as prescribed by Protocol A or Protocol B, the boxes produce bits that pass the user's test with good probability, but still contain little min-entropy. Our goal is to show that such boxes must violate the theorems' assumptions: they must be signaling.

In order to demonstrate this we introduce two ingredients. The first is a decomposition of the protocol into *phases*, which are consecutive sequences of a fixed number $k = O(\log n + \log 1/\varepsilon)$ of rounds of interaction between the user and the boxes. Each box always receives identical inputs throughout all rounds in a given phase. The purpose of the decomposition into phases is to enable the user to perform a robust verification of the CHSH condition: his final test will enforce that in *every* phase, a significantly larger than 3/4 fraction of pairs of outputs satisfy the CHSH condition (with respect to the corresponding pair of outputs). This lets us argue about the *Hamming distance* between Alice and Bob's k-bit outputs in any phase: if the CHSH constraint was of the form $a \oplus b = 0$ then the outputs should be close in Hamming distance, whereas if it had the form $a \oplus b = 1$ then they should be far apart.

The second ingredient builds upon the first. Consider the following simple *guessing game*: two players, Alice and Bob, each get a uniformly random bit as input. They win the game if Alice outputs a bit that equals Bob's input. Clearly, any strategy with success probability larger than 1/2 must involve communication between Alice and Bob. Ignoring quantum adversaries for now, suppose given a pair of boxes violating the conclusion of Theorem 2. Then such boxes can be used to devise a successful strategy in the guessing game: a contradiction of the no-signaling assumption. The main point is that if box $\mathcal{B}$'s output is not random enough, then in a certain block of the protocol it is likely to output a particular k-bit string almost deterministically. In that case, by using the CHSH condition Alice, given access to $\mathcal{A}$, can guess $\mathcal{B}$'s input $y \in \{0,1\}$ based on whether $\mathcal{A}$'s k-bit output in that block is "close" or "far" in Hamming distance from that particular string. This provides a way for Alice to guess $\mathcal{B}$'s input with probability greater than 1/2, violating the no-signaling condition placed on the boxes. This style of reasoning can be used to establish that $\mathcal{B}$'s output must have high min-entropy, thus yielding Theorem 6.

Proving security in the presence of a quantum adversary involves additional challenges. Indeed, indication that dealing with such adversaries may present substantial new difficulties may already be found in the area of strong extractor constructions: there are examples of such constructions, secure against classical adversaries, that dramatically fail in the presence of quantum adversaries with even smaller prior information [12]. We need to rule out the following catastrophic scenario: the adversary Eve inserted an undetectable "back-door" by entangling $\mathcal{A}$ and $\mathcal{B}$ together with her own, private, laboratory. Eve knows how the protocol proceeds, and how the boxes will behave (her only unknown are the $\text{poly}\log n$ random bits of seed used by the user). Based on this she repeatedly makes a specific measurement on her system, which reliably produces the same output bits as $\mathcal{A}$ and $\mathcal{B}$: while $\mathcal{B}$'s outputs may appear random in isolation, they are totally insecure![5]

Interestingly, the proof of Theorem 1 makes crucial use of the properties of a specific construction of a quantum-proof extractor, based on Trevisan's construction and the t-XOR code, that was first outlined in [9]. This construction is used to prove a key information-theoretic lemma, which we state informally below. The lemma gives an operational interpretation to a random variable having small smooth min-entropy conditioned on a quantum system, and may be of independent interest.

LEMMA 3 (INFORMAL). *Let X be a random variable distributed over m-bit strings, and for every $x \in \{0,1\}^m$, ρ_x a quantum state. Let $\varepsilon = \Omega(\text{poly}^{-1}(m))$, and suppose that the smooth min-entropy of X, given ρ_X, is $K = H_\infty^\varepsilon(X|\rho_X)$. Then there exists a collection of $O(K \log m)$ subsets of $O(\log m)$ positions of X each, and a measurement on ρ_X, depending on the parities of the bits of X in each of those subsets, that produces a string Y such that with inverse-polynomial probability, Y agrees with X in a fraction at least $1 - \frac{1}{\log m}$ of positions.*

Lemma 3 formalizes the intuition that if $H_\infty^\varepsilon(X|\rho_X) = K$, then given access to ρ_X one is only "missing" roughly K bits of information about X: there exists specific "advice bits" (the parities of the bits of X in each of the $O(K \log m)$ subsets) such that, given these advice bits, one can measure ρ_X and recover most of X with inverse-polynomial success probability.[6] The proof of Lemma 3

[5] This is so even after the application of the extractor by the user, as if Eve knows the whole input to the extractor, but not the seed, she still has about $n - O(\text{poly}\log n)$ bits of information on the n output bits.

[6] Note that in the range of large K (at least inverse-polynomial in m), an inverse-polynomial success is much higher than the inverse-exponential probability of guessing correctly the whole of X that one would get by measuring ρ_X directly, without using any "advice" bits.

is based directly on the proof of security of Trevisan's extractor against quantum adversaries presented in [8]. Since however it does not follow as a black-box, we give a detailed outline of the proof of the lemma in Appendix A.

Finally, note that while the *proof* of Lemma 3 requires the use of a specific construction of an extractor secure against quantum adversaries, as a result we can show that the bits output in Protocol B contain a large amount of conditional min-entropy — and hence *any* good quantum-proof extractor can be applied to those bits in order to obtain near-uniform bits.

Related work.

Two concurrent and independent papers, the first by Fehr, Gelles and Schaffner [11] and the second by Pironio and Massar [21], consider the problem of randomness expansion in the presence of *classical* adversaries. Both papers give a more rigorous proof of the results originally presented in [20], in particular concluding that the protocol introduced in that paper is composable, provided the adversary only possesses classical side information. Using this they are able to obtain a protocol with exponential randomness expansion; however their protocol requires the use of at least two pairs of boxes that are not entangled (in contrast, our protocol requires a single pair of boxes, and is secure even in the presence of arbitrary entanglement between them and any potential adversary).

Recent work by Colbeck and Renner [7] studies a related question, that of improving the quality of a given source of weak randomness. Specifically, they show that if one is given access to a so-called Santha-Vazirani source then one can produce bits that are guaranteed to be statistically close to uniform by using the violation of a specific Bell inequality by a pair of untrusted no-signaling devices.

Organization of the paper.

We begin with some preliminaries in Section 2. In Section 3 we introduce the *guessing game*, an important conceptual tool in the proofs of both Theorem 6 and Theorem 8. In Section 4 we prove Theorem 6, while Theorem 8 is proven in Section 5. The proof of Lemma 3 mostly follows from known results, and is given in Appendix A.

2. PRELIMINARIES

Notation.

Given two n-bit strings x, y we let $d_H(x,y) = \frac{1}{n}\sum_{i=1}^{n} |x_i - y_i|$ denote their relative Hamming distance. For $i \in [n]$, we let x_i be the i-th bit of x, and $x_{<i}$ its $(i-1)$-bit prefix.

Classical random variables.

Given a random variable $X \in \{0,1\}^n$, its min-entropy is

$$H_\infty(X) = -\log \max_x \Pr(X = x).$$

For two distributions p, q on a domain D, their statistical distance is

$$\|p - q\|_1 := (1/2)\sum_{x \in D} |p(x) - q(x)|_1.$$

This notion of distance can be extended to random variables with the same range in the natural way. Given $\varepsilon > 0$, the smooth min-entropy of a random variable X is

$$H_\infty^\varepsilon(X) = \sup_{Y,\, \|Y-X\|_1 \le \varepsilon} H_\infty(Y).$$

The following simple and well-known claim will be useful.

CLAIM 4. *Let $\alpha, \varepsilon > 0$ and X a random variable such that $H_\infty^\varepsilon(X) \le \alpha$. Then there exists a set B such that $\Pr(X \in B) \ge \varepsilon$ and for every $x \in B$, it holds that $\Pr(X = x) \ge 2^{-\alpha}$.*

PROOF. Let B be the set of x such that $\Pr(X = x) \ge 2^{-\alpha}$, and suppose $\Pr(X \in B) < \varepsilon$. Define Y so that $\Pr(Y = x) = \Pr(X = x)$ for every $x \notin B$, $\Pr(Y = x) = 0$ for every $x \in B$. In order to normalize Y, introduce new values z such that $\Pr(X = z) = 0$, and extend Y by defining $\Pr(Y = z) = 2^{-\alpha-1}$ until it is properly normalized. Then $\|Y - X\|_1 < \varepsilon$ and $H_\infty(Y) > \alpha$, contradicting the assumption on the smooth min-entropy of X. □

Quantum states.

Let X be a register containing a classical random variable, which we also call X, and E a register containing a quantum state, possibly correlated to X. Then the whole system can be described using the cq-state (cq stands for classical-quantum) $\rho_{XE} = \sum_x p_X(x)|x\rangle\langle x| \otimes \rho_x$, where for every x ρ_x is a density matrix, i.e. a positive matrix with trace 1. Given such a state, the guessing entropy $p_{guess}(X|E)$ is the maximum probability with which one can predict X, given access to E. Formally, it is defined as

$$p_{guess}(X|E)_\rho = \sup_{\{M_x\}} \sum_x p_X(x)\mathrm{Tr}\big(M_x \rho_x\big),$$

where the supremum is taken over all projective operator-valued measurements (POVMs) on E.[7] The conditional min-entropy can be defined through the guessing entropy as $H_\infty(X|E)_\rho = -\log p_{guess}(X|E)_\rho$ [17]. We will often omit the subscript ρ, when the underlying state is clear. The appropriate distance measure on quantum states is the trace distance, which derives from the trace norm $\|A\|_{tr} = \mathrm{Tr}\big(\sqrt{A^\dagger A}\big)$. This lets us define a notion of smooth conditional min-entropy: $H_\infty^\varepsilon(X|E)_\rho = \sup_{\sigma_{XE},\, \|\sigma_{XE} - \rho_{XE}\|_{tr} \le \varepsilon} H_\infty(X|E)_\sigma$, where here the supremum is taken over all sub-normalized cq-state σ_{XE}. As in the purely classical setting, it is known that this measure of conditional min-entropy is the appropriate one from the point of view of extracting uniform bits [22]: if $H_\infty^\varepsilon(X|E) = K$ then $K - O(\log 1/\varepsilon)$ bits can be extracted from X that are ε-close to uniform, even from the point of view of E.

