Fully device independent quantum key distribution

Umesh Vazirani* Thomas Vidick†

Abstract

We give the first device-independent proof of security of a protocol for quantum key distribution that guarantees the extraction of a linear amount of key even when the devices are subject to a constant rate of noise. Our only assumptions are that the laboratories in which each party holds his or her own device are spatially isolated, and that both devices as well as the eavesdropper, are bound by the laws of quantum mechanics.

1 Introduction

Quantum key distribution [BB84, Eke91] together with its proof of security [May01, SP00] appeared to have achieved the holy grail of cryptography — unconditional security, or a scheme whose security was based solely on the laws of physics. However, practical implementations of QKD protocols necessarily involve imperfect devices [BBB+92, MHH+97], and it was soon realized that these imperfections could be exploited by a malicious eavesdropper to break the "unconditional" security of QKD (see e.g. [SK09] for a review).

Mayers and Yao [MY98] put forth a vision for restoring unconditional security in the presence of imperfect or even maliciously designed devices, by subjecting them to tests that they fail unless they behave consistently with "honest" devices. The fundamental challenge they put forth was of *device-independent quantum key distribution* (DIQKD): establishing the security of a QKD protocol based only on the validity of quantum mechanics, the physical isolation of the devices and the passing of certain statistical tests. The germ of the idea for device-independence may already be seen in Ekert's entanglement-based protocol for QKD [Eke91]. Barrett, Hardy, and Kent [BHK05] used intuition from the recently discovered monogamy of non-signalling correlations [BLM+05] to show that the near-maximal violation of a specific Bell inequality, based on the chained inequalities of Braunstein and Caves [BC90], by an arbitrary pair of spatially isolated devices could be used to guarantee the generation of a random bit secure against any non-signalling eavesdropper.[1]

A long line of research on DIQKD seeks to make the qualitative argument from [BHK05] quantitative, devising protocols that extract an amount of key that is linear in the number of uses of the devices, and is secure against increasingly general eavesropping strategies. Initial works [AGM06, AMP06, SGB+06] give

*Department of Computer Science, UC Berkeley, California. Supported by ARO Grant W911NF-09-1-0440, NSF Grant CCF-0905626 and Templeton Foundation Grant 21674. Email vazirani@eecs.berkeley.edu

†Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Supported by NSF Grant 0844626. Email vidick@csail.mit.edu.

[1] Barrett et al. give a protocol in which the users have access to n pairs of devices, such that the no-signalling condition holds in-between each pair. This assumption is used in order to estimate the amount of Bell inequality violation that characterizes any single use of the devices.

efficient and noise-tolerant protocols that are secure against individual attacks by non-signalling eavesdroppers. Subsequent work [MRC+09, Mas09] and [HRW10] also proved security against collective attacks. Other works [ABG+07, PAB+09, MRC+09, HR10, MPA11] obtains better key rates under the stronger assumption that the eavesdropper is bound by the laws of quantum mechanics. All these results, however, could only be established under restrictive *independence* assumptions on the devices, e.g. in recent work [HR10, MPA11] a proof of security based on collected statistics requires that the n uses of each device are causally independent: measurements performed at successive steps of the protocol commute with each other.

Very recently two papers [BCK12b, RUV12] announced proofs of security of DIQKD without requiring any independence assumption between the different uses of the devices Unfortunately, although the approaches in [BCK12b, RUV12] are very different both implied protocols are inefficient and cannot tolerate noisy devices. The protocol used in [BCK12b] is very similar to the one originally introduced in [BHK05], and requires a large number of uses of a pair of noise-free devices in order to generate a single bit of key. In the case of [RUV12] the noise intolerance comes as a consequence of the very strong testing that is performed: building on work on making the self-checking idea of Mayers and Yao robust [MMMO06, MYS12], the authors show that the shared quantum state and operators of the two untrusted devices can be completely characterized by performing certain statistical tests (CHSH tests). It is unclear whether such strong testing can be achieved in a manner that is robust to noise.

A major stumbling block of DIQKD is the difficulty of dealing with the noise inherent in even the best devices, without making any independence assumptions. Indeed, a good DIQKD protocol should differentiate devices that are honest but noisy from devices that may attempt to take advantage of a protocol's necessary tolerance to noise in order to leak information to an eavesdropper by introducing correlations in their errors [BCK12a]. The protocols in [BCK12b, RUV12] do not achieve this, since they cannot tolerate any constant noise rate.

On a quite different though related front, a recent line of work investigates the possibility of generating certifiable randomness. Although the goal is different, abstractly these works may be viewed as device-independent results in which the devices are used in multiple rounds, without imposing independence assumptions on successive rounds. Building on an observation made in [Col06], Pironio et al. [PAM+10] devised a protocol in which the generation of randomness could be certified solely by testing for a sufficiently large Bell inequality violation. In [FGS11, PM11] it was further shown that the randomness generated was secure against an arbitrary classical adversary. Concurrently, [VV12] gave a protocol that was secure even against a quantum adversary. It is tempting to use this protocol as a basis for DIQKD by drawing an analogy between the quantum adversary in the randomness protocol and the eavesdropper in QKD. This is not straightforward though, to begin with because in randomness generation the users' inputs may be kept private whereas in DIQKD they are revealed to the adversary in the classical post-processing phase; this opens the way for a much broader set of eavesdropping strategies. Moreover, the strong security guarantees in [VV12] seem to crucially depend upon the zero error-tolerance of the protocol, which provides a way to enforce extremely strong correlations between the devices' outputs.

To summarize, existing lines of work demonstrate the device-independent security of protocols assuming either that each successive use of the devices is independent, or that they are noiseless. This raises the question: is device-independent QKD even *possible* without independence assumptions in a realistic, noise-tolerant scenario?

1.1 Results

We answer this question in the affirmative by giving the first complete device-independent proof of security of quantum key distribution that tolerates a constant noise rate and guarantees the generation of a linear amount of key. Our only assumption on the devices is that they can be modeled by the laws of quantum mechanics, and that they are spatially isolated from each other and from any adversary's laboratory. In particular, we emphasize that the devices may have quantum memory. While our proof builds upon ideas from the work on certifiable randomness generation mentioned above, our protocol is closely related to Ekert's entanglement-based protocol [Eke91].

In the protocol,[2] the users Alice and Bob make n successive uses of their respective devices. At each step, Alice (resp. Bob) privately chooses a random input $x_i \in \{0,1,2\}$ (resp. $y_i \in \{0,1\}$) for her device, collecting an output bit a_i (resp. b_i). If the devices were honestly implemented they would share Bell states $|\psi\rangle = 1/\sqrt{2}|00\rangle + 1/\sqrt{2}|11\rangle$, and measure their qubits according to the following strategy: if $x_i = 0$ measure in the computational basis, if $x_i = 1$ measure in the Hadamard basis and if $x_i = 2$ measure in the $3\pi/8$-rotated basis. If $y_i = 0$ measure in the $\pi/8$-rotated basis and if $y_i = 1$ measure in the $3\pi/8$-rotated basis.

To test the devices, after the n steps have been completed, the users select a random subset $\mathbf{B} \subseteq \{1,\ldots,n\}$ of size $|\mathbf{B}| = \gamma n$, where $\gamma > 0$ is a small constant, and publicly announce their inputs and outputs in $\mathbf{B}$. Let $z_i = 1$ if and only if $a_i \neq 2$ and $a_i \oplus b_i \neq x_i \wedge y_i$, or $(a_i, b_i) = (2,1)$ and $a_i \neq b_i$. The users jointly compute the noise rate $\eta := (1/|\mathbf{B}|)\sum_{i\in\mathbf{B}} z_i - (1 - \mathtt{opt})$, where $\mathtt{opt} = (2\cos^2 \pi/8 + 1)/3$.[3]. If $\eta \geq 0.5\%$ they abort. If not, they announce their remaining input choices. Let $\mathbf{C} \subseteq \{1,\ldots,n\}$ be the steps in which $(a_i, b_i) = (2,1)$. The users conclude by performing standard information reconciliation and privacy amplification steps on their outputs in $\mathbf{C}$, extracting a key of length κn for some $\kappa = \kappa(\eta, \varepsilon)$, where ε is the desired security parameter.

Theorem 1 (Informal)**.** *Let n be a large enough integer and $\varepsilon = 2^{-c_0 n}$, where $c_0 > 0$ is a small constant. Suppose given a pair of spatially isolated quantum devices $\mathcal{A}$ and $\mathcal{B}$ such that the probability that the protocol aborts is at most ε. Then conditioned on not aborting, at the end of the protocol Alice and Bob can extract a shared key K of length κn, where $\kappa \approx 1.4\%$, such that $\|\rho_{KE} - \widetilde{\mathrm{Id}}_{\kappa n} \otimes \rho_E\|_1 \leq \varepsilon$, where E designates the environment and $\widetilde{\mathrm{Id}}_{\kappa n}$ the totally mixed state on κn qubits.*

This informal statement hides a tradeoff between the parameters ε, η, and κ: the larger the security parameter ε and the smaller the noise rate η, the higher the key rate κ. Provided ε is chosen large enough, as $\eta \to 0$ our proof guarantees a secure key rate $\kappa \approx 2.5\%$, which with our setting of parameters corresponds to about 15% of the raw key. The maximum noise rate for which we may extract a positive key length is $\eta_{max} \approx 1.2\%$. This is worse than the optimal key rates obtained under the causal independence assumption [MPA11], but still quite reasonable.

1.2 Proof overview and techniques

We now give a slightly more technical introduction to our results. As shown by Renner [Ren05], in order to generate an ε-secure key over κn bits, for some $\kappa > 0$, it is enough to show that the smooth min-entropy

[2] See Figures 1 and 2 for a more detailed description.

[3] Note that this corresponds to estimating the average amount by which the devices' outputs in $\mathbf{B}$ differ from a maximal violation of a Bell inequality based on the CHSH inequality [CHSH69, BC90]. One can easily verify that $\mathtt{opt}$ corresponds to the largest possible violation of that inequality allowed by quantum mechanics

$H_\infty^\varepsilon(B|E) \approx \kappa n$, where B is the string of bits produced by Bob's device and E the side information held by the adversary Eve. For the purposes of this overview we focus on establishing the latter criterion.

The non-smooth variant of the quantum conditional min-entropy is best understood through its relationship with the guessing probability as $H_\infty(B|E) = -\log P_{guess}(B|E)$ [KRS09].[4] The key idea in establishing security in a device-independent scenario is to exploit the tension between high guessing probability (Eve can predict Bob's outputs) and the violation of a Bell inequality (as estimated in the protocol). This connection was first made quantitatively precise in [ABG$^+$07], who established a formula relating the incompatibility between determinism and nonlocality in the case of a single use of the devices (see also [PAM$^+$10] for a similar formula in the no-signalling case, and [NPA08] for a generalization).

In [HR10, MPA11] it is shown how this approach can be extended to repeated uses of the devices under an assumption of causal independence. This assumption is used to derive a product rule for the guessing probability:

$$P_{guess}(B_1 \dots B_n|E) \lessapprox \prod_i P_{guess}(B_i|E). \tag{1}$$

A sufficiently large guessing probability then implies the existence of an i_0 such that $P_{guess}(B_{i_0}|E) \geq 1 - \ln(2)\kappa$, and one can conclude using the techniques in [PAM$^+$10]. It is not hard to see, however, that (1) completely fails without the assumption of independence; this failure is the main difficulty in establishing security in the most general scenario. In order to get around this our proof combines three main ingredients.

1. We first observe that the use of the guessing probability is *too strong* a criterion for obtaining privacy, and instead focus on establishing a lower bound on its robust analogue, the ε-*smooth* min-entropy. The use of this criterion, instead of the more usual guessing probability, already constitutes a major departure between our work and previous analyses of DIQKD. We prove the following operational interpretation: if $H_\infty^\varepsilon(B|E) \leq \kappa n$, then there exists roughly κn bits of classical side information V about B such that $P_{guess}(B|VE) \geq \text{poly}(\varepsilon/n)$. Note the two key differences with the condition $P_{guess}(B|E) \geq 2^{-\kappa n}$ that follows from the weaker assumption $H_\infty(B|E) \leq \kappa n$: additional side information, given by V, is required, but provided ε is not too small the guessing probability is now much larger.

This characterization of the smooth min-entropy was already sketched in [VV12]. It is based on a so-called "quantum reconstruction paradigm", which takes its source in the theory of extractors. Originally developed by Trevisan [Tre01] in the design and analysis of a family of classical extractors with very short seed, it was extended to the setting of quantum side information in order to extend the security of Trevisan's extractor to the case of quantum adversaries in [DV10, DPVR12]. Interestingly, the bits of V can be chosen as a subset of the bits of an encoding of the string B using a list-decodable code. In [VV12] the use of a *local* list-decodable code was crucial; here any good code will do.

2. Recall that our goal is to obtain a contradiction between this characterization of the hypothetical adversary and the non-signalling condition that follows from the spatial isolation of the devices. As we discussed in the introduction, the key difficulty that needs to be addressed is the possibility for a combination of correlations and noise in the devices' successive output bits. In [VV12] the absence of noise was used to guarantee that the "advice bits" V could be obtained from the output of either device, and so Alice's device could provide the advice bits to Eve allowing them (by using the fact that the devices' outputs pass the CHSH tests) to jointly guess Bob's output and thus violate no signalling. Here we cannot make this assumption, and instead work directly with the following strategy for the adversary: Eve first *guesses* a string $\hat{V}$ uniformly at random, hoping that $\hat{V} = V$; she then measures E according to the measurement that optimizes

[4] We refer to Section 2 for complete definitions.

$P_{guess}(B|\hat{V}E)$. Let $\hat{E}$ be the random variable describing her outcome. Since Eve does not require any input she can perform her measurement even *before* the string B is obtained from the device, and we may apply Baye's rule to obtain

$$\prod_i \Pr(B_i = \hat{E}_i | B_{i-1} = \hat{E}_{i-1}, \ldots, B_1 = \hat{E}_1) = P_{guess}(B|\hat{V}E)$$
$$\geq P_{guess}(B|VE) \Pr(\hat{V} = V) \geq \text{poly}(\varepsilon/n) 2^{-\kappa n}. \quad (2)$$

From (2) we may deduce, as we previously did from (1), the existence of an index i_0 such that $\Pr(B_{i_0} = \hat{E}_{i_0} | B_{i_0-1} = \hat{E}_{i_0-1}, \ldots, B_1 = \hat{E}_1) \geq 1 - \ln(2)\kappa - o_n(1)$, provided ε is chosen large enough.

Here we encounter a new difficulty: compared to (1), the conditioning on $B_{i_0-1} = \hat{E}_{i_0-1}, \ldots, B_1 = \hat{E}_1$ performed in (2) could, in general, drastically influence the joint distribution $P_{A_{i_0} B_{i_0} X_{i_0} Y_{i_0}}$ of the devices' inputs and outputs in round i_0. For instance, $\hat{E}$ could have the property of matching the string B in the first few bits if and only if inputs to the i_0-th round (which are publicly revealed in the protocol) are, say, $(0,0)$. If this is the case there is no hope to exploit the no-signalling condition: devices whose inputs are fixed can reproduce *any* joint distribution.[5]

3. The third and last step of the proof consists in using the specific structure of Eve's measurement in order to bound the disturbance introduced by the conditioning. Recall that Eve first guesses the string $\hat{V}$ uniformly at random, and then measures her system E. Crucially however, the event $\hat{V} = V$, though it has very small probability $2^{-\kappa n}$, is independent from any of the other events we consider. Hence the conditioning in (2) only affects the distribution $P_{A_{i_0} B_{i_0} X_{i_0} Y_{i_0}}$ to the extent that conditioning on success of the second step in Eve's measurement does.

The fact that this second step has a relatively high success probability $\text{poly}(\varepsilon/n)$ enables us to complete the proof. To show that conditioning on a (somewhat) high-probability event cannot influence the distribution $P_{A_i B_i X_i Y_i}$ in many rounds i we use the chain rule for mutual information and Pinsker's inequality. Together these two ingredients guarantee that $P_{A_i B_i X_i Y_i}$ remains close to no-signalling for a large fraction of rounds i.[6] We may then use techniques from [PAM+10] to derive a contradiction between no-signalling and the violation of a Bell inequality (in our setting, the CHSH inequality): it must be that $H_\infty^\varepsilon(B|E) > \kappa n$, and hence an ε-secure key of length $\approx \kappa n$ can be extracted from B (after the usual information reconciliation step).

1.3 Perspective

We have not attempted to optimize the relationship between the parameters κ, η and ε describing the key rate, the noise rate and the security parameter respectively, and it is likely that the explicit dependency stated in Theorem 8 can be improved by tightening our arguments. It is an interesting question to find out whether our approach can lead to a trade-off as good as the one that has been shown to be achievable under additional assumptions on the devices [MPA11]. One possibility for improvement would be to bias the users' input distribution towards the pair of inputs $(2,1)$ from which the raw key is extracted, as was done in e.g. [AMP06]: indeed, only a very small fraction of the rounds are eventually required to estimate the violation of the CHSH condition.

[5] Note that this difficulty does not arise in the context of randomness certification, as in that context one may assume, as was done in [VV12], that the inputs to the devices are not revealed to the adversary.

[6] Interestingly, a similar argument was used in the proof of a parallel repetition theorem for no-signalling strategies [Hol09, Lemma 9.6].

Our proof crucially makes use of quantum mechanics to model the devices and the adversary. Can one obtain a fully device-independent proof of security of QKD against adversaries that are only restricted by the no-signalling principle? Barrett et al. [BCK12b] recently showed that such security is achievable in principle; however their protocol is highly inefficient and does not tolerate noisy devices.

It would be interesting to compare our characterization of the smooth min-entropy, proven with the help of the quantum reconstruction paradigm, with other operational interpretations of this quantity [KRS09]. One could also contrast it with the recently proven chain rules for min-and max entropy [VDTR12].

Putting our results in a broader context, a major ongoing challenge remains to bridge the gap between theoretical security proofs of quantum key distribution and practical implementations. The model of device-independence only addresses one — admittedly major — aspect of this gap, and it will be interesting to investigate whether other discrepancies, such as e.g. the possibility for side-channel attacks, or weaknesses stemming from the possible mishandling of their devices by the users (such as re-using devices that should have been securely discarded [BCK12a]), can also be incorporated in security proofs.

Organization of the paper. We start with some preliminaries in Section 2, introducing our notation, the information-theoretic quantities that will be used. We also summarize the main parameters of our protocol, which is described in Figures 1 and 2. In Section 3 we formally state our result and outline the security proof. The two main ingredients are the analysis of Protocol B, which is given in Section 4, and the "quantum reconstruction paradigm" introduced in Section 5. Finally, Section 6 contains probabilistic and information-theoretic lemmas used in some of the proofs.

Acknowledgments. We thank Anthony Leverrier for many useful comments on a preliminary version of this manuscript.

2 Preliminaries

We assume familiarity with basic concepts and standard notation in quantum information, and refer to the books [NC00, Wil11] for detailed introductions.

Notation. We use roman capitals $A, B, \ldots, X$ both to refer to random variables and the registers, classical or quantum, that contain them. Calligraphic letters $\mathcal{A}, \mathcal{B}, \ldots, \mathcal{X}$ are used to refer to the underlying Hilbert space. $\mathrm{D}(\mathcal{X})$ denotes the set of density operators (non-negative matrices with trace 1) on $\mathcal{X}$. For an arbitrary matrix A on $\mathcal{X}$ we let $\|A\|_1 = \mathrm{Tr}\sqrt{AA^\dagger}$ denote its Schatten 1-norm. ln denotes the natural logarithm and log the logarithm in base 2. For $x \in [0,1]$, $H(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function. We abuse notation and also write $H(\rho) = -\mathrm{Tr}(\rho\ln(\rho))$, when $\rho \in \mathrm{D}(\mathcal{X})$ is a density matrix, for the von Neumann entropy.