The CHSH game.

The following game was originally introduced by Clause, Horne, Shimony and Holt [4] to demonstrate the non-locality of quantum mechanics. Two collaborating but non-communicating parties, Alice and Bob, are each given a bit $x, y \in \{0,1\}$ distributed uniformly at random. Their

[7] A POVM $\{M_x\}$ is given by a set of positive matrices which sum to identity. We refer the reader to the standard textbook [19] for more details on the basics of quantum information theory.

goal is to produce bits a, b respectively such that $a \oplus b = x \wedge y$. It is not hard to see that classical parties (possibly using shared randomness) have a maximum success probability of 3/4 in this game. In contrast, quantum mechanics predicts that the following strategy, which we will sometimes refer to as the "honest" strategy, achieves a success probability of $\cos^2(\pi/8) \approx 0.85$. Alice and Bob share an EPR pair $|\Psi\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. Upon receiving her input, Alice measures either in the computational ($x = 0$) or the Hadamard ($x = 1$) basis. Bob measures in the computational basis rotated by either $\pi/8$ ($y = 0$) or $3\pi/8$ ($y = 1$). One can then verify that, for every pair of inputs (x, y), this strategy produces a pair of correct outputs with probability exactly $\cos^2(\pi/8)$.

3. THE GUESSING GAME

Consider the following simple guessing game. In this game, there are two cooperating players, Alice and Bob. At the start of the game Bob receives a single bit $y \in \{0, 1\}$ chosen uniformly at random. The players are then allowed to perform arbitrary computations, but are not allowed to communicate. At the end of the game Alice outputs a bit a, and the players win if $a = y$.

Clearly, any strategy with success probability larger than $\frac{1}{2}$ indicates a violation of the no-communication assumption between Alice and Bob. At the heart of the proofs of both Theorem 6 and Theorem 8 is a reduction to the guessing game. Assuming there existed a pair of boxes violating the conclusions of either theorem, we will show how these boxes may be used to devise a successful strategy in the guessing game, contradicting the no-signaling assumption placed on the boxes.

To illustrate the main features of the strategies we will design later, consider the following simplified setting. Let $\mathcal{A}, \mathcal{B}$ be a given pair of boxes taking inputs $X, Y \in \{0, 1\}$ and producing outputs $A, B \in \{0, 1\}^k$ respectively, where k is a parameter. Assume the following two properties hold. First, if the input to $\mathcal{B}$ is $Y = 0$ then its output B is essentially deterministic, in the sense that $B = b_0$ with high probability. Second, whatever their inputs, the boxes' outputs satisfy the CHSH constraint on average with a slightly higher probability than could any classical boxes: there is a fixed $\delta > 0$ such that a fraction at least $3/4 + \delta$ of $i \in [k]$ are such that $A_i \oplus B_i = X \wedge Y$. Then we claim that there is a strategy for Alice and Bob in the guessing game, using $\mathcal{A}$ and $\mathcal{B}$, that succeeds with probability strictly larger than 1/2.

Alice and Bob's strategy is the following. Alice is given access to $\mathcal{A}$ and Bob to $\mathcal{B}$. Upon receiving his secret bit y, Bob inputs it to $\mathcal{B}$, collecting outputs $b \in \{0, 1\}^k$. Alice chooses an $x \in \{0, 1\}$ uniformly at random, and inputs it to $\mathcal{A}$, collecting outputs $a \in \{0, 1\}^k$. Let b_0 be the k-bit string with the highest probability of being output by $\mathcal{B}$, conditioned on $y = 0$. Alice makes a decision as follows: she computes the relative Hamming distance $d = d_H(a, b_0)$. If $d < 1/4$ she claims "Bob's input was 0". Otherwise, she claims "Bob's input was 1".

By assumption, if Bob's secret bit was $y = 0$, then his output is almost certainly b_0. By the CHSH constraint, independently of her input Alice's output a lies in a Hamming ball of radius $1/4 - \delta$ around b_0. So in this case she correctly claims "Bob's input was 0".

In the case that Bob's secret bit was $y = 1$, the analysis is more interesting. Let b be the actual output of $\mathcal{B}$. Let a_0 and a_1 be $\mathcal{A}$'s output in the two cases $x = 0$ and $x = 1$ respectively. We claim that the Hamming distance $d_H(a_0, a_1) \geq 1/2 + 2\delta$. This is because by the CHSH constraint, $d_H(a_0, b) \leq 1/4 - \delta$, while $d_H(a_1, b) \geq 3/4 + \delta$. Applying the triangle inequality

$$d_H(a_0, a_1) \geq \big|d_H(a_1, b) - d_H(a_0, b)\big| \geq 1/2 + 2\delta,$$

as claimed. Hence both a_0 and a_1 cannot lie in the Hamming ball of radius 1/4 around the fixed string b_0 (observe that this argument makes no use of the actual location of b!). Thus in the case $y = 1$, Alice correctly claims "Bob's input was 1" with probability at least 1/2.

Overall Alice and Bob succeed in the guessing game with probability at least 3/4, implying the boxes $\mathcal{A}$, $\mathcal{B}$ allowed them to communicate, and hence do not satisfy the no-signaling condition.

Clearly there is a lot of slack in the above reasoning, since for contradiction it suffices to succeed in the guessing game with any probability strictly greater than 1/2. By being more careful it is possible to allow Bob's output on $y = 0$ to not be fully deterministic, as well as allow for a small probability that the boxes' outputs may not satisfy the CHSH constraint:

LEMMA 5. *Let $\beta, \gamma > 0$ be such that $\gamma/2 + 3\beta < 1/4$, and k an integer. Suppose given a pair of boxes $\mathcal{A}, \mathcal{B}$, taking inputs $X, Y \in \{0, 1\}$ and producing outputs $A, B \in \{0, 1\}^k$ each. Suppose the following conditions hold:*

1. *When given input 0, the distribution of outputs of $\mathcal{B}$ has low min-entropy: there exists a $b_0 \in \{0, 1\}^k$ such that $\Pr(B = b_0 | Y = 0) \geq 1 - \gamma$,*
2. *The boxes' outputs fall in the "quantum regime" of the CHSH inequality: there exists a constant $\delta > 0$ such that*

$$\Pr\big(d_H\big(A \oplus B, (X \wedge Y, \ldots, X \wedge Y)\big) > 1/4 - \delta\big) \leq \beta,$$

where the probability is taken over the choice of uniformly random X, Y, and the boxes' internal randomness.

Then there is a strategy for Alice and Bob, using $\mathcal{A}$ and $\mathcal{B}$, which gives them success probability strictly greater than 1/2 in the guessing game.

PROOF. Alice and Bob's strategy in the guessing game is as described above. Let b_0 be the k-bit string that is most likely to be output by $\mathcal{B}$, conditioned on $y = 0$.

We first show that, if Bob's input was $y = 0$, then Alice claims that Bob had a 0 with probability at least $1 - \gamma - 2\beta$. By the first condition in the lemma, Bob obtains the output b_0 with probability at least $1 - \gamma$. Moreover, by the second condition the CHSH constraint will be satisfied with probability at least $1 - 2\beta$ on average over Alice's choice of input, given that Bob's input was $y = 0$. Given $y = 0$, whatever the input to $\mathcal{A}$ the CHSH constraint implies that $d_H(a, b) < 1/4$. Hence by a union bound Alice will obtain an output string a at relative Hamming

distance at most $1/4$ from b_0 with probability at least $1-\gamma-2\beta$.

Next we show that, in case Bob's input in the guessing game is $y=1$, Alice claims that Bob had a 1 with probability at least $\frac{1}{2}\big(1-8\beta\big)$. The second condition in the lemma implies that for any of the two possible choices for Alice's input $X=x\in\{0,1\}$, it holds that

$$\Pr_{ABY}\Big(\, d_H\big(A\oplus B,(X\wedge Y),\ldots,X\wedge Y)\big) > \frac{1}{4}-\delta \,|X=x\Big) \leq 2\beta. \quad (1)$$

Let b' be Bob's output, and suppose that b' is such that for every $x\in\{0,1\}$, (1) holds conditioned on $B=b'$, with the 2β replaced by a 4β. It follows from (1) and Markov's inequality that this condition holds with probability at least $1-4\beta$ over b'.

If Alice chooses $x=0$ then the CHSH constraint indicates that the corresponding a_0 should be such that $d_H(a_0,b')<1/4-\delta$, while in case she chooses $x=1$ her output a_1 should satisfy $d_H(a_1,b')>3/4+\delta$. By the triangle inequality,

$$d_H(a_0,a_1) \;\geq\; \big|d_H(a_1,b')-d_H(a_0,b')\big| \;>\; 1/2+2\delta,$$

so that whatever the value of b', at most one of a_0 or a_1 can be at distance less than $1/4$ from b_0. Since Alice's input is chosen uniformly at random, taking into account the choice of b' we have shown that with probability at least $(1-8\beta)/2$ Alice will choose an input that will make her correctly claim that Bob had a 1.

The two bounds proven above together show that Alice's probability of correctly guessing Bob's input in the guessing game is at least

$$p_{succ} \;\geq\; \frac{1}{2}\big(1-\gamma-2\beta\big)+\frac{1}{2}\frac{1-8\beta}{2} \;=\; \frac{1}{2}+\Big(\frac{1}{4}-3\beta-\frac{\gamma}{2}\Big),$$

which is greater than $1/2$ whenever $3\beta+\gamma/2<1/4$, proving Lemma 5. □

4. A PROTOCOL WITH EXPONENTIAL RANDOMNESS EXPANSION

In this section we prove Theorem 2, which can be stated formally as follows.

THEOREM 6. *Let $\varepsilon>0$ be given, and n an integer. Let $(\mathcal{A},\mathcal{B})$ be an arbitrary pair of no-signaling boxes used to execute Protocol A, as described in Figure 2, B the random variable describing the bits output by $\mathcal{B}$, and* CHSH *the event that the boxes' outputs are accepted in the final test described in the protocol. Then for all large enough n at least one of the following holds :*

- *Either* $H_\infty^\varepsilon(B|\mathrm{CHSH})\geq n$,
- *Or* $\Pr\big(\mathrm{CHSH}\big)\leq\varepsilon$.

Moreover, inputs in Protocol A can be generated using $\tilde{O}(\log n\log(1/\varepsilon))$ uniformly random bits, and it makes $O(n(\log n+\log(1/\varepsilon)))$ uses of the boxes.

Protocol A is described in Figure 2. It uses two main ideas in order to save on the randomness used by the user to select inputs to the boxes. The first idea is to restrict

Protocol A

1. Let n,ε be parameters given as input. Set $m=500n$, $\Delta=200\lceil\ln(1/\varepsilon)\rceil$, $\ell=m/\Delta$ and $k=100\lceil\log n+\log 1/\varepsilon\rceil$.
2. Choose $T\subseteq[m]$ uniformly at random by selecting each position independently with probability $1/\ell$.
3. Repeat, for $i=1,\ldots,m$:
   - 3.1 If $i\notin T$, then
     - 3.1.1 Set $x=y=0$ and choose x,y as inputs for k consecutive steps. Collect outputs $a,b\in\{0,1\}^k$.
     - 3.1.2 If $a\oplus b$ has more than $\lceil 0.2k\rceil$ 1's then reject and abort the protocol. Otherwise, continue.
   - 3.2 If $i\in T$,
     - 3.2.1 Pick $x,y\in\{0,1\}$ uniformly at random, and set x,y as inputs for k consecutive steps. Collect outputs $a,b\in\{0,1\}^k$.
     - 3.2.2 If $a\oplus b$ differs from $x\wedge y$ in more than $\lceil 0.2k\rceil$ positions then reject and abort the protocol. Otherwise, continue.
4. If all steps accepted, then accept.

Figure 2: Protocol A uses $O(\log n\log 1/\varepsilon)$ bits of randomness and makes $O(n(\log n+\log 1/\varepsilon))$ uses of the boxes. Theorem 6 (in Section 4) shows that n bits of randomness are produced with confidence $1-\varepsilon$. The threshold $0.2k$ in steps 3.1.2 and 3.2.2 is arbitrary, and any value strictly lower than $k/4$ would work.

the inputs to $(0,0)$ most of the time.[8] Only a few randomly placed checks (the Bell blocks) are performed in order to verify that the boxes are generating their inputs honestly. There are about $O(\log 1/\varepsilon)$ such blocks. Note that the boxes usually do not know when they are being checked: for instant, if the input is $(0,1)$ then even though box $\mathcal{B}$ knows that it is in a Bell round, box $\mathcal{A}$ by itself cannot differentiate that particular round from one in which both inputs are 0. This implies in particular that the strategy it uses to determine its output cannot be different from what it would have been had the inputs been the more frequent $(0,0)$.

The second main idea, as already explained in the introduction, consists in decomposing the protocol into *phases* (we also use *blocks* when specifically referring to the sequence of inputs or outputs in a given phase), which are consecutive sequences of a fixed number $k = O(\log n + \log 1/\varepsilon)$ of rounds of interaction between the user and the boxes, and whose purpose of the decomposition in phases is to enable the user to perform a robust verification of the CHSH condition.

Altogether, Protocol A only requires the use of random bits in order to select the position of the Bell blocks, as well as to select inputs in these blocks. The $O(\log 1/\varepsilon)$ Bell blocks can be chosen among the $O(n)$ rounds using $\tilde{O}(\log n \log 1/\varepsilon)$ random bits (see e.g. [16]), and corresponding uniformly distributed inputs may be generated using an additional $O(\log 1/\varepsilon)$ random bits.

Before proceeding, we should verify that "honest" boxes, which play the optimal quantum strategy for the CHSH game independently in every round, are accepted by the user with high probability. Indeed, we have seen that such boxes will satisfy the CHSH constraint independently with probability $\cos^2 \pi/8$ in each round. Hence when one considers a block of k successive rounds, the probability that the CHSH constraint is *not* satisfied in more than 20% of those rounds will be exponentially small in k. Precisely, a simple Chernoff bound shows that the probability that the honest strategy satisfies the CHSH condition in less than 80% of any k successive rounds is at most $\exp\big(-(\cos^2 \pi/8 - 0.80)^2 k/2\big)$. Given our choice of $k = 100\lceil \log n + \log 1/\varepsilon \rceil$, it can be verified that for large enough n this expression is smaller than ε/m, where $m = 500n$ is the total number of blocks in the protocol. By a union bound, such boxes will fail to produce correlations satisfying the user in even just one of these blocks with probability at most ε.

Modeling events in the protocol.

Let $x = (x_i), y = (y_i), a = (a_i), b = (b_i) \in \big(\{0,1\}^k\big)^m$ denote the boxes' respective input and output strings in an execution of Protocol A, as described in Figure 2. Let X, Y, A, B be the corresponding random variables. The boxes' behavior is characterized by a probability distribution $p_{AB|XY}(a,b|x,y)$. The iterative structure of the protocol implies that $p_{AB|XY}$ can be factored as follows:

$$p_{AB|XY}(a,b|x,y) = \prod_{i=1}^{mk} p_{A_iB_i|X_iY_iH_i}(a_i,b_i|x_i,y_i,h_i),$$

where for any $i \in [mk]$, $H_i = (A_{<i}, B_{<i}, X_{<i}, Y_{<i})$ and $h_i = (a_{<i}, b_{<i}, x_{<i}, y_{<i})$. We impose a single additional condition on $p_{AB|XY}$: that it obeys the no-signaling condition in every block, that is for every k-round block $S_i \subseteq \{0,1\}^{mk}$, where $i \in [m]$, $a_{S_i}, x_{S_i}, y_{S_i}, y'_{S_i}$ and h_{S_i}, it holds that

$$\sum_{b_{S_i} \in \{0,1\}^k} p_{A_{S_i}B_{S_i}|Z_i}(a_{S_i}, b_{S_i}|z_i) = \sum_{b_{S_i} \in \{0,1\}^k} p_{A_{S_i}B_{S_i}|Z_i}(a_{S_i}, b_{S_i}|z'_i),$$

where we used $Z_i = X_{(k-1)i+1}Y_{(k-1)i+1}H_{(k-1)i+1}$ and $z_i = x_{(k-1)i+1}, y_{(k-1)i+1}, h_{(k-1)i+1}$ and $z'_i = x_{(k-1)i+1}, y'_{(k-1)i+1}, h_{(k-1)i+1}$ as shorthands, and a symmetric condition holds when marginalizing over a_{S_i}. For $i \in [m]$, let CHSH_i be the event that $d_H(A_{S_i} \oplus B_{S_i}, X_{S_i} \wedge Y_{S_i}) \leq 0.2$, and $\mathrm{CHSH} = \bigwedge_i \mathrm{CHSH}_i$. We will also use the shorthand $\mathrm{CHSH}_{<i} = \bigwedge_{j<i} \mathrm{CHSH}_j$. Finally, we let $T_j \in [m]$ be a random variable denoting the j-th Bell block, i.e. the j-th element of the set T chosen by the user in step 2. of Protocol A.

Claim 7. *Let n be an integer, and $2^{-n/10} < \varepsilon < 1/100$. Suppose that both conditions (i) $H^{\varepsilon}_{\infty}(B|CHSH) \leq n$, and (ii) $\Pr(\mathrm{CHSH}) \geq \varepsilon$ hold. Then for all large enough n there exists an index j_0 and a subset G of output strings satisfying $\Pr(B \in G) \geq \varepsilon^3$ such that the following hold.*

- *Conditioned on $\mathcal{B}$'s input in the j_0-th Bell block T_{j_0} being 0, its output in that block is essentially deterministic: $\forall b \in G$,*

$$\Pr(B_{T_{j_0}} = b_{T_{j_0}} | \mathrm{CHSH}_{<T_{j_0}}, B_{<T_{j_0}} = b_{<T_{j_0}}, Y_{T_{j_0}} = 0) \geq 0.92, \quad (2)$$

- *The CHSH condition is satisfied with high probability in the j_0-th Bell block T_{j_0}: $\forall b \in G$,*

$$\Pr(\mathrm{CHSH}_{T_{j_0}} | \mathrm{CHSH}_{<T_{j_0}}, B_{<T_{j_0}} = b_{<T_{j_0}}) \geq 0.95. \quad (3)$$

Proof. As in Protocol A, set $m = 500n$, $\ell = m/\Delta$ and $\Delta = 200\lceil \ln(1/\varepsilon) \rceil$. Let BAD be the set of strings $b \in \big(\{0,1\}^k\big)^m$ such that $\Pr(B = b|\mathrm{CHSH}) > 2^{-n}$. Assumption (i) together with Claim 4 show that $\Pr(\mathrm{BAD}|\mathrm{CHSH}) \geq \varepsilon$. Using (ii) and Baye's rule we get that for every $b = (b_1, \ldots, b_m) \in \mathrm{BAD}$,

$$\begin{aligned} &\Pr(B = b, \mathrm{CHSH}) \\ &= \prod_{i=1}^{m} \Pr(B_i = b_i, \mathrm{CHSH}_i | \mathrm{CHSH}_{<i}, B_{<i} = b_{<i}) \\ &> 2^{-n}\varepsilon. \end{aligned}$$

[8] This idea was already used in [20], and led to their protocol with quadratic $\sqrt{n \log 1/\varepsilon} \to n$ expansion of randomness.

Taking logarithms on both sides,

$$\begin{aligned}&\sum_{i=1}^{m} -\log \Pr(B_i = b_i, \mathrm{CHSH}_i | \mathrm{CHSH}_{<i}, B_{<i} = b_{<i}) \\ &< n + \log(1/\varepsilon) \\ &\leq (1 + 1/10)n,\end{aligned}$$

assuming as in the statement of the claim that ε is not too small. By an averaging argument at least 9/10 of all $i \in [m]$ are such that a fraction at least 6/10 (in probability) of all $b \in \mathrm{BAD}$ are such that

$$\begin{aligned}\Pr(B_i = b_i | \mathrm{CHSH}_{<i}, B_{<i} = b_{<i}) &\\ \geq 2^{-(100/4)(1+1/10)(n/m)} &\\ \geq 2^{-28/500} \geq 0.96.& \qquad (4)\end{aligned}$$

Since in the protocol Bob's input is a 0 with probability at least 1/2 irrespective of the type of block, we may ensure that (4) holds (with a slightly smaller probability) conditioned on $Y_i = 0$, an event that is independent from both $\mathrm{CHSH}_{<i}$ and $B_{<i} = b_{<i}$ (for any $b_{<}i$):

$$\Pr(B_i = b_i | \mathrm{CHSH}_{<i}, B_{<i} = b_{<i}, Y_i = 0) \geq 0.92. \qquad (5)$$

Let S be the set of $i \in [m]$ such that (5) holds for a fraction at least 6/10 of $b \in \mathrm{BAD}$. S is a random variable of size $|S| \geq 9m/10$.