Information theoretic quantities. We summarize key properties of some standard quantities in quantum information theory that will be used throughout. Given a density matrix $\rho \in \mathrm{D}(\mathcal{A})$, its von Neuman entropy is $H(\rho) := -\mathrm{Tr}(\rho\ln\rho)$. Given a classical-quantum state $\rho_{XA} = \sum_x p_x|x\rangle\langle x| \otimes \rho_x \in \mathrm{D}(\mathcal{X}\mathcal{A})$, where for every x $\rho_x \in \mathrm{D}(\mathcal{A})$, we define the conditional entropy as $H(A|X)_\rho := \sum_x p_x H(\rho_x)$. Given a state ρ_{ABX}, where X is classical, the conditional mutual information is defined as

$$I(A:B|X)_\rho := H(A|X)_\rho + H(B|X)_\rho - H(AB|X)_\rho.$$

We will make use of the following quantum analogue of the classical Pinsker's inequality (see e.g. Theorem 11.9.1 in [Wil11] for a proof): for any $\rho_{AB} \in \mathrm{D}(\mathcal{AB})$,

$$\left\|\rho_{AB} - \rho_A \otimes \rho_B\right\|_1^2 \leq (2\ln 2)\, I(A:B)_\rho. \tag{3}$$

The most important information measure in our context is the quantum conditional min-entropy, first introduced in [Ren05], and defined as follows.

Definition 2. *Let ρ_{AB} be a bipartite density matrix. The* min-entropy *of A conditioned on B is defined as*

$$\mathrm{H}_{min}(A|B)_\rho := \max\{\lambda \in \mathbb{R} : \exists \sigma_B \in \mathrm{D}(\mathcal{B}) \text{ s.t. } 2^{-\lambda}\,\mathrm{Id}_A \otimes \sigma_B \geq \rho_{AB}\}.$$

We will often drop the subscript ρ when there is no doubt about what underlying state is meant. The smooth min-entropy is defined as follows.

Definition 3. *Let $\varepsilon \geq 0$ and ρ_{AB} a bipartite density matrix. The* ε-smooth min-entropy *of A conditioned on B is defined as*

$$\mathrm{H}^\varepsilon_{min}(A|B)_\rho := \max_{\tilde{\rho}_{AB} \in B(\rho_{AB},\varepsilon)} \mathrm{H}_{min}(A|B)_{\tilde{\rho}},$$

where $B(\rho_{AB},\varepsilon)$ is a ball of sub-normalized states of radius ε around ρ_{AB}.[7]

The CHSH condition. Our results are based on the violation of a specific Bell inequality. The inequality we use is a simple extension of the CHSH inequality, similar to the so-called "chained inequality" for two inputs [BC90].

Let $\mathcal{A}$ and $\mathcal{B}$ designate two spatially isolated devices. There are three possible choices of inputs $x \in \{0,1,2\}$ to $\mathcal{A}$, and two possible inputs $y \in \{0,1\}$ to $\mathcal{B}$. Each of the 6 possible pairs of inputs is chosen with uniform probability $1/6$. The devices produce outputs $a, b \in \{0,1\}$ respectively. In case both inputs were in $\{0,1\}$, the constraint on the outputs is the CHSH parity constraint $a \oplus b = x \wedge y$ [CHSH69]. If the inputs are $(2,1)$ the constraint is that the outputs (a,b) should satisfy $a \oplus b = 0$. Finally, for the remaining pair of inputs $(2,0)$ all pairs of outputs are valid. We will refer to this set of constraints collectively as "the CHSH condition".

Let $\mathtt{opt}$ be the maximum probability with which any two isolated devices, obeying the laws of quantum mechanics, may produce outputs satisfying the CHSH condition. It is not hard to show that $\mathtt{opt} = (2/3)\cos^2 \pi/8 + (1/3)$, which is achieved using the following strategy. The devices are initialized in a single EPR pair, each device holding one qubit. On input 0, $\mathcal{A}$ is measured in the computational basis, and on input 1 it is measured in the Hadamard basis. On input 0, $\mathcal{B}$ is measured in the computational basis rotated by $\pi/8$. If $\mathcal{A}$ gets input 2, or if $\mathcal{B}$ gets input 1, they are measured in the computational basis rotated by $3\pi/8$.

Parameters. For convenience, we summarize here the main parameters of the key distribution protocol described in Figures 1 and 2.

- m is the total number of rounds in the protocol (in each round, an input to each of $\mathcal{A}, \mathcal{B}$ is chosen, and an output is collected).

[7] Theoretically any distance measure could be used to define an ε-ball. As has become customary, we use the *purified distance*, $P(\rho,\sigma) := \sqrt{1 - F(\rho,\sigma)^2}$, where $F(\cdot,\cdot)$ is the fidelity.

Protocol A

1. Let m and $\varepsilon, C_\eta > 0$ be parameters given as input. Let C_γ be the constant specified in Theorem 8, and set $\gamma = (C_\gamma / C_\eta^2) \ln(1/\varepsilon)/m$.

2. Alice and Bob run Protocol B for m steps, choosing inputs $x \in \{0,1,2\}^m$ (resp. $y \in \{0,1\}^m$) and obtaining outcomes $a \in \{0,1\}^m$ (resp. $b \in \{0,1\}^m$). Let η be the observed error rate.

3. Alice and Bob publicly reveal their choices of inputs. Let $\mathbf{C}$ be the set of rounds i in which $(x_i, y_i) = (2,1)$. If $||\mathbf{C}| - m/6| > 10\sqrt{\ln(1/\varepsilon)}$ they abort the protocol.

4. Alice and Bob perform information reconciliation on their outputs in $\mathbf{C}$, with Bob sending a message of $\ell \leq H(2\eta)|\mathbf{C}| + \log(2/\varepsilon)$ bits to Alice.

5. Let $\kappa = \kappa(\eta)$ be as specified in Theorem 8. Alice and Bob perform privacy amplification using e.g. two-universal hashing, extracting a shared key of length $(\kappa - H(2\eta) - O(\log(1/\varepsilon)/m))|\mathbf{C}|$ from the common $|\mathbf{C}|$-bit string they obtained at the end of the previous step.

Figure 1: The device-independent key distribution protocol, Protocol A

- $\mathbf{B}$ are the rounds selected to perform parameter estimation. They are chosen uniformly at random under the constraint that $|\mathbf{B}| = \gamma m$, for some $\gamma > 0$ specified in the protocol. .

- η is the error rate, as measured in the rounds in $\mathbf{B}$: η is such that the fraction of rounds in $\mathbf{B}$ satisfying the CHSH condition is at least $\mathtt{opt} - \eta$.

- $\mathbf{C} \subseteq [m]$ are the *check* rounds. Those are rounds in which the inputs to $(\mathcal{A}, \mathcal{B})$ are $(2,1)$. Since the inputs are chosen uniformly at random, the number of check rounds $|\mathbf{C}|$ is highly concentrated around $m/6$.

- The target min-entropy rate κ. This is the rate of min-entropy that the users Alice and Bob expect to be present in the check rounds, provided the protocol did not abort. Once information reconciliation and privacy amplification have been performed, a secret key of length roughly $(\kappa - H(2\eta))|\mathbf{C}|$ will be produced.

- ε is the security parameter: the statistical distance from uniform of the extracted key (conditioned on the eavesdropper's side information). Precisely, if K denotes the system containing the extracted key of length ℓ_K, we will obtain that $\|\rho_{K\mathcal{E}'} - \rho_{U_{\ell_K}} \otimes \rho_{\mathcal{E}'}\|_1 \leq \varepsilon$, where $\mathcal{E}'$ is a register containing all the side information available to an arbitrary quantum eavesdropper in the protocol, and $\rho_{U_{\ell_K}}$ is the totally mixed state on ℓ_K qubits.

3 Analysis of the key distribution protocol

The analysis of Protocol A, and the proof of Theorem 1, is performed in two steps. The first, main step consists in proving a lower bound on the quantum smooth conditional min-entropy $H_{min}^{\varepsilon}(B_{\mathbf{C}}|XYA_{\mathbf{B}}B_{\mathbf{B}}\mathcal{E})$ of the outputs obtained by Bob in the check rounds $\mathbf{C}$. This lower bound will depend on the "error rate" η that

Protocol B

1. Let m, γ and C_η be parameters given as input.
2. Repeat, for $i = 1, \ldots, m$:
   2.1 Alice picks $x_i \in \{0,1,2\}$, and Bob picks $y_i \in \{0,1\}$, uniformly at random. They input x_i, y_i into their respective device, obtaining outputs $a_i, b_i \in \{0,1\}$ respectively.
3. Alice chooses a random subset $\mathbf{B} \subseteq [m]$ of size γm and shares it publicly with Bob. Alice and Bob announce their input/output pairs in $\mathbf{B}$, and compute the fraction of pairs satisfying the CHSH condition. Let $(\mathtt{opt} - \eta')$ be this fraction. Set $\eta := \max(\eta', C_\eta)$.

Figure 2: Theorem 8 shows that, at the end of protocol B, the bits $B_{\mathbf{C}}$ generated by Bob's device in the check rounds $\mathbf{C}$ both have high smooth min-entropy, conditioned on the adversary's arbitrary quantum side information.

is estimated by the users in the sub-protocol B (see Figures 1 and 2 for a description of protocols A and B respectively). Here the lower bound is taken conditioned on the state of an arbitrary quantum adversary (whom we will call Eve and refer to indiscriminately as "the adversary" or "the eavesdropper") in the protocol, who has access to the information $X, Y, A_{\mathbf{B}}, B_{\mathbf{B}}$ revealed publicly in the course of the protocol, as well as to a quantum system $\mathcal{E}$ which may initially have been correlated with the systems $\mathcal{A}$, $\mathcal{B}$ of the devices. Such an estimate is stated in Theorem 8 in Section 3.3 below.