We apply the same reasoning once more, focusing on the CHSH constraint being satisfied in a Bell block. Let T be a random variable containing the indices of the blocks that have been designated as Bell blocks in the protocol. Let $N = |T \cap S|$. We may write N as the sum of Boolean random variables N_j, where $N_j = 1$ if and only if the j-th element of S falls in T. Since for every i the i-th block is chosen to be a Bell block independently with probability $1/\ell$ (independently of past events such as $\mathrm{CHSH}_{<i}$ and $B_{<i} = b_{<i}$), the random variables N_j, for $j \leq |S|$, are independent. Recall that $|S| \geq 9m/10$, and by a Chernoff bound

$$\begin{aligned}\Pr\left(N_1 + \cdots + N_{9m/10} \geq 9m/10 \cdot 1/(2\ell)\right) &\\ \geq 1 - e^{-(9m/10\ell)(1/2)^2/2} &\\ \geq 1 - e^{-\Delta/10} \geq 1 - \varepsilon^3,&\end{aligned}$$

given that $m/\ell = \Delta$. Let K denote the event that this bound holds: $\Pr(K) \geq 1 - \varepsilon^3$, and conditioned on K it holds that $N = |S \cap T| \geq 9m/(20\ell) \geq 9\Delta/20$. Starting from $\Pr(\mathrm{CHSH}|\mathrm{BAD}) \geq \varepsilon^2$, further conditioning on K gives

$$\begin{aligned}\Pr(\mathrm{CHSH}|\mathrm{BAD}, K) &= \frac{\Pr(\mathrm{CHSH}, K|\mathrm{BAD})}{\Pr(K|\mathrm{BAD})} \\ &\geq \varepsilon^2 - \varepsilon^3 \geq \varepsilon^2/2.\end{aligned}$$

Using Baye's rule as before we then obtain

$$\begin{aligned}\sum_{i \in T \cap S} -\log \Pr(\mathrm{CHSH}_i | \mathrm{CHSH}_{<i}, \mathrm{BAD}_{<i}, K) &\\ \leq 2\log(2/\varepsilon).&\end{aligned}$$

Using the lower bound on N this implies that there exists an $i \in T \cap S$ such that

$$\begin{aligned}\Pr(\mathrm{CHSH}_i | \mathrm{CHSH}_{<i}, \mathrm{BAD}_{<i}, K) &\geq 2^{-2\log(2/\varepsilon)/N} \\ &\geq 0.978, \qquad (6)\end{aligned}$$

given the choice of Δ made in the claim. Given our assumption on ε, removing the conditioning on K in this equation at most decreases the lower bound to 0.975. Let $i \in T \cap S$ be a Bell block for which (6) holds. By Markov's inequality, for a fraction at least 1/2 of $b \in \mathrm{BAD}$ it holds that

$$\Pr(\mathrm{CHSH}_i | \mathrm{CHSH}_{<i}, B_{<i} = b_{<i}) \geq 0.95. \qquad (7)$$

By the union bound, at iteration i (7) will hold simultaneously with (5) for a subset G of BAD of size at least

$$\Pr(G) = \Pr(G|\mathrm{BAD}) \Pr(\mathrm{BAD}) \geq \big(6/10 - 1/2\big)\varepsilon^2 \geq \varepsilon^3,$$

given our choice of parameters. Eq. (7) implies (3) in the claim, and (5) implies (2). □

In order to conclude the proof of Theorem 6 it remains to show how the special block identified in Claim 7 can lead to a successful strategy in the guessing game.

Consider the following strategy for Alice and Bob in the guessing game. In a preparatory phase (before Bob receives his secret bit y), Alice and Bob run the protocol with the boxes $\mathcal{A}$ and $\mathcal{B}$, up to the T_{j_0}-th block (excluded). Bob communicates $\mathcal{B}$'s outputs up till that block to Alice. Together they check that the CHSH constraint is satisfied in all blocks preceding the T_{j_0}-th; if not they abort. They also verify that Bob's outputs are the prefix of a string $b \in G$; if not they abort. The guessing game can now start: Alice and Bob are separated and Bob is given his secret input y.

Given the conditioning that Alice and Bob have performed, once they are ready to start the game boxes $\mathcal{A}$ and $\mathcal{B}$ satisfy both conditions of Lemma 5. Condition 1. in Lemma 5 holds with $\gamma = 0.08$ as a consequence of item 1 in Claim 7 and condition 2 in Lemma 5 follows from item 2 in Claim 7 with $\beta = 0.05$. Since $\gamma/2 + 3\beta = 0.19 < 1/4$, Lemma 5 lets us conclude that the boxes $\mathcal{A}$ and $\mathcal{B}$ must be signaling in the T_{j_0}-th block, a contradiction. This finishes the proof of Theorem 6.

5. SECURITY IN THE PRESENCE OF A QUANTUM ADVERSARY

In this section we prove our main theorem, which can be formally stated as follows.

THEOREM 8. *Let n and $\alpha > 0$ be given, and set $\varepsilon = n^{-\alpha}$. Let $(\mathcal{A}, \mathcal{B})$ be an arbitrary pair of no-signaling boxes used to execute Protocol B, described in Figure 3, and assume that $\mathcal{A}$ and $\mathcal{B}$ can be described by two isolated but possibly entangled quantum-mechanical systems. Let* CHSH *be the event that the boxes' outputs are accepted in the protocol, and B' the random variable describing the bits output by $\mathcal{B}$, conditioned on* CHSH. *Let $\mathbf{E}$ denote an arbitrary quantum system, possibly entangled with $\mathcal{A}$ and $\mathcal{B}$, but such that no communication occurs between $\mathcal{A}, \mathcal{B}$ and $\mathbf{E}$ once the protocol starts. Then for all large enough n at least one of the following holds:*

- *Either $H_\infty^\varepsilon(B'|\mathbf{E}) \geq n$,*

Protocol B

1. Let n and $\alpha > 0$ be given as input. Set $\ell = n^{10+8\alpha}$, $k = 100\lceil \log \ell + \log 1/\varepsilon \rceil$ and $m = \lceil C\ell \log^2 \ell \rceil$, where $C > 0$ is a large constant.
2. Choose $T \subseteq [m]$ uniformly at random by selecting each position independently with probability $1/\ell$.
3. Repeat, for $i = 1, \ldots, m$:
   - 3.1 If $i \notin T$, then
     - 3.1.1 Set $x = y = (A, 0)$ and choose x, y as inputs for k consecutive steps. Collect outputs $a, b \in \{0,1\}^k$.
     - 3.1.2 If $a \neq b$ then reject and abort the protocol. Otherwise, continue.
   - 3.2 If $i \in T$,
     - 3.2.1 Pick $x \in \{(A,0),(A,1)\}$ and $y \in \{(A,0),(B,0)\}$ uniformly at random, and set x, y as inputs for k consecutive steps. Collect outputs $a, b \in \{0,1\}^k$.
     - 3.2.2 If either $a = b$ and $x = y$, or $d_H(a,b) \leq 0.16$ and $y = (B,0)$, or $d_H(a,b) \in [0.49, 0.51]$ and $x = (A,1)$ and $y = (A,0)$ then continue. Otherwise reject and abort the protocol.
4. If all steps accepted, then accept.

Figure 3: Protocol B uses $\tilde{O}(\log^3 n)$ bits of randomness and makes $\mathrm{poly}(n)$ uses of the boxes. Theorem 8 shows that n bits of randomness are produced, with confidence $\varepsilon = n^{-\alpha}$.

- *Or* $\Pr\big(\mathrm{CHSH}\big) \leq \varepsilon$.

Moreover, inputs in Protocol B may be generated using only $\tilde{O}(\log^3 n)$ bits of randomness.

We first give an overview of the proof, describing the main steps, in the next section. The formal proof is given in Section 5.3.

5.1 The protocol

Theorem 8 is based on Protocol B, a variant of Protocol A which replaces the use of the CHSH game by the following "extended" variant. In this game each box may receive one of four possible inputs, which we label as $(A,0),(A,1),(B,0),(B,1)$. An input such as "$(A,1)$" to either box means: "perform the measurement that $\mathcal{A}$ would have performed in the honest CHSH strategy, in case its input had been a 1". The advantage of working with this game is that there exists an optimal strategy (the one directly derived from the honest CHSH strategy) in which both players always output identical answers when their inputs are equal.

Protocol B follows the same structure as Protocol A. Inputs are divided into groups of $k = O(\log^2 n)$ identical inputs. There are $m = O(n^{10+8\alpha} \log^2 n)$ successive blocks. Each round of the protocol selects inputs to the boxes coming from the "extended CHSH" game. That game has four questions per party: $(A,0),(A,1),(B,0),(B,1)$. We expect honest boxes to apply the following strategy. They share a single EPR pair, and perform the same measurement if provided the same input. On input $(A,0)$ the measurement is in the computational basis $\{|0\rangle, |1\rangle\}$, and on input $(A,1)$ it is in the Hadamard basis $\{|+\rangle, |-\rangle\}$, with the outcome $|+\rangle$ being associated with the output '0'. On input $(B,0)$ the measurement is in the basis $\{\cos^2(\pi/8)|0\rangle + \sin^2(\pi/8)|1\rangle, \sin^2(\pi/8)|0\rangle - \cos^2(\pi/8)|1\rangle\}$, with the first vector being associated with the outcome '0'.