The second step consists in showing that there exists appropriate protocols for the information reconciliation and privacy amplification steps, Steps 4 and 5 in Protocol A respectively, such that the lower bound on the conditional min-entropy from the first step guarantees the security (distance from uniform from the point of view of the adversary) and correctness (Alice and Bob should obtain the same key) of the key that is extracted. This step is standard, and all the ingredients required already appear in the literature. We summarize the result as Lemma 4 in Section 3.2 below.

Theorem 1 follows immediately by combining Theorem 8 and Lemma 4.

3.1 Probability space

Before stating and proving our technical results, it will be useful to formally define all the random variables and events that will be necessary for their analysis.

Modeling the devices. Fix a pair of spatially isolated devices $(\mathcal{A}, \mathcal{B})$. Alice holds system $\mathcal{A}$, and Bob system $\mathcal{B}$. In addition, the adversary Eve holds a system $\mathcal{E}$, arbitrarily correlated with but spatially isolated from $\mathcal{A}$ and $\mathcal{B}$. Let $\rho_{A_1 B_1 \mathcal{E}}$ be the density matrix describing the joint state of the system at the start of the protocol.

We define the following random variables and events. $X \in \{0,1,2\}^m$ and $Y \in \{0,1\}^m$ are two uniformly distributed random variables, used to represent the inputs to $\mathcal{A}, \mathcal{B}$ respectively. $A, B \in \{0,1\}^m$ are random variables denoting the outputs produced by the devices, when sequentially provided their respective inputs X, Y. We will always use $\mathbf{C} \subseteq [m]$ to denote the set of "check" rounds, in which $(X_i, Y_i) = (2,1)$, and $\mathbf{B} \subseteq [m]$ the set of "Bell" rounds chosen by Alice and Bob to perform parameter estimation.

Let $\rho_{\mathcal{A}_i\mathcal{B}_i}$ denote the reduced state of devices $\mathcal{A}$ and $\mathcal{B}$ in the i-th round of the protocol (before they have been provided their input). Formally,

$$\rho_{\mathcal{A}_i\mathcal{B}_i} \propto \Big(\prod_{j<i} M_{X_j}^{A_j} \otimes N_{Y_j}^{B_j}\Big)\, \rho_{\mathcal{A}_1\mathcal{B}_1}\, \Big(\prod_{j<i} \big(M_{X_j}^{A_j}\big)^\dagger \otimes \big(N_{Y_j}^{B_j}\big)^\dagger\Big),$$

where $\{M_{X_j}^{A_j}\}$ and $\{N_{Y_j}^{B_j}\}$ are the Krau operators corresponding to the measurement on device $\mathcal{A}$ and $\mathcal{B}$ in round j respectively, and $\rho_{\mathcal{A}_i\mathcal{B}_i}$ is normalized. Here $\rho_{\mathcal{A}_1\mathcal{B}_1} = \mathrm{Tr}_{\mathcal{E}}(\rho_{\mathcal{A}_0\mathcal{B}_0\mathcal{E}})$ is the reduced state of the devices at the start of the protocol. It is important to note that for any i the state $\rho_{\mathcal{A}_i\mathcal{B}_i}$ may depend on a measurement that is performed on system $\mathcal{E}$, as soon as a particular outcome of that measurement is fixed (or conditioned on).

Measuring the violation of the CHSH condition. Given a set $S \subseteq [m]$ and $\delta > 0$, $\mathrm{CHSH}_{\mathcal{A}\mathcal{B}}(S,\delta)$ is the event that the tuple (X, Y, A, B) satisfies the CHSH condition (as described in Section 2) in a fraction at least $\mathtt{opt} - \delta$ of the rounds indicated by S. If S is omitted, $\mathrm{CHSH}_{\mathcal{A}\mathcal{B}}(\delta) = \mathrm{CHSH}_{\mathcal{A}\mathcal{B}}([m],\delta)$. Letting $Z \in \{0,1\}^m$ be the indicator random variable of the CHSH condition *not* being satisfied in any given round, we can write

$$\mathrm{CHSH}_{\mathcal{A}\mathcal{B}}(S,\delta) \equiv \Big\{\frac{1}{|S|}\sum_{i\in S} Z_i \leq (1-\mathtt{opt}) + \delta\Big\}.$$

We also define $\mathrm{VIOL}_{\mathcal{A}\mathcal{B}}(i)$, where $i \in [m]$, to express the expected amount by which the CHSH condition in round i is satisfied:

$$\mathrm{VIOL}_{\mathcal{A}\mathcal{B}}(i) = \mathrm{E}[\,Z_i\,] - (1-\mathtt{opt}),$$

where here the expectation is taken over the choice of inputs (X_i, Y_i) in round i, and over the randomness in the devices' own measurements in round i. Note that $\mathrm{VIOL}_{\mathcal{A}\mathcal{B}}(i)$ implicitly depends on the specific state of the devices in round i, which may be affected by previous input/outputs $(X_{<i}, Y_{<i}, A_{<i}, B_{<i})$ obtained in the protocol as well as on other events that may be conditioned on. Hence an expression such as $\Pr(\mathrm{VIOL}_{\mathcal{A}\mathcal{B}}(i) < \delta | E)$, for some event E, indicates the average probability, over all possible $e \in E$, that the devices satisfy the CHSH condition in round i with probability at least $\mathtt{opt} - \delta$, provided their inputs are distributed according to the conditional distribution $(X_i, Y_i)|E = e$, and when performed on the post-measurement state of $\mathcal{A}\mathcal{B}$ in round i conditioned on $E = e$.

For any $\delta > 0$ we let $\mathrm{VIOL}_{\mathcal{A}\mathcal{B}}(\delta)$ be the event that $(1/m)\sum_i \mathrm{VIOL}_{\mathcal{A}\mathcal{B}}(i) \leq \delta$.

The adversary. We include a description of random variables that depend on the adversary, holding the quantum system $\mathcal{E}$. The adversary is described in Lemma 9 below; to understand the events below it may be useful to read that lemma's statement first.

Let $E \in \{0,1\}^{|\mathbf{C}|}$ be the random variable that describes the outcome of the measurement on $\mathcal{E}$ described in Lemma 9. Note that this outcome depends on the "advice" that is given to the adversary. We use $\hat{X}, \hat{Y}$ to denote the inputs that are given to the adversary, and $\hat{\textsc{Adv}} \in \{0,1\}^{\alpha m}$ to denote the additional advice bits. Note that these random variables need not equal the actual values $X, Y, \textsc{Adv}$: in general, the adversary's measurement is well-defined for any given advice bits, and E is used to denote its outcome irrespective of whether the advice given was "correct" or not. For any $i \in [m]$, define $\mathrm{GUESS}_{\mathcal{B}\mathcal{E}}(i) \in \{0,1\}$ to be 1 if and only if, either $i \in \mathbf{C}$ and $E_i = B_i$, or $i \notin \mathbf{C}$, and let $\mathrm{GUESS}_{\mathcal{B}\mathcal{E}} = \wedge_i \mathrm{GUESS}_{\mathcal{B}\mathcal{E}}(i)$.

3.2 Information reconciliation and privacy amplification

For convenience, we let $\mathcal{E}' := XYA_{\mathbf{B}}B_{\mathbf{B}}\mathcal{E}$ denote the side information available to the eavesdropper. We show the following lemma, whose proof follows from standard arguments in the analysis of QKD protocols (see e.g. [Ren05]). We provide the relevant details below.

Lemma 4. *Let $\gamma, \varepsilon > 0$. Let $\varepsilon' = 2e^{-\gamma|\mathbf{C}|/400}$. Suppose that, after Step 2 of Protocol A, the condition $\mathrm{H}^{\varepsilon}_{min}(B_{\mathbf{C}}|\mathcal{E}') \geq \kappa|\mathbf{C}|$ is satisfied. Then with probability at least $1-\varepsilon'$, at the end of the protocol Alice and Bob have a common shared key that is 2ε-close to uniform and has length $\mathrm{H}^{\varepsilon}_{min}(B_{\mathbf{C}}|\mathcal{E}') - H(1.1\eta)|\mathbf{C}| - 4\log(1/\varepsilon)$.*

Information reconciliation. We first analyze the information reconciliation step. The following lemma states the conditions that are required for there to exist a satisfactory information reconciliation procedure.

Lemma 5 (Lemma 6.3.4 in [Ren05])**.** *Let $A, B \in \{0,1\}^k$ be two random variables, and $\varepsilon > 0$. Suppose Alice holds A, and Bob holds B. There is an information reconciliation protocol in which Bob communicates $\ell \leq H^{\varepsilon}_{max}(B|A) + \log(2/\varepsilon)$ bits of information about B to Alice and is such that with probability at least $1-\varepsilon$ Alice and Bob both know B at the end of the protocol.*

To apply Lemma 5 it suffices to prove an upper bound on the conditional max-entropy $H^{\varepsilon}_{max}(B_{\mathbf{C}}|A_{\mathbf{C}})$. By definition of the rounds $\mathbf{C}$, the CHSH condition in those rounds imposes that $A_i = B_i$ for all $i \in \mathbf{C}$. Hence, were it not for errors, we would have $H^{\varepsilon}_{\max}(B|A) = 0$. The following claim shows that the bound on the error rate that results from the estimation performed in the rounds $\mathbf{B}$ in Step 3 of Protocol B is enough to guarantee a good upper bound on the conditional max-entropy.