5.2 Proof overview

As in the proof of Theorem 6 we prove Theorem 8 by contradiction, through a reduction to the guessing game. In the non-adversarial case the crux of the reduction consisted in identifying a special block $j_0 \in [m]$ in which $\mathcal{B}$'s output B was essentially deterministic, conditioned on past outputs. In the adversarial setting, however, B may be perfectly uniform, and such a block may not exist. Instead, we start by assuming for contradiction that the min-entropy of Bob's output conditioned on Eve's information is small: $H_\infty^\varepsilon(B|\mathbf{E}) \leq n$.

Previously in the guessing game Alice tried to guess Bob's secret input $y \in \{0,1\}$. She did so by using her prediction for $\mathcal{B}$'s outputs, together with the CHSH constraint and her own box $\mathcal{A}$'s outputs. Here we team up Alice and Eve. Alice will provide Eve with some information she obtained in previous blocks of the protocol, and based on that information Eve will attempt to make an accurate prediction for $\mathcal{B}$'s outputs in the special block. Alice will then use that prediction to guess y, using as before the CHSH constraint and her own box $\mathcal{A}$'s outputs.

The reconstruction paradigm.

We would like to show that, under our assumption on $H_\infty^\varepsilon(B|\mathbf{E})$, Eve can perform the following task: accurately predict (part of) B, given auxiliary information provided by Alice. We accomplish this by using the "reconstruction" property of certain extractor constructions originally introduced by Trevisan [27]. Recall that an extractor is a function which maps a string B with large min-entropy (conditioned on side information contained in the quantum register $\mathbf{E}$) to a (shorter) string Z that is statistically close to uniform even from the point of view of an adversary holding $\mathbf{E}$. The reconstruction proof technique proceeds as follows: Suppose an adversary breaks the extractor. Then there exists another adversary who, given a small subset of the bits of the extractor's input as "advice", can reconstruct the *whole* input. Hence the input's entropy must have been at most the number of advice bits given.

For the purposes of constructing extractors, one would then take the contrapositive to conclude that, provided the input has large enough entropy, the extractor's output must be indistinguishable from uniform, thereby proving security. Here we work *directly* with the reconstruction procedure. Suppose that B has low min-entropy, conditioned on Eve's side information. If we were to apply an extractor to B in order to extract *more* bits than its conditional min-entropy, then certainly the output would

not be secure: Eve would be able to distinguish it from a uniformly random string. The reconstruction paradigm states that, as a consequence, there is a strategy for Eve that successfully predicts the *entire* string B, given a subset of its bits as advice — exactly what is needed from Eve to facilitate Alice's task in the guessing game.

The t-XOR extractor.

At this stage we are faced with two difficulties. The first is that the reconstruction paradigm was developed in the context of classical adversaries, who can repeat predictive measurements at will. Quantum information is more delicate, and may be modified by the act of measuring. The second has to do with the role of the advice bits: since they come from $\mathcal{B}$'s output B we need to ensure that, in the guessing game, Alice can indeed provide this auxiliary information to Eve, *without* communicating with Bob.

In order to solve both problems we focus on a specific extractor construction, the t-XOR extractor E_t (here t is an integer such that $t = O(\log^2 n)$). For our purposes it will suffice to think of E_t as mapping the mk-bit string B to a string of $r \ll n$ bits, each of which is the parity of a certain subset of t out of B's mk bits. Which parities is dictated by an extra argument to the extractor, its seed, based on the use of combinatorial designs. Formally,

$$\begin{aligned} E_t \,:\, \{0,1\}^{mk} \times \{0,1\}^s &\to \{0,1\}^r \\ (b,y) &\mapsto \big(C_t^1(b,y),\ldots,C_t^r(b,y)\big), \end{aligned}$$

where $C_t^i(b,y)$ is the parity of a specific subset of t bits of x, depending on both i and y.

Suppose that Eve can distinguish the output of the extractor $Z = E_t(B,Y)$ from a uniformly random string with success probability ε. In the first step of the reconstruction proof, a hybrid argument is used to show that Eve can predict the parity of t bits of B chosen at random with success ε/r, given access to the parities of $O(r)$ other subsets of t bits of B as advice. This step uses specific properties of the combinatorial designs.

The next step is the most critical. One would like to argue that, since Eve can predict the parity of a random subset of t of B's bits, she can recover a string that agrees with *most* of the t-XORs of B. One could then appeal to the approximate list-decoding properties of the t-XOR code in order to conclude that Eve may deduce a list of guesses for the string B itself. Since, however, Eve is quantum, the fact that she has a measurement predicting *any* t-XOR does not imply she has one predicting *every* t-XOR: measurements are destructive and distinct measurements need not be compatible. This is a fundamental difficulty, which arises e.g. in the analysis of random access codes [1]. To overcome it one has to appeal to a subtle argument due to Koenig and Terhal [18]. They show that without loss of generality one may assume that Eve's measurement has a specific form, called the *pretty-good measurement*. One can then argue that this specific measurement may be refined into one that predicts a guess for the whole list of t-XORs of B, from which a guess for B can be deduced by list-decoding the t-XOR code.

The security of the t-XOR extractor against quantum adversaries was first shown by Ta-Shma [25], and later improved in [9, 8]. As such, the argument above is not new. Rather, our contribution is to observe that it proves *more* than just the extractor's security. Indeed, summarizing the discussion so far we have shown that, if $H_\infty^\varepsilon(B|\mathbf{E}) \leq n$, then there is a measurement on $\mathbf{E}$ which, given a small amount of information about B as advice, reconstructs a good approximation to the *whole* string B with success probability $\mathrm{poly}(\varepsilon/r)$. This fact is what we already described in Lemma 3 in the introduction, and it can be formally stated as follows.

LEMMA 9. *Let ρ_{XE} be a state such that X is a classical random variable distributed over m-bit strings, and E is an arbitrarily correlated quantum system. Let $\varepsilon = \Omega(m^{-c})$, where $c > 0$ is an arbitrary constant, and $K = H_\infty^\varepsilon(X|E)$. Then there exists a subset $V \subseteq [m]$ of size $v = |V| = O(K \log^2 m)$, and for every v-bit string z a measurement M_z on E such that, with probability at least $\Omega(\varepsilon^6/m^6)$, M_{X_V} produces a string Y that agrees with X in a fraction at least $1 - \frac{1}{\log m}$ of positions.*

The lemma is proved in Appendix A. Crucially, the bits of information required as advice are localized to a small subset of bits of B, of the order of the number of bits of information Eve initially has about that string. This property holds thanks to the specific extractor we are using, which is *local*: every bit of the output only depends on few bits of the input.

Completing the reduction to the guessing game.

In the guessing game it is Alice who needs to hand the advice bits to Eve. Indeed, if Bob, holding box $\mathcal{B}$, was to hand them over, they could leak information about his secret input y: some of the advice bits may fall in blocks of the protocol that occur *after* the special block j_0 in which Bob is planning to use his secret y as input. This leak of information defeats the purpose of the guessing game, which is to demonstrate signaling between $\mathcal{A}$ and $\mathcal{B}$.

Hence the "extended" variant of the CHSH game introduced in Protocol B: since in most blocks the inputs to both $\mathcal{A}$ and $\mathcal{B}$ are identical, by the extended CHSH constraint enforced in the protocol their outputs should be identical. The relatively few advice bits needed by Eve occupy a fixed set of positions, and with good probability all Bell blocks will fall outside of these positions, in which case Alice can obtain the advice bits required by Eve directly from $\mathcal{A}$'s outputs.

The proof of Theorem 8 is now almost complete, and one may argue as in Lemma 5 that Alice and Eve together will be able to successfully predict Bob's secret input in the guessing game, contradicting the no-signaling assumption placed on $\mathcal{A}$ and $\mathcal{B}$. A more detailed proof of the theorem is given in the next section.

5.3 Proof of Theorem 8

We proceed to formally prove Theorem 8, performing a reduction to the guessing game through the use of Lemma 9. Let n and $\alpha > 0$ be given, $\varepsilon = n^{-\alpha}$, and ℓ, m, k as specified in Protocol B (described in Figure 3.)

Modeling.

Let $x = (x_i), y = (y_i), a = (a_i), b = (b_i) \in \big(\{0,1\}^k\big)^m$ denote the boxes' respective input and output strings in Protocol B, and denote the corresponding random variables by $X = (X_i), Y = (Y_i), A = (A_i), B = (B_i) \in \big(\{0,1\}^k\big)^m$. In contrast to Section 4, here we require the behavior of the boxes $\mathcal{A}, \mathcal{B}$ to follow the laws of quantum mechanics. Let $\rho_{\mathbf{ABE}}$ denote the state of the system at the start of the protocol. Here $\mathbf{A}, \mathbf{B}$ denote registers held by $\mathcal{A}, \mathcal{B}$ respectively, while $\mathbf{E}$ denotes a register held by the environment (the potential eavesdropper Eve). We could take $\rho_{\mathbf{ABE}}$ to be pure, but this will not be necessary.

At every step $i \in [km]$ of the protocol, Alice and Bob each make a binary-outcome measurement $\{A^0_{i,x}, A^1_{i,x}\}$ and $\{B^0_{i,y}, B^1_{i,y}\}$ respectively. In general their measurement may also depend on past inputs and outputs, but without loss of generality we may assume that those are recorded in the post-measurement state $\rho^{h_i}_{\mathbf{ABE}}$ resulting from $\mathcal{A}$ and $\mathcal{B}$'s measurements in previous rounds (here, as in Section 4, $h_i = (a_{<i}, b_{<i}, x_{<i}, y_{<i})$ denotes the protocol history). For every $i \in [km]$ we may then write

$$p_{A_iB_i|X_iY_iH_i}(a_i, b_i|x_i, y_i, h_i) = \mathrm{Tr}\big((A^{a_i}_{i,x_i} \otimes B^{b_i}_{i,y_i} \otimes \mathbb{I}_{\mathbf{E}})\, \rho^{h_i}_{\mathbf{ABE}}\big). \quad (8)$$

Assuming the conditions of Lemma 9 are met, let V be the fixed subset of $\{0,1\}^{mk}$, of size v, whose existence is guaranteed in its conclusion. For every v-bit string z let $\{M^e_v\}_{e\in\{0,1\}^{mk}}$ be the measurement on register $\mathbf{E}$ whose existence is also guaranteed in the lemma. V and M depend on the state $\rho_{\mathbf{XYE}}$ and the measurements performed by $\mathcal{A}$ and $\mathcal{B}$, but not on the specific execution of the protocol performed by the user: as long as $\mathcal{A}$ and $\mathcal{B}$ are fixed they are, too. We introduce two new random variables to model the outcomes obtained from performing the measurement $\{M^e_z\}$ on register $\mathbf{E}$, for different choices of z. We use $E^A = (E^A_i) \in \big(\{0,1\}^k\big)^m$ to denote the outcome when the string z is the string a_V taken from $\mathcal{A}$'s outputs, and $E^B = (E^B_i) \in \big(\{0,1\}^k\big)^m$ to denote its outcome when it is the string b_V taken from $\mathcal{B}$'s output.

Let G^A be the event that $d_H(E^A, B) < f_e$, and G^B the event that $d_H(E^B, B) < f_e$, where $f_e > 0$ is a parameter to be specified later. Let $j \in T$ be an index that runs over the blocks that have been designated as Bell blocks in Step 2. of the protocol (T itself is a random variable). Given a Bell block j, let G^A_j be a boolean random variable such that $G^A_j = 1$ if and only if either $d_H(E^A_j, B) < 0.01$ and $Y_j = (A, 0)$, or $d_H(E^A_j, B) < 0.17$ and $Y_j = (B, 0)$. Define G^B_j symmetrically with respect to E^B instead of E^A.

Finally, for $i \in [m]$, let CHSH_i be the following event:

$$\mathrm{CHSH}_i = \begin{cases} A_i = B_i & \text{if } X_i = Y_i, \\ d_H(A_i, B_i) \le 0.16 & \text{if } Y_i = (B, 0), \\ d_H(A_i, B_i) \in [0.49, 0.51] & \text{if } X_i = (A, 1) \\ & \text{and } Y_i = (A, 0). \end{cases}$$

Honest CHSH boxes as described above satisfy CHSH_i with probability $1 - 2^{-\Omega(k)}$. Let $\mathrm{CHSH} = \bigwedge_i \mathrm{CHSH}_i$.

We prove Theorem 8 by contradiction. Assume that both the theorem's conclusions are violated, so that (i) $H^\varepsilon_\infty(B'|\mathbf{E}) \ge n$, where B' is a random variable describing the distribution of $\mathcal{B}$'s outputs conditioned on CHSH, and $\Pr\big(\mathrm{CHSH}\big) \le \varepsilon$. Here $\varepsilon = n^{-\alpha}$, where $\alpha > 0$ is a parameter.

The first step is to apply Lemma 9 with $X = B'$. The conclusion of the lemma is that there exists a subset $V \subseteq [km]$ of size $|V| = O(n \log^2 n)$ such that, letting $f_e = 1/(\log mk)$, we have $p_s := \Pr(G^B|\mathrm{CHSH}) = \Omega(\varepsilon^7/n^6) = \Omega(n^{-6-7\alpha})$, where G^B denotes the event that Eve correctly predicts B on a fraction at least $1 - f_e$ of positions. Since in Protocol B the Bell blocks form only a very small fraction of the total, a priori it could still be that Eve's prediction is systematically wrong on all Bell blocks, preventing us from successfully using them in the guessing game.

The following claim shows Eve's errors cannot be concentrated in the Bell blocks. The intuition is the following. If $\mathcal{B}$'s input in a Bell block is $(A, 0)$ then nothing distinguishes this block from most others, so that Eve's prediction has no reason of being less correct than average. However, blocks in which its input is $(B, 0)$ are distinguished. We rule out the possibility that Eve's errors are concentrated in such blocks by appealing to the no-signaling condition between Eve and $\mathcal{A}$. Indeed, about half of Bell blocks in which $\mathcal{B}$'s input is $(B, 0)$ are such that $\mathcal{A}$'s input for the same block is $(A, 0)$: looking only at $\mathcal{A}$'s inputs they are indistinguishable from most other blocks. We will argue that, as long as the CHSH constraint is satisfied, Eve might as well have been given the advice bits by Alice, in which case there is no reason for her to make more errors than average in those blocks.

Claim 10. *Let T be the set of Bell blocks selected in Protocol B. Then there exists a constant $c_e < 10^{-3}$ such that the following holds.*

$$\Pr\Big(\mathrm{E}_{j\in T}\big[\, G^A_j \,\big] > 1 - \frac{c_e}{\log n}, \mathrm{CHSH}\Big) = \Omega(p_s\varepsilon) = \Omega\big(n^{-6-8\alpha}\big).$$

Proof. By definition, $\Pr\big(G^B\big) \ge p_s \Pr(\mathrm{CHSH}) \ge p_s\varepsilon$. Conditioned on G^B, by Markov's inequality it must be that $d_H(E^B, B) < 0.01$ on a fraction at least $1 - 100 f_e$ of blocks in which the input to B was $(A, 0)$. Let $f'_e = 100 f_e$. Let $\eta = 2^{-10^{-5} f'_e |T|/(2\cdot 100^2)}$, and assume C chosen large enough so that $\eta \le p_s\varepsilon/6 = \Omega(n^{-6-8\alpha})$. This is possible since $|T|$ is sharply concentrated around $C \log^2 \ell$ and $f'_e = \Omega(1/\log \ell)$.

Among the blocks in which Eve's prediction is correct, nothing distinguishes those Bell blocks in which $\mathcal{B}$'s input is $(A, 0)$: indeed, we may think of those only being designated as Bell blocks *after* Eve has made her prediction. By a Chernoff bound the probability that more than a fraction $2f'_e$ of such blocks fall into those for which G^B_j does not hold is upper-bounded by η. Hence the following holds

$$\Pr\big(\mathrm{E}_{j\in T:\, Y_j=(A,0)}\, G^B_j > 1 - 2f'_e | G^B\big) \ge 1 - \eta. \quad (9)$$

Since V is a fixed subset of $[km]$ of size $|V| = O(n \log^2 m)$, the probability that any of the randomly chosen $O(\log^2 \ell)$ Bell blocks intersects it is at most $O(n^{2-(10+8\alpha)} \log^4 n)$ for large enough n. This quantity is much smaller than (our upper bound on) η, and for the remainder of the proof we will neglect the chance of this happening.

Conditioning further on CHSH can only blow-up the error by a factor $1/\Pr(\mathrm{CHSH}|G^B) \leq 1/(p_s\varepsilon)$. In that case $G^A = G^B$ (Eve's prediction only depends on the advice bits she is given, and these bits are the same when taken from either $\mathcal{A}$ and $\mathcal{B}$'s outputs whenever the CHSH condition holds), so we obtain:

$$\begin{aligned}&\frac{\Pr\left(\mathrm{E}_{j\in T:Y_j=(A,0)}\, G_j^A > 1-2f'_e, \mathrm{CHSH}|G^A\right)}{\Pr\left(\mathrm{CHSH}|G^A\right)}\\ &= \Pr\left(\mathrm{E}_{j\in T:\, Y_j=(A,0)}\, G_j^A > 1-2f'_e|G^A, \mathrm{CHSH}\right)\\ &\geq 1-\eta/(p_s\varepsilon). \end{aligned} \tag{10}$$

Suppose Eve makes more than a fraction $5f'_e$ of errors in predicting $\mathcal{A}$'s output on those Bell blocks in which its input is $(A,0)$. By a Chernoff bound with probability at least $1-\eta$ the input to $\mathcal{B}$ will also be $(A,0)$ in at least 40% of those blocks. Indeed, since Eve now receives her advice bits from $\mathcal{A}$'s outputs, by the no-signaling condition we may think of the choice of $\mathcal{B}$'s inputs as being made *after* both Alice and Eve have completed their measurements. Whenever this condition holds, Eve's prediction will be wrong on a total fraction more than $2f'_e$ of $\mathcal{B}$'s $(A,0)$-input Bell blocks, contradicting (10). Indeed, whenever CHSH holds, if the input to both boxes is $(A,0)$ then Eve being correct in predicting $\mathcal{B}$'s output is equivalent to her being correct in predicting $\mathcal{A}$'s output. Hence the following holds:

$$\begin{aligned}&\Pr\left(\mathrm{E}_{j\in T:\, X_j=(A,0)}\, G_j^A > 1-5f'_e, \mathrm{CHSH}|G^A\right)\\ &\quad\geq \Pr\left(\mathrm{E}_{j\in T:\, Y_j=(A,0)}\, G_j^A > 1-2f'_e, \mathrm{CHSH}|G^A\right) - \eta\\ &\quad\geq (1-\eta/(p_s\varepsilon))\Pr\left(\mathrm{CHSH}|G^A\right) - \eta\\ &\quad\geq (1-2\eta/(p_s\varepsilon))\Pr\left(\mathrm{CHSH}|G^A\right), \end{aligned} \tag{11}$$

where the second inequality follows from(10) and the last uses $\Pr(\mathrm{CHSH}|G^A) \geq p_s\varepsilon$. As previously, since $G^A \wedge \mathrm{CHSH} = G^B \wedge \mathrm{CHSH}$, (11) implies the following:

$$\begin{aligned}&\Pr\left(\mathrm{E}_{j\in T:\, X_j=(A,0)}\, G_j^B > 1-5f'_e|G^B, \mathrm{CHSH}\right)\\ &\qquad\geq 1-2\eta/(p_s\varepsilon). \end{aligned} \tag{12}$$

Next, suppose Eve makes a prediction that is wrong on a fraction at least $14f'_e$ of the Bell blocks, irrespective of Bob's inputs. Then again with high probability at least 40% of the inputs to $\mathcal{A}$ in those blocks will be $(A,0)$, implying that Eve is wrong on more than a fraction $5f'_e$ of $\mathcal{A}$'s $(A,0)$ inputs, contradicting (12). Hence the following is proven just as (11) was:

$$\Pr\left(\mathrm{E}_{j\in T}\, G_j^B > 1-14f'_e|G^B, \mathrm{CHSH}\right) \geq 1-3\eta/(p_s\varepsilon). \tag{13}$$

Hence

$$\Pr\left(\mathrm{E}_{j\in T}\, G_j^A > 1-14f'_e|G^A, \mathrm{CHSH}\right) \geq 1-3\eta/(p_s\varepsilon),$$

which is greater than $1/2$ given our choice of η. Removing all conditioning, whenever Eve is given advice bits by Alice, it holds that

$$\Pr\left(\mathrm{E}_{j\in T}\, G_j^A > 1-14f'_e, \mathrm{CHSH}\right) \geq \Omega(p_s\varepsilon).$$

□

Based on Claim 10 we can show an analogue of Claim 7 which will let us complete the reduction to the guessing game. Claim 10 shows that with probability $\Omega(p_s\varepsilon)$ Eve's prediction will be correct on a fraction at least $1-c_e/\log n$ of Bell blocks. Since there are $O(\log^2 n)$ such blocks in Protocol B, with the same probability Eve only makes errors on a total number $w_e = O(\log n)$ of Bell blocks. Group the Bell blocks in groups of $20w_e$ successive blocks, and let k be an index that runs over such groups; there are $O(\log n)$ of them. Let G_k^A be the event that Eve's prediction is correct in at least 99% of the Bell blocks in group k: $G_k^A = 1$ if and only if $\mathrm{E}_{j\sim k} G_j^A \geq 0.99$, where the average is taken over the Bell blocks comprising group k. By Markov's inequality, it follows from Claim 10 that $\Pr(\wedge_k G_k^A, \mathrm{CHSH}) = \Omega(p_s\varepsilon)$.