Claim 6. *Suppose Alice and Bob do not abort after Step 3 in Protocol B. Let $\mathbf{C}$ be the set of check rounds, as designated in Step 4 of Protocol A. Then $H^{\varepsilon'}_{max}(B_{\mathbf{C}}|C_{\mathbf{C}}) \leq H(1.1\eta)|\mathbf{C}|$, where $\varepsilon' = 2e^{-\gamma|\mathbf{C}|/400}$.*

Proof. Fix the set $\mathbf{C}$. The set $\mathbf{B}$ chosen by Alice and Bob to perform parameter estimation contains a fraction at least $\gamma/2$ of the rounds in $\mathbf{C}$, except with probability at most $e^{-\gamma|\mathbf{C}|/8}$. The protocol is aborted as soon as more than an η fraction of those rounds are such that $a_i \neq b_i$. Hence with probability at least $1 - e^{-\gamma|\mathbf{C}|/200}$ the total fraction of errors in $\mathbf{C}$ is at most 1.1η. In particular, with probability at least $1 - e^{-\gamma|\mathbf{C}|/400}$ over $A_{\mathbf{C}}$, with probability at least $1 - e^{-\gamma|\mathbf{C}|/400}$, $B_{\mathbf{C}}$ will take on at most $2^{H(1.1\eta)|\mathbf{C}|}$ values. □

Privacy amplification. The following lemma states the existence of a good protocol for privacy amplification.

Lemma 7 (Lemma 6.4.1 in [Ren05])**.** *Suppose the information reconciliation protocol requires at most ℓ bits of communication. Then for any $\varepsilon > 0$ there is a privacy amplification protocol based on two-universal hashing which extracts $\mathrm{H}^{\varepsilon}_{min}(B_{\mathbf{C}}|\mathcal{E}') - \ell - 2\log(1/\varepsilon)$ bits of key.*

Lemma 4 now follows directly by combining Claim 6 with Lemma 7 and the assumption on the conditional min-entropy placed in the lemma.

3.3 A lower bound on the conditional min-entropy

The main result of this section is a lower bound on $\mathrm{H}^{\varepsilon}_{min}(B_{\mathbf{C}}|XYA_{\mathbf{B}}B_{\mathbf{B}}\mathcal{E})$, the quantum smooth conditional min-entropy of the raw key, given by the bits of $\mathcal{B}$'s output string B that falls in the check blocks $\mathbf{C}$.

Theorem 8. *Let $C_\eta > 0$ be given. There exists positive constants C_ε, C_γ (possibly depending on C_η) such that the following hold. Let m be an integer and $\varepsilon \geq 2^{-C_\varepsilon m}$ be given. Let $\gamma = (C_\gamma / C_\eta^2) \ln(1/\varepsilon)/m$ and η be as specified in Protocol A (Figure 1) and Protocol B (Figure 2) respectively. Let κ be any constant such that $\kappa < (\sqrt{2}-1)/(4\ln(2)) - (4/\ln(2))\eta$.*

Suppose that the devices $\mathcal{A}$, $\mathcal{B}$ are such that with probability at least ε the protocol does not abort. Let $\mathcal{E}$ be an auxiliary system held by an eavesdropper, who may also learn (X, Y) and $(A_{\mathbf{B}}, B_{\mathbf{B}})$. Then, conditioned on the protocol not aborting, it holds that

$$\mathrm{H}_{min}^{\varepsilon}(B_{\mathbf{C}}|XYA_{\mathbf{B}}B_{\mathbf{B}}\mathcal{E}) \geq \kappa|\mathbf{C}| - O\big(\ln(1/\varepsilon)\big).$$

We note that the precise relation between the parameters κ and η stated in the theorem is the one that we obtain from our proof; however we have not attempted to optimize it fully and it is likely that one may be able to derive a better dependency. It is also clear from the proof that one may trade off the different constants between each other, depending on whether one is interested in the maximum possible key rate in the presence of very small noise, or to the opposite if one wishes to tolerate as much noise as possible.

The proof of Theorem 8 is based on two lemmas. We state both lemmas, and derive the theorem from them, below; the lemmas themselves are proved in Section 5 and Section 4 respectively.

Our first lemma states that, if the min-entropy condition in the conclusion of the theorem is not satisfied, then there must exist a measurement on the system $\mathcal{E}$, depending on $X, Y, A_{\mathbf{B}}$ and $B_{\mathbf{B}}$, together with some additional "advice" bits of information about $B_{\mathbf{C}}$, whose outcome $E \in \{0,1\}^{|\mathbf{C}|}$ agrees with $B_{\mathbf{C}}$ with reasonable probability.

Lemma 9. *Let $\kappa > 0$ and suppose that $\mathrm{H}_{min}^{\varepsilon}(B_{\mathbf{C}}|XYA_{\mathbf{B}}B_{\mathbf{B}}\mathcal{E}) < \kappa|\mathbf{C}|$. Then there exists an $\alpha = \kappa|\mathbf{C}|/m + 2\gamma + O(\log(m/\varepsilon)/m)$, and a function $f : \{0,1\}^{|\mathbf{C}|} \to \{0,1\}^{(\alpha-2\gamma)m}$ such that, given the bits $\mathrm{ADV} = f_{\mathrm{ADV}}(B_{\mathbf{C}})A_{\mathbf{B}}B_{\mathbf{B}} \in \{0,1\}^{\alpha m}$ together with the inputs X, Y, there exists a measurement on $\mathcal{E}$ that outputs a string $e \in \{0,1\}^{|\mathbf{C}|}$ such that with probability (over the randomness in B and in the measurement) at least $C_E(\varepsilon/m)^6$, where C_E is a universal constant, the equality $e = b_{\mathbf{C}}$ holds.*

The proof of Lemma 9 is based on a "reconstruction"-type argument from [DPVR12]. A very similar argument was already used to establish an analogous lemma in [VV12]. We give the proof of Lemma 9 in Section 5.

Our second lemma states the existence of a "good" round $i_0 \in [m]$ in which both the CHSH condition is satisfied, and the outcome E_{i_0} of the measurement described in Lemma 9 agrees with B_{i_0}, with good probability. Note also the additional condition (4) in the lemma, which states that systems $\mathcal{A}$ and $\mathcal{B}$ are each close to being independent from the random variables X_{i_0}, Y_{i_0} describing the choice of inputs in round i_0. This condition is crucial for condition (5), on the CHSH violation, to be of any use: indeed, without (4) it could in principle be that the conditioning on specific outcomes in previous rounds, including the adversary's outcomes, completely determines the choice of inputs in the i_0-th round (since the adversary's measurement may depend on inputs X, Y to all rounds). As we will see in the proof of Lemma 27, (4) implies that the distribution that arises from the devices' measurements on the states $\rho_{\mathcal{AB}}^{xy}$ is, while not necessarily quantum, still no-signalling, and this will be sufficient for our purposes. (We refer to Section 3.1 for a description of the events $\mathrm{CHSH}_{\mathcal{AB}}$ and $\mathrm{VIOL}_{\mathcal{AB}}$ appearing in the statement of the lemma.)

Lemma 10. *Let $\hat{\mathrm{ADV}}$ be uniformly distributed in $\{0,1\}^{\alpha m}$, and $\eta, \varepsilon > 0$ be such that the following holds:*

$$\Pr\left(\mathrm{CHSH}_{\mathcal{AB}}(\eta) \wedge \mathrm{GUESS}_{\mathcal{BE}}|\mathrm{ADV} = \hat{\mathrm{ADV}}\right) \geq \varepsilon,$$

and let $\alpha = |\textsc{Adv}|/m$. *Then there exists a universal constant* $C_\nu > 0$, *a* $\nu \leq C_\nu \sqrt{\log(1/\varepsilon)/m}$, *an* $i_0 \in [m]$ *and a set* $G_{i_0} \subseteq (\{0,1,2\} \times \{0,1\} \times \{0,1\}^3)^{i_0-1}$ *such that for every* $(x,y,a,b,e) \in G_{i_0}$, *there is a choice of* $\hat{x}_{>i_0}, \hat{y}_{>i_0}$ *such that the following hold:*

$$\max\left\{ \left\| \rho_{\mathcal{A}_{i_0} X_{i_0} Y_{i_0}} - \rho_{\mathcal{A}_{i_0}} \otimes \left(\frac{1}{6} \sum_{x,y} |x,y\rangle\langle x,y| \right) \right\|_1 , \left\| \rho_{\mathcal{B}_{i_0} X_{i_0} Y_{i_0}} - \rho_{\mathcal{B}_{i_0}} \otimes \left(\frac{1}{6} \sum_{x,y} |x,y\rangle\langle x,y| \right) \right\|_1 \right\} \leq \nu, \tag{4}$$

$$\mathrm{VIOL}_{\mathcal{AB}}(i_0) \leq 3\eta + \nu, \tag{5}$$

$$\Pr(\mathrm{GUESS}_{\mathcal{BE}}(i_0)) \geq 1 - 12\ln(2)\alpha - \nu, \tag{6}$$

where in (4) *the state* $\rho_{\mathcal{A}_{i_0}\mathcal{B}_{i_0} X_{i_0} Y_{i_0}}$ *is the (normalized) state of the corresponding systems in round* i_0, *conditioned on* (x,y,a,b,e), *and similarly in* (5) *and* (6) *the violation is estimated conditioned on previous input/outputs to the devices being* (x,y,a,b), *and on Eve making her measurement with inputs* $(x_{<i_0}, 2, \hat{x}_{>i_0})$ *and* $(y_{<i_0}, 1, \hat{y}_{>i_0})$ *and advice string* $\hat{\textsc{Adv}}$ chosen uniformly at random, *and obtaining outcomes* e *as her prediction in rounds* $\mathbf{C} \cap \{1, \ldots, i_0 - 1\}$.

The proof of Lemma 10 in given in Section 4. Based on these two lemmas, we give the proof of Theorem 8.

Proof of Theorem 8. Let (X, Y, A, B) be random variables describing Alice and Bob's choice of inputs to $\mathcal{A}$ and $\mathcal{B}$ respectively, and the outputs obtained, in an execution of Protocol A. Let $E = E(\hat{\textsc{Adv}})$ be the random variable that describes the outcome of the measurement on $\mathcal{E}$ described in Lemma 9, when the advice bits $\hat{\textsc{Adv}}$ are selected uniformly at random (independently from A and B). Let $\textsc{Adv} = f_{\textsc{Adv}}(B_{\mathbf{C}}) A_{\mathbf{B}} B_{\mathbf{B}}$ denote the "correct" advice bits.

The proof proceeds by contradiction. Assume that there existed a pair of devices $(\mathcal{A}, \mathcal{B})$ such that

$$\Pr\left(\mathrm{CHSH}_{\mathcal{AB}}(\mathbf{B}, \eta)\right) \geq \varepsilon \qquad \text{and} \qquad \mathrm{H}^{\varepsilon}_{min}(B_{\mathbf{C}} | XY A_{\mathbf{B}} B_{\mathbf{B}} \mathcal{E}) < \kappa |\mathbf{C}|, \tag{7}$$

where $\varepsilon, \eta, \kappa$ are as in the statement of the theorem. Let $\mathrm{GUESS}_{\mathcal{BE}}(\hat{\textsc{Adv}})$ denote the event that $E = B_{\mathbf{C}}$. Using Lemma 9, we deduce from (7) that the following must hold:

$$\begin{aligned} &\Pr\left(\mathrm{CHSH}_{\mathcal{AB}}(\mathbf{B}, \eta) \wedge \mathrm{GUESS}_{\mathcal{BE}}(\hat{\textsc{Adv}}) | \hat{\textsc{Adv}} = \textsc{Adv}\right) \\ &\quad = \Pr\left(\mathrm{GUESS}_{\mathcal{BE}}(\hat{\textsc{Adv}}) | \mathrm{CHSH}_{\mathcal{AB}}(\mathbf{B}, \eta), \hat{\textsc{Adv}} = \textsc{Adv}\right) \Pr\left(\mathrm{CHSH}_{\mathcal{AB}}(\mathbf{B}, \eta) | \hat{\textsc{Adv}} = \textsc{Adv}\right) \\ &\quad \geq C_E (\varepsilon/m)^6 \cdot \varepsilon, \end{aligned} \tag{8}$$

where C_E is the constant from Lemma 9. Since the rounds $\mathbf{B}$ are chosen uniformly at random, Claim 11 below states that, for any $0 \leq \beta \leq 1$:

$$\Pr\left(\mathrm{CHSH}_{\mathcal{AB}}((1+\beta)\eta) | \mathrm{CHSH}_{\mathcal{AB}}(\mathbf{B}, \eta)\right) \geq 1 - e^{-2\beta^2 \eta^2 \gamma m}, \tag{9}$$

where $\gamma = |\mathbf{B}|/m$. Choose $\beta = 1/3$, and let $\eta' := 4\eta/3$. Provided C_γ is chosen large enough, and since by definition $\eta \geq C_\eta$ the choice of γ made in the theorem is such that $\gamma \geq \log(2m^6/C_E\varepsilon^7)/((2/9)\eta^2 m)$, so that $e^{-2\beta^2\eta^2\gamma m} \leq C_E \varepsilon^7/(2m^6)$. Hence we obtain the following by combining (8) and (9):

$$\Pr\left(\mathrm{CHSH}_{\mathcal{AB}}(\eta') \wedge \mathrm{GUESS}_{\mathcal{BE}}(\hat{\textsc{Adv}}) | \, \hat{\textsc{Adv}} = \textsc{Adv}\right) \geq C_E(\varepsilon^7/(2m^6)) =: \varepsilon'. \tag{10}$$

We may now apply Lemma 10. Let $\nu = C_\nu \sqrt{\log(1/\varepsilon')/m}$, and $i_0 \in [m]$ be the "good" round that is promised by the lemma. We proceed to show that the existence of such a round leads to a contradiction by appealing to the guessing lemma, Lemma 27.

Consider the following setup. Alice, Bob and Eve prepare their devices by selecting a random string of inputs $\hat{X}, \hat{Y}$ for Eve, except that $\hat{X}_{i_0} = 2$ and $\hat{Y}_{i_0} = 1$ always. Eve guesses the advice bits $\hat{\text{ADV}}$ at random and makes a prediction $E = e$. Alice and Bob then use their devices up to round $i_0 - 1$ by choosing inputs $(X_{<i_0}, Y_{<i_0}) = (\hat{X}_{<i_0}, \hat{Y}_{<i_0})$. They verify that the resulting outputs $a_{<i_0}, b_{<i_0}$ are such that $(x_{<i_0}, y_{<i_0}, a_{<i_0}, b_{<i_0}, e_{<i_0}) \in G_{i_0}$; if not they abort. Upon having succeeded in this conditioning they separate and play the guessing game. Alice holds system $\mathcal{A}$, while Bob holds system $\mathcal{B}$.

Lemma 10 shows that all conditions in Lemma 27 are satisfied: it must be that

$$12\ln(2)\alpha + \nu \geq \Big(\frac{\sqrt{2}-1}{2} - 6\eta' - 2\nu\Big) - 75\nu.$$

By definition, provided the constant C_ν is large enough we have $\alpha \leq \kappa/6 + 2\gamma + \nu$, where we used that $|\mathbf{C}| \leq m/6 + 10\sqrt{\ln(1/\varepsilon)}$, as enforced in the protocol, and $\eta' = 4/3\eta$. Re-arranging terms and using the definition of ν and γ we obtain the condition

$$\kappa \;>\; \frac{\sqrt{2}-1}{4\ln(2)} - \frac{4}{\ln(2)}\eta - O\Big(\frac{\log(1/\varepsilon)}{C_\eta^2 m}\Big),$$

which, given the choice of κ made in the theorem, is a contradiction provided C_ε is chosen small enough (possibly depending on C_η). □

Claim 11. *Let $\eta, \gamma > 0$. The following holds for any $0 \leq \beta \leq 1$:*

$$\Pr_S\big(\text{CHSH}((1+\beta)\eta) | \text{CHSH}(S,\eta)\big) \;\geq\; 1 - e^{-2\beta^2\eta^2\gamma m},$$

where the probability is taken over the choice of a random subset $S \subseteq [m]$ of size $|S| = \gamma m$.

Proof. Consider a given run of the protocol. Suppose that the fraction of rounds in which the CHSH condition is not satisfied is at least $(1 - \texttt{opt}) + (1+\beta)\eta$. By a Chernoff bound, a randomly chosen set $S \subseteq [m]$ will of size γm will have at least $((1 - \texttt{opt}) + \eta)\gamma m$ of its rounds with inputs corresponding to the CHSH condition being violated, except with probability at most $e^{-2\beta^2\eta^2\gamma m}$. □

4 Proof of Lemma 10

This section is devoted to the proof of Lemma 10. Let D be the event $\text{CHSH}_{\mathcal{AB}}(\eta) \wedge \text{GUESS}_{\mathcal{BE}}$: the main assumption of the lemma states that $\Pr(D | \text{ADV} = \hat{\text{ADV}}) \geq \varepsilon$. We first prove two preliminary claims which establish that, provided ε is not too small, conditioning on D does not affect either the distribution of inputs (X_i, Y_i) or the reduced density matrices of the inner state of each device's system in most rounds i by too much.

Claim 12. *Suppose that, in Protocol B, Alice and Bob choose inputs $(X,Y) \in \{0,1,2\}^m \times \{0,1\}^m$ uniformly at random, obtaining outcomes $A, B \in \{0,1\}^m$. Suppose that $\mathcal{E}$ is measured using Eve's guessing measurement (as described in Lemma 9) with inputs $(\hat{X}, \hat{Y}) = (X,Y)$ and advice bits $\hat{\text{ADV}} = \text{ADV}$, resulting in an outcome $E \in \{0,1\}^{|\mathbf{C}|}$. Let $P_{X_i Y_i}$ be the marginal distribution of the inputs in the i-th round,*

conditioned on $(X_{<i}, Y_{<i}, A_{<i}, B_{<i}, E_{<i}) = (x_{<i}, y_{<i}, a_{<i}, b_{<i}, e_{<i}) \in D_{<i}$*, the projection of* D *on the first* $(i-1)$ *coordinates. Then the following bound holds on expectation over* $(x_{<i}, y_{<i}, a_{<i}, b_{<i}, e_{<i})$*:*

$$\frac{1}{m}\sum_i \left\| P_{X_iY_i} - U_{3\times 2} \right\|_1 \leq \sqrt{\frac{\log(1/\varepsilon)}{2m}},$$

where $U_{3\times 2}$ *is the uniform distribution on* $\{0,1,2\} \times \{0,1\}$*.*

Proof. The Shannon entropy $H(X,Y) = \log(6)m$, and conditioned on D, $H(X,Y|D) \geq \log(6)\, m - \log(1/\varepsilon)$. Applying the chain rule,

$$\frac{1}{m}\sum_i H(X_i, Y_i | X_{<i}, Y_{<i}, D_{<i}) \geq \log(6) - \frac{\log(1/\varepsilon)}{m}.$$

Using the classical Pinsker's inequality as $\|P_{X_iY_i} - U_{3\times 2}\|_1 \leq \sqrt{(\log(6) - H(X_i,Y_i))/2}$ and Jensen's inequality we get

$$\frac{1}{m}\sum_i \left\| P_{X_iY_i} - U_{3\times 2} \right\|_1 \leq \sqrt{\frac{\log(1/\varepsilon)}{2m}},$$

proving the claim. □

The fact that D depends both on the choice of inputs (X,Y) and on the adversary's measurement outcome implies that conditioning on D could not only bias the distribution of (X,Y) but also introduce correlations between (X,Y) and the reduced state $\rho_{\mathcal{AB}}$ of the devices. The following claim shows that, if D is an event with large enough probability, the correlations introduced by this conditioning do not affect the reduced state on either $\mathcal{A}$ or $\mathcal{B}$ by too much, for most rounds i.