Claim 11. *For all large enough n there exists a Bell block $j_0 \in T$ such that, in that block, it is highly likely that both Eve's prediction (when given advice bits from $\mathcal{A}$'s output) is correct and the CHSH constraint is satisfied, conditioned on this being so in past iterations:*

$$\Pr(G_{j_0}^A, CHSH_{j_0}|CHSH_{j<j_0}, G_{k<k_0}^A) \geq 0.98, \tag{14}$$

where k_0 is the index of the group containing the j_0-th Bell block.

Proof. By the chain rule, since there are $O(\log n)$ groups there will exist a group k_0 in which Eve's prediction is correct, and the CHSH condition is satisfied, with probability at least 0.99, when conditioned on the same holding of all previous groups. Since by definition Eve being correct in the group means that she is correct in 99% of that group's blocks, there is a specific block j_0 in which she is correct with probability at least 0.98. □

The reduction to the guessing game should now be clear, and follows along the same lines as the proof of Theorem 6 given in Section 4. Alice and Bob run protocol B, including the selection of all Bell blocks T, with the boxes $\mathcal{A}$ and $\mathcal{B}$, up to the j_0-th Bell block (excluded). Bob communicates $\mathcal{B}$'s outputs up till that block to Alice. They check that the CHSH constraint is satisfied in all blocks previous to the j_0-th; if not they abort. The guessing game can now start: Alice and Bob are separated and Bob is given his secret input y. If $y = 0$ then he chooses $(A,0)$ as input to $\mathcal{B}$ in the j_0-th block; otherwise he chooses $(B,0)$. He then completes the protocol honestly. Alice chooses an input $x \in \{(A,0),(A,1)\}$ at random for the j_0-th block, and then completes the protocol honestly.

In order to help her guess Bob's input, Alice has access to the eavesdropper Eve. Alice gives the bits a_V taken from $\mathcal{A}$'s output string a as advice bits to Eve. Eve makes a prediction e for Bob's output. Alice checks that the event $G_{<k_0}^A$ is satisfied. If not she aborts. If so, by Claim 11 we know that both CHSH_{j_0} and $G_{j_0}^A$ are satisfied with probability at least 0.98, so this must be so with probability at least 0.92 for each of the four possible pair of inputs (x,y) given to $\mathcal{A}$ and $\mathcal{B}$ in the j_0-th block.

Alice makes her prediction as follows: if either $\mathcal{A}$'s input was $(A,0)$ and its output agrees with Eve's prediction on at least a 0.99 fraction of positions (in the j_0-th block), or $\mathcal{A}$'s input was $(A,1)$ and its output agrees with Eve's prediction on a fraction of positions that is between 0.48

and 0.52 she claims "Bod had a 0". Otherwise she claims "Bob had a 1".

Clearly if Bob is using $(A, 0)$ as his input then Alice will predict correctly with probability at least 0.92, since in that case $G^A_{j_0}$ implies that Eve predicts $\mathcal{B}$'s output with at most 1% of error. If he is using $(B, 1)$ then $G^A_{j_0}$ implies that Eve's prediction will be within 0.17 relative Hamming distance of $\mathcal{B}$'s output in block j_0. By the CHSH constraint $\mathcal{A}$'s output must also be within 0.16 of $\mathcal{B}$'s output, whatever input Alice chooses. Hence $\mathcal{A}$'s output is always within $0.43 < 0.49$ of $\mathcal{B}$'s, meaning Alice will correctly claim Bob had a 1 whenever her input is $(A, 1)$. Hence in that case she correctly predicts Bob's input with probability at least 0.92/2.

Overall, conditioned on Alice not aborting her prediction is correct with probability at least 0.69 over the choice of a random input for Bob, indicating a violation of the no-signaling assumption on the boxes and proving Theorem 8.

Acknowledgments.

We thank Matthew Coudron for useful comments on a preliminary draft of this paper.

6. REFERENCES

[1] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM*, 49(4):496–511, 2002.

[2] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[3] J. S. Bell. On the problem of hidden variables in quantum theory. *Rev. Mod. Phys.*, 38:447–452, 1966.

[4] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, 1969.

[5] R. Colbeck. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, Trinity College, University of Cambridge, Nov. 2009.

[6] R. Colbeck and A. Kent. Private randomness expansion with untrusted devices. *Journal of Physics A: Mathematical and Theoretical*, 44(9):095305, 2011.

[7] R. Colbeck and R. Renner. Free randomness amplification. arXiv:1105.3195, 2011.

[8] A. De, C. Portmann, R. Renner, and T. Vidick. Trevisan's extractor in the presence of quantum side information. Technical report arXiv:0912.5514, 2009.

[9] A. De and T. Vidick. Near-optimal extractors against quantum storage. In *Proceedings of the 42nd ACM STOC*, pages 161–170, New York, NY, USA, 2010.

[10] A. Einstein, P. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

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

[12] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM Journal of Computing*, 38(5):1695–1708, 2008.

[13] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from parvaresh-vardy codes. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity*, pages 96–108, Washington, DC, USA, 2007.

[14] T. Hartman and R. Raz. On the distribution of the number of roots of polynomials and explicit weak designs. *Random Structures and Algorithms*, 23(3):235–263, 2003.

[15] R. Impagliazzo, R. Jaiswal, and V. Kabanets. Approximately list-decoding direct product codes and uniform hardness amplification. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 187 –196, oct. 2006.

[16] A. Knuth, D. Yao. *Algorithms and Complexity: New Directions and Recent Results,* Chapter *The complexity of nonuniform random number generation*. Academic Press, 1976.

[17] R. König, R. Renner, and C. Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, 2009.

[18] R. König and B. Terhal. The bounded storage model in presence of a quantum adversary. *IEEE Transactions on Information Theory*, 54(2):749–762, 2008.

[19] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[20] S. Pironio, A. Acin, S. Massar, A. B. 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):10, 2009.

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

[22] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, Swiss Federal Institute of Technology Zurich, Sept. 2005.

[23] M. Santha and U. V. Vazirani. Generating quasi-random sequences from slightly-random sources. In *Proceedings of the 25th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 434–440, Washington, DC, USA, 1984.

[24] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the European Association for Theoretical Computer Science*, 77:67–95, June 2002.

[25] A. Ta-Shma. Short seed extractors against quantum storage. In *Proceedings of the 41st annual ACM*

Symposium on Theory of Computing (STOC), pages 401–408, New York, NY, USA, 2009.

[26] G. Taylor and G. Cox. Behind intel's new random-number generator. *IEEE Spectrum*, September 2011.

[27] L. Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48:860–879, July 2001.

[28] D. Zuckerman. General weak random sources. In *Proceedings of the 31st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 534–543, 1990.

APPENDIX

A. PROOF OF LEMMA 9

In this appendix we give the proof of Lemma 9. The proof crucially uses properties of a specific extractor construction based on Trevisan's construction paradigm [27]. A specific construction based on this paradigm was first shown to be secure in the presence of quantum bounded-storage adversaries in [25]. The construction we use here was shown secure in the more general setting of quantum bounded-information adversaries in [8]. We first describe the extractor.

A.1 The t-XOR extractor

The t-XOR extractor E_t, parametrized by an integer t, follows Trevisan's general construction paradigm [27]. It is based on two main ingredients, the t-XOR code and a combinatorial design construction due to Hartman and Raz [14]. For us, only the details of the t-XOR code will be important.

The t-XOR code.

Given integers m and $t \leq m$, let $C_t : \{0,1\}^m \to \{0,1\}^{\binom{m}{t}}$ map an m-bit string to the string of parities of all subsets of t out of its m bits. Two properties of this encoding will be relevant for us. The first is that it is locally computable: each bit of the codeword only depends on t bits of the input. The second is that it is approximately list-decodable (see Lemma 16 below).

Combinatorial designs.

Given integers s, m, r and $\rho > 0$, a collection of subsets $S_1, \ldots, S_r \subseteq [s]$ is called a (s, m, r, ρ) weak design if for all $i \in [r]$, $|S_i| = m$ and for all j, $\sum_{i<j} 2^{|S_i \cap S_j|} \leq \rho(r-1)$. For our purposes it will suffice to note that Hartman and Raz [14] proved the existence of a $(s, m, r, 1+\gamma)$ design for every m, $0 < \gamma < 1/2$, $s = O(m^2 \log 1/\gamma)$ and $r > s^{\Omega(\log s)}$.

The t-XOR extractor.

We define the extractor that we will use in the proof of Lemma 9.