Claim 13. *Consider the same situation as described in Claim 12. Let* $\rho_{\mathcal{A}_iX_iY_i}$ *denote the reduced density of the joint state of systems* $\mathcal{A}$ *(in round* i*) and* X_i, Y_i*, conditioned on* $(X_{<i}, Y_{<i}, A_{<i}, B_{<i}, E_{<i}) = (x_{<i}, y_{<i}, a_{<i}, b_{<i}, e_{<i}) \in D_{<i}$*. Then the following holds on expectation over* $(x_{<i}, y_{<i}, a_{<i}, b_{<i}, e_{<i})$*:*

$$\frac{1}{m}\sum_i \left\| \rho_{\mathcal{A}_iX_iY_i} - \rho_{\mathcal{A}_i} \otimes \left(\frac{1}{6}\sum_{x,y} |x,y\rangle\langle x,y|\right) \right\|_1 \leq 4\sqrt{\log(1/\varepsilon)/m}.$$

Moreover, the same bound holds when $\mathcal{A}_i$ *is replaced by* $\mathcal{B}_i$*.*

Proof. We use Claim 26. Alice's sequential measurements are taken to be the ones performed on $\mathcal{A}$, while Bob's measurement is the combination of the measurements on $\mathcal{B}$, together with Eve's measurement, on inputs X, Y and advice bits $\hat{\text{ADV}} = \text{ADV}$ obtained from B. We set $\mathbf{X}$ in the claim to be XY here, and the outcomes $\mathbf{B}$ in the claim to BE here. Together with the assumption $\Pr(D|\hat{\text{ADV}} = \text{ADV}) \geq \varepsilon$, the claim shows that

$$\frac{1}{m}\sum_i I\big(\mathcal{A}_i; X_iY_i | D_{<i}\big)_{\rho_{\mathcal{A}_iX_iY_i}} \leq \frac{\log(1/\varepsilon)}{m}.$$

Using Pinsker's inequality (3) together with Jensen's inequality,

$$\frac{1}{m}\sum_i \left\| \rho_{\mathcal{A}_iX_iY_i} - \rho_{\mathcal{A}_i} \otimes \left(\frac{1}{6}\sum_{xy} |x,y\rangle\langle x,y|\right) \right\|_1 \leq 4\sqrt{\log(1/\varepsilon)/m},$$

where we used Claim 12 to show that the marginal distribution of (X_i, Y_i) is close to uniform on $\{0,1,2\} \times \{0,1\}$, even conditioned on $D_{<i}$. □

Let $\mathrm{PROD}_{\mathcal{ABXY}}(\delta)$ be the event that the bound in Claim 13 holds up to error δ, both as stated and when the system $\mathcal{A}$ is replaced by $\mathcal{B}$. The following claim states a bound that is similar to our initial assumption $\Pr(D|\hat{\textsc{Adv}} = \textsc{Adv}) \geq \varepsilon$, except that it adds the additional "tensor product form" condition $\mathrm{PROD}_{\mathcal{ABXY}}$. It also replaces the event that the CHSH condition is satisfied in a large fraction of rounds by the event that, on average over $i \in [m]$, the CHSH condition is *likely* to be satisfied in the i-th round.

Claim 14. *There exists a $\nu \leq 5\sqrt{\log(1/\varepsilon)/m}$ such that*

$$\Pr\left(\mathrm{PROD}_{\mathcal{ABXY}}(\nu) \wedge \mathrm{VIOL}_{\mathcal{AB}}(\eta+\nu) \wedge \mathrm{GUESS}_{\mathcal{BE}} | \hat{\textsc{Adv}} = \textsc{Adv}\right) \geq \varepsilon/20.$$

Proof. Let $Z_i \in \{0,1\}$ be 1 if and only if the CHSH condition is not satisfied in round i. By definition, $\mathrm{E}[Z_i] = (1-\mathtt{opt}) + \mathrm{VIOL}_{\mathcal{AB}}(i)$. Let $W_i = \mathrm{E}[Z_i] - Z_i$ and $W_{\leq i} = W_1 + \cdots + W_i$. $(W_{\leq i})_i$ is a Martingale, and by Azuma's inequality (see Lemma 24), for any $\beta > 0$

$$\begin{aligned}\Pr\left(\frac{1}{m}\sum_i \mathrm{VIOL}_{\mathcal{AB}}(i) + (1-\mathtt{opt}) > \frac{1}{m}\sum_i Z_i + \beta\right) &= \Pr\left(\frac{1}{m}\sum_i W_i > \beta\right) \\ &\leq e^{-\beta^2 m/2}.\end{aligned}$$

Since the string $\hat{\textsc{Adv}}$ is chosen by the adversary uniformly at random, we may further condition the equations above on $\hat{\textsc{Adv}} = \textsc{Adv}$ without affecting their validity. Note that the event $\mathrm{CHSH}_{\mathcal{AB}}(\eta)$ is equivalent to $\frac{1}{m}\sum_i Z_i \leq (1-\mathtt{opt}) + \eta$. Choosing $\beta = \sqrt{2\log(8/\varepsilon)/m}$, so that $e^{-\beta^2 m/2} < \varepsilon/8$, and using the assumption $\Pr(D|\hat{\textsc{Adv}} = \textsc{Adv}) \geq \varepsilon$ to further condition on $D = \mathrm{CHSH}_{\mathcal{AB}}(\eta) \wedge \mathrm{GUESS}_{\mathcal{BE}}$ we get

$$\Pr\left(\frac{1}{m}\sum_i \mathrm{VIOL}_{\mathcal{AB}}(i) > \eta + \beta \middle| D, \hat{\textsc{Adv}} = \textsc{Adv}\right) \leq 1/8. \tag{11}$$

To conclude, note that Claim 13 together with Markov's inequality implies that conditioned on D the event $\mathrm{PROD}_{\mathcal{ABXY}}(\beta')$ holds with probability at least $1/5$ for $\beta' = 5\sqrt{\log(1/\varepsilon)/m}$. Together with (11) and the assumption $\Pr(D|\hat{\textsc{Adv}} = \textsc{Adv}) \geq \varepsilon$, this proves the claim. □

Proof of Lemma 10. Let ν be as in Claim 14, and for a parameter $\delta > 0$ define

$$\mathrm{GOOD}(i,\delta) = (\mathrm{VIOL}_{\mathcal{AB}}(i) \leq (\eta+\nu)/\delta) \wedge (\mathrm{PROD}_{\mathcal{ABXY}}(i) \leq 2\nu/\delta),$$

where $\mathrm{PROD}_{\mathcal{ABXY}}(i) \leq 2\nu/\delta$ denotes the event that

$$\left\|\rho_{\mathcal{A}_i X_i Y_i} - \rho_{\mathcal{A}_i} \otimes \left(\frac{1}{6}\sum_{x,y} |x,y\rangle\langle x,y|\right)\right\|_1 \leq 2\nu/\delta \quad \text{and} \quad \left\|\rho_{\mathcal{B}_i X_i Y_i} - \rho_{\mathcal{B}_i} \otimes \left(\frac{1}{6}\sum_{x,y} |x,y\rangle\langle x,y|\right)\right\|_1 \leq 2\nu/\delta.$$

Choosing $\delta = 1/3$, by Claim 14 together with Markov's inequality there must exist a subset $S \subseteq [m]$ of size at least $|S| \geq m/2$ such that

$$\Pr\left(\wedge_{i\in S}\left(\mathrm{GOOD}(i,1/3) \wedge \mathrm{GUESS}_{\mathcal{BE}}(i)\right) | \hat{\textsc{Adv}} = \textsc{Adv}\right) \geq \varepsilon/20.$$

Removing the conditioning on $\hat{\textsc{Adv}} = \textsc{Adv}$ (so that the adversary now guesses an advice string $\hat{\textsc{Adv}}$ uniformly at random), $\varepsilon/20$ is replaced by $2^{-\alpha m}\varepsilon/20$. Applying Baye's rule, we find that there must exist an $i_0 \in S$ such that

$$\Pr\left(\mathrm{GOOD}(i_0,1/3) \wedge \mathrm{GUESS}_{\mathcal{BE}}(i_0) | \wedge_{i\in S, i<i_0}\left(\mathrm{GOOD}(i,1/3) \wedge \mathrm{GUESS}_{\mathcal{BE}}(i)\right)\right) \geq 2^{-2\alpha}(\varepsilon/20)^{2/m}.$$

To conclude the proof of the lemma, it suffices to note that, once inputs and outputs to the devices in rounds prior to i_0 have been fixed, the event $\mathrm{GOOD}(i_0, 1/3)$ is deterministic. Finally, in order for the adversary's guess in round i_0 to be meaningful, we further condition on inputs in round i_0 to be the pair $(2,1)$, which (given the condition $\mathrm{PROD}_{\mathcal{ABXY}}(i_0)$) happens with probability $1/6 \pm 6 \cdot 9\nu$. □

5 The quantum reconstruction paradigm

In this section we prove a general lemma, Lemma 21 in Section 5.2 below, from which Lemma 9 is deduced in Section 5.3. We start with some useful preliminary definitions and known results.

5.1 Combinatorial preliminaries

We first define extractors.

Definition 15. *A function $Ext : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a* quantum-proof *(or simply* quantum*)* (k,ε)-strong extractor *if for all states ρ_{XE} classical on X with $\mathrm{H}_{min}(X|E) \geq k$, and for a uniform seed $Y \in \{0,1\}^d$, we have*

$$\frac{1}{2}\left\|\rho_{Ext(X,Y)YE} - \rho_{U_m} \otimes \rho_Y \otimes \rho_E\right\|_1 \leq \varepsilon,$$

where ρ_{U_m} is the fully mixed state on a system of dimension 2^m.

We will use list-decodable codes.

Definition 16. *A code $C : \{0,1\}^n \to \{0,1\}^{\bar{n}}$ is said to be (ε, L)-list-decodable if every Hamming ball of relative radius $1/2 - \varepsilon$ in $\{0,1\}^{\bar{n}}$ contains at most L codewords.*

There exist list-decodable codes with the following parameters.

Lemma 17. *For every $n \in \mathbb{N}$ and $\delta > 0$ there is a code $C_{n,\delta} : \{0,1\}^n \to \{0,1\}^{\bar{n}}$, which is $(\delta, 1/\delta^2)$-list-decodable, with $\bar{n} = \mathrm{poly}(n, 1/\delta)$. Furthermore, $C_{n,\delta}$ can be evaluated in time $\mathrm{poly}(n, 1/\delta)$ and $\bar{n}$ can be assumed to be a power of 2.*

For example, Guruswami et al. [GHSZ02] combine a Reed-Solomon code with a Hadamard code, obtaining such a list-decodable code with $\bar{n} = O(n/\delta^4)$.