DEFINITION 12. *Let m, r, t, s be given integers such that $t = O(\log m)$ and $s = O(\log^4 n)$. Then $E_t : \{0,1\}^m \times \{0,1\}^s \to \{0,1\}^r$ maps $(x, y) \in \{0,1\}^m \times \{0,1\}^s$ to $C_t(x)_{y_{S_1}}, \ldots, C_t(x)_{y_{S_r}}$, where $(S_1, \ldots, S_r)$ is a $(s, t \log m, r, 5/4)$ design and y_{S_i} designates the bits of y indexed by S_i, interpreted as a t-element subset of $[m]$.*

While, as shown in Corollary 5.11 in [8], E_t is a strong extractor with good parameters, we will not use this fact directly. Rather, we will use specific properties that arise from the "reconstruction paradigm"-based *proof* that it is an extractor secure against quantum adversaries. Indeed, one may argue that Lemma 9 is implicit in the proof of security of E_t given in [8]. Since it does not follow directly from the mere statement that E_t is an extractor, we give more details here. We will show the following lemma, which is more general than Lemma 9.

LEMMA 13. *Let m, r, t be integers, $\varepsilon > 0$, and suppose that $t = O(\log^2 m)$. Let ρ_{XE} be a cq-state such that X is a random variable distributed over m-bit strings. Let U_r be uniformly distributed over r-bit strings, and suppose that*

$$\left\|\rho_{Ext(X,Y)E} - \rho_{U_r} \otimes \rho_E\right\|_{tr} > \varepsilon, \tag{15}$$

i.e. an adversary Eve holding register E can distinguish the output of the extractor from a uniformly random r-bit string. Then there exists a fixed subset $V \subseteq [m]$ of size $|V| = O(tr)$ such that, given the string X_V as advice, with probability at least $\Omega(\varepsilon^2/r^2)$ over the choice of $x \sim p_X$ and her own randomness Eve can output a list of $\ell = O(r^4/\varepsilon^4)$ strings $\tilde{x}^1, \ldots, \tilde{x}^\ell$ such that there is an $i \in [\ell]$, $d_H(\tilde{x}^i, x) \leq (2/t)\ln(4r/\varepsilon)$.

It is not hard to see why Lemma 13 implies Lemma 9. First note that if r is chosen in Lemma 13 so that $r > 2H_\infty^\varepsilon(X|E)$ then the assumption (15) is automatically satisfied as a consequence of the data processing inequality.[9] The conclusion of Lemma 9 then follows from that of Lemma 13 by having Eve output a random string out of her ℓ predictions, and choosing $t = \Omega(\log^2 m)$ to ensure that $(2/t)\ln(4r/\varepsilon) \leq 1/\log m$.

In the remainder of this section we sketch the proof of Lemma 13. The first step, explained in Section A.2, consists in using a hybrid argument to show that, given (15), Eve can predict a random t-XOR of X's bits with reasonable success probability, given sufficiently many "advice bits" about X. In the second step, detailed in Section A.3, we show using an argument due to Koenig and Terhal [18] that this implies the adversary can in fact recover most t-XORs of X, simultaneously. Finally, in Section A.4 we use the list-decoding properties of the XOR code to show that as a consequence the adversary can with good probability produce a string that agrees with X on a large fraction of coordinates.

A.2 The hybrid argument

Suppose that (15) holds. Proposition 4.4 from [8] shows that a standard hybrid argument, together with properties of Trevisan's extractor (specifically the use of the seed through combinatorial designs), can be used to show the following claim.

CLAIM 14. *There exists a subset $V \subseteq [m]$ of size $|V| = O(tr)$ such that, given the bits X_V, Eve can predict a random t-XOR of the bits of X with advantage ε/r. For-*

[9] The extra randomness coming from the seed of the extractor will be small, as its size can be taken to be $s = O(\log^4 m)$.

mally,

$$\left\|\rho_{C_t(X)_Y YVE} - \rho_{U_1}\otimes\rho_Y\otimes\rho_{VE}\right\|_{tr} > \frac{\varepsilon}{r}, \qquad (16)$$

where Y is a random variable uniformly distributed over $\left[\binom{m}{t}\right]$ and V is a register containing the bits of X indexed by V.

A.3 Recovering all t-XORs.

The next step in the proof of Lemma 13 is to argue that Eq. (16) implies that an adversary given access to $E' = VE$ can predict not only a random XOR of X, but a string Z of length $\binom{m}{t}$ such that Z agrees with the string $C_t(X)$ of all t-XOR's of X in a significant fraction of positions. Classically this is trivial, as one can just repeat the single-bit prediction procedure guaranteed by (16) for all possible choices Y of the t bits whose parity one is trying to compute. In the quantum setting it is more subtle. We will follow an argument from [18] showing that (16) implies that there is a single measurement, independent of Y, that one can perform on E and using the (classical) result of which one can predict the bits $C_t(X)_Y$ with good success on average (over the measurement's outcome and the choice of Y).

Claim 15. *Suppose* (16) *holds. Then there exists a measurement $\mathcal{F}$, with outcomes in $\{0,1\}^m$, such that*

$$\Pr_{x\sim p_X,\, y\sim U_{t\log m}}\left(C_t(x)_Y = C_t(\mathcal{F}(VE))_y\right) \geq \frac{1}{2} + \frac{\varepsilon^2}{4r^2}, \qquad (17)$$

where $\mathcal{F}(VE)$ denotes the outcome of $\mathcal{F}$ when performed on the cq-state ρ_{VE}.

Proof. Our argument closely follows the proof of Theorem III.1 from [18]. Given an arbitrary cq-state ρ_{ZQ}, define the non-uniformity of Z given Q as

$$d(Z\leftarrow Q) := \left\|\rho_{ZQ} - \rho_{U_z}\otimes\rho_Q\right\|_{tr}.$$

Let ρ_x denote the state contained in registers VE, conditioned on $X = x$. For a fixed string y, define two states

$$\rho_0^y := \sum_{x:\, C_t(x)_y=0} p_X(x)\,\rho_x,$$

$$\rho_1^y := \sum_{x:\, C_t(x)_y=1} p_X(x)\,\rho_x.$$

Then, by definition $d\big(C_t(X)_y \leftarrow VE\big) = \left\|\rho_0^y - \rho_1^y\right\|_{tr}$ is the adversary's maximum success probability in distinguishing those states ρ_x which correspond to an XOR of 0 from those which correspond to an XOR of 1. Let $\mathcal{E}_y = \left\{E_y^0, E_y^1\right\}$ be the pretty good measurement corresponding to the pair of states $\left\{\rho_0^y, \rho_1^y\right\}$:

$$E_y^0 = \rho_{VE}^{-1/2}\rho_0^y\,\rho_{VE}^{-1/2} \qquad \text{and} \qquad E_y^1 = \rho_{VE}^{-1/2}\rho_1^y\,\rho_{VE}^{-1/2},$$

where $\rho_{VE} = \sum_x P_X(x)\rho_x$. Lemma 2 from [18] (more precisely, Eq. (19)), shows that the following holds as a consequence of (16):

$$\sqrt{\mathrm{E}_y\left[\,2\,d\big(C_t(X)_y \leftarrow \mathcal{E}^y(VE)\big)\right]} + d(C_t(X)_Y \leftarrow Y) > \frac{\varepsilon}{r}, \qquad (18)$$

where $\mathcal{E}^y(VE)$ is the result of the POVM $\mathcal{E}^y$ applied on ρ_{VE}, and $d(C_t(X)_Y \leftarrow Y)$ is the distance from uniform of the one-bit extractor's output, in the absence of the adversary. We may as well assume this term to be small: indeed, if it is more than $\varepsilon/(2r)$ then (17) is proved without even having to resort to the quantum system E. Hence (18) implies

$$\mathrm{E}_y\left[\,d\big(C_t(X)_y \leftarrow \mathcal{E}_{pgm}^y(VE)\big)\,\right] > \frac{\varepsilon^2}{2r^2},$$

which can be equivalently re-written as

$$\mathrm{E}_y\left[\,\mathrm{Tr}\big(E_y^0\,\rho_y^0\big) + \mathrm{Tr}\big(E_y^1\,\rho_y^1\big)\,\right] > \frac{1}{2} + \frac{\varepsilon^2}{4r^2}. \qquad (19)$$

Following the argument in [18], we define a new measurement $\mathcal{F}$ with outcomes in $\{0,1\}^m$ and POVM elements $F^x = P_X(x)\rho_{VE}^{-1/2}\rho_x\,\rho_{VE}^{-1/2}$. The important point to notice is that for $z\in\{0,1\}$ we have $E_y^z = \sum_{x:\, C_t(x)_y=z} F^x$, hence (19) can be re-written as

$$\mathrm{E}_y\left[\sum_{b:\, C_t(b)_y=0}\mathrm{Tr}\big(F^x\,\rho_y^0\big) + \sum_{b:\, C_t(b)_y=1}\mathrm{Tr}\big(F^x\,\rho_y^1\big)\right] > \frac{1}{2} + \frac{\varepsilon^2}{4r^2},$$

which is exactly (17). □

A.4 List-decoding the XOR code.

The following lemma (for a reference, see [15], Lemma 42) states the list-decoding properties of the t-XOR code C_t that are important for us.

Lemma 16. *For every $\eta > 2t^2/2^m$ and $z\in(\{0,1\}^m)^t$, there is a list of $\ell \leq 4/\eta^2$ elements $x^1,\ldots,x^\ell\in\{0,1\}^m$ such that the following holds: for every $z'\in\{0,1\}^m$ which satisfies*

$$\Pr_{\{y_1,\ldots,y_t\}\in\binom{m}{t}}\left[z_{(y_1,\ldots,y_t)} = \oplus_{i=1}^t z'_{y_i}\right] \geq \frac{1}{2} + \eta,$$

there is an $i\in[\ell]$ such that

$$\Pr_{y\sim\mathcal{U}_N}\left[x_y^i = z'_y\right] \geq 1-\delta,$$

with $\delta = (1/t)\ln(2/\eta)$.

Claim 15 implies that, with probability at least $\varepsilon^2/(8r^2)$ over the choice of x and over Eve's own randomness, when measuring her system with $\mathcal{F}$ she will obtain a string $\tilde{z}$ whose t-XORs agree with those of x with probability at least $1/2+\varepsilon^2/(8r^2)$. Lemma 16 shows that in that case she can recover a list of at most $2^8r^4/\varepsilon^4$ "candidate" strings $\tilde{z}^i$ such that there exists at least one of these strings which agrees with x at a (possibly adversarial) fraction $1-\delta$ of positions, where $\delta = (2/t)\ln(4r/\varepsilon)$ given our choice of parameters. Hence Lemma 13 is proved.