We will also use the notion of weak design, as defined in [RRV02].

Definition 18. *A family of sets $S_1, \cdots, S_m \subset [d]$ is a* weak (t, r, m, d)-design *if*

1. *For all i, $|S_i| = t$.*
2. *For all i, $\sum_{j=1}^{i-1} 2^{|S_j \cap S_i|} \leq rm$.*

There exists designs with the following parameters.

Lemma 19 ([RRV02, Lemma 17])**.** *For every $t, m \in \mathbb{N}$ there exists a weak $(t, 1, m, d)$-design $S_1, \ldots, S_m \subset [d]$ such that $d = t\left\lceil \frac{t}{\ln 2} \right\rceil \lceil \log 4m \rceil = O(t^2 \log m)$. Moreover, such a design can be found in time $\mathrm{poly}(m, d)$ and space $\mathrm{poly}(m)$.*

Finally, we describe Trevisan's extractor construction.

Definition 20. *For a one-bit extractor $C : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}$, and for a weak (t, r, m, d)-design $S_1, \cdots, S_m \subset [d]$, we define the m-bit extractor $Ext_C : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ as*

$$Ext_C(x, y) := C(x, y_{S_1}), \ldots, C(x, y_{S_m}).$$

5.2 The reconstruction lemma

The following lemma is implicit in the proof of security of Trevisan's extractor construction paradigm against quantum adversaries given in [DPVR12]. A similar lemma also appeared in [VV12, Lemma 13], where the code C was specialized to the t-XOR code. For completeness, we state and sketch the proof of a more general variant of that lemma.

Lemma 21. *Let n, m, r, t, L be integers and $\varepsilon > 0$. Let $C : \{0,1\}^n \to \{0,1\}^{\bar{n}}$ be a $(\varepsilon^2/(8m^2), L)$-list-decodable code, where $\bar{n} = 2^t$. Let Ext_C be the extractor obtained by combining C with a (t, r, m, d) design as in Definition 20.*

Let ρ_{XE} be a state such that X is a random variable distributed over n-bit strings. Let U_m be uniformly distributed over m-bit strings, and suppose that

$$\left\|\rho_{Ext_C(X,Y)YE} - \rho_{U_m} \otimes \rho_Y \otimes \rho_E\right\|_1 > \varepsilon, \tag{12}$$

where Y is uniformly distributed over $\{0,1\}^d$. Then there exists fixed strings $y_1, \ldots, y_{rm} \in \{0,1\}^t$ such that, given the $\{(y_i, C(X)_{y_i})\}$ as advice, with probability at least $\varepsilon^2/(8m^2)$ over the choice of $x \sim p_X$ and her own randomness an "adversary" Eve holding system E can produce a string z such that $d_H(z, C(x)) \leq 1/2 - \varepsilon^2/(8m^2)$. In particular, Eve can recover L strings $\tilde{x}_i \in \{0,1\}^n$ such that there exits i, $\tilde{x}_i = x$.

Proof. Proposition 4.4 from [DPVR12] 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 22. *Assume* (12) *holds. Then there exists strings $y_1, \ldots, y_{rm} \in \{0,1\}^t$, and for every $y \in \{0,1\}^t$ a binary measurement, depending on the $\{(y_i, C(X)_{y_i})\}$, on E that outputs $C(X, y)$ with probability at least $1/2 + \varepsilon/m$ on average over y. Formally,*

$$\left\|\rho_{C_t(X)_Y YVE} - \rho_{U_1} \otimes \rho_Y \otimes \rho_{VE}\right\|_1 > \frac{\varepsilon}{m}, \tag{13}$$

where Y is a random variable uniformly distributed over $\{0,1\}^t$ and V is a classical register containing the $\{(y_i, C(X)_{y_i})\}$.

The next step is to argue that Eq. (13) implies that an adversary given access to $E' = VE$ can predict not only a random bit of $C(X)$, but a string Z of length m such that Z agrees with $C(X)$ in a significant fraction of positions. This follows from an argument given in [KT08], and the following claim is proved exactly as [VV12, Claim 15].

Claim 23. *Suppose* (13) *holds. Then there exists a measurement $\mathcal{F}$, with outcomes in $\{0,1\}^n$, such that*

$$\Pr_{x \sim p_X, y \sim U_t}\left(C(x)_y = C(\mathcal{F}(VE))_y\right) \geq \frac{1}{2} + \frac{\varepsilon^2}{4m^2}, \tag{14}$$

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

To conclude the argument, we use the error-correction properties of C to argue that Eve can decode her string $C(\mathcal{F}(VE))$ into an educated guess of x. Claim 23 shows that, on expectation over x, Eve's string is at Hamming distance $1/2 - \varepsilon^2/(4m^2)$ from the encoding of x. In particular, the distance will be at most $1/2 - \varepsilon^2/(8m^2)$ for a fraction at least $\varepsilon^2/(8m^2)$ of $x \sim p_X$. Since, by assumption, C is $(\varepsilon^2/(8m^2), L)$-list-decodable, for those x Eve can narrow down the possibilities to at most L distinct values. □

5.3 Proof of Lemma 9

The proof of Lemma 9 follows immediately from Lemma 21 and an appropriate choice of parameters. Let E denote the system made of the combination of $XYA_{\mathbf{B}}B_{\mathbf{B}}\mathcal{E}$, and let $n = |\mathbf{C}|$. The assumption of the lemma is that $\mathrm{H}^{\varepsilon}_{min}(B_{\mathbf{C}}|E) < \kappa n$. Let $m = \kappa n + 1$. Let $C = C_{n,\delta}$, where $\delta = \varepsilon^2/(32m^2)$, be a $(\delta, 1/\delta^2)$ list-decodable code, as promised by Lemma 17. Let Ext_C be constructed from C and a $(t, 1, m, d)$ design, where $t = \log \bar{n}$ and $d = O(t^2 \log m)$, as promised by Lemma 19.

It follows from the data processing inequality (see e.g. [KR11, Lemma V.1 (ii)]), our assumed upper bound on $\mathrm{H}^{\varepsilon}_{min}(B_{\mathbf{C}}|E)$, and our choice of m that Eq. (12) holds with $(\varepsilon/2)$ in place of ε. Thinking of Eve as simply outputting one of her L guesses $\tilde{x}_i$ chosen at random, we obtain that Eve's guess will be successful with probability at least $\varepsilon^2/(32Lm^2)$. Overall, Eve needs m bits of advice, given which she can predict x with success probability $O(\varepsilon^6/m^6)$, given our choice of parameters.

6 Additional lemmas

Lemma 24 (Azuma-Hoeffding inequality)**.** *Let (X_k) be a martingale such that $|X_k - X_{k-1}| \leq c_k$ for all k. Then for all integers m and all $t \geq 0$,*

$$\Pr\left(X_m - X_0 \geq t\right) \leq e^{-t^2/(2\sum_k c_k^2)}.$$

Lemma 25. *Let $\varepsilon, \delta, \eta, \beta > 0$ and m an integer such that $e^{-2\beta^2\delta m} < \varepsilon/2$. Let X be a random variable defined over m-bit strings. Suppose that $\Pr(\sum_i X_i \leq \eta m) \geq \varepsilon$. Then there exists a set $G \subseteq \{0,1\}^m$ such that $\Pr(G) \geq \varepsilon/2$ and for all x in G, for a fraction $\geq 1 - \delta$ of indices $i \in [m]$,*

$$\Pr(X_i = 0 | X_{<i} = x_{<i}) \geq 1 - \eta - \beta.$$

As a consequence, for a fraction at least $1 - 2\delta$ of $i \in [m]$ there exists a set $G_i \subseteq G$ such that $\Pr(G_i|G) \geq 1/2$ and for every $x_{<i} \in G_i$,

$$\Pr(X_i = 0 | X_{<i} = x_{<i}) \geq 1 - \eta - \beta.$$

Proof. For every $i \in [m]$ define

$$B_i = \left\{(x_1, \ldots, x_{i-1}, \ldots, x_m) |\ \Pr(X_i = 1 | X_{<i} = x_{<i}) \geq \eta + \beta\right\},$$

let

$$B = \left\{x \Big| \sum_{i : x \in B_i} 1 \geq \delta m\right\},$$

and suppose towards a contradiction that $\Pr(B) \geq 1 - \varepsilon/2$. Let $\hat{B} = \{x \in B | \sum_i x_i \leq \eta m\}$. By definition, for every $x \in B$ and at least a δ-fraction of indices i it holds that $\Pr(X_i = 1 | X_{<i} = x_{<i}) \geq \eta + \beta$. Hence the probability that $x \in B$ has less than η indices j at which $x_j = 1$ is at most $e^{-2\beta^2\delta m}$, i.e. $\Pr(\hat{B}|B) \leq e^{-2\beta^2\delta m}$. This shows that

$$\Pr\left(\sum_i X_i > \eta m\right) \geq \Pr(B)\left(1 - \Pr(\hat{B}|B)\right) \geq (1 - \varepsilon/2)\left(1 - e^{-2\beta^2\delta m}\right) > 1 - \varepsilon$$

given our assumption on $\varepsilon, \delta, \eta, \beta$ and m; a contradiction.

For the "consequence", for any $x \in G$ and $i \in [m]$ let $Y_{x,i} = 1$ if and only if the condition

$$\Pr(X_i = 0 | X_{<i} = x_{<i}) \geq 1 - \eta - \beta$$

is satisfied. We have shown $\mathrm{E}_{x\in G, i\in[m]}\left[Y_{x,i}\right] \geq 1-\delta$. The result is then a consequence of Markov's inequality. □

Claim 26. *Let $\rho = \rho_{\mathcal{AB}}$ be a bipartite state shared between Alice and Bob. Suppose Bob chooses $x \in \mathbf{X}^m$ according to distribution (p_x), and applies a measurement with Krauss operators $\{N_x^b\}_{b\in\mathcal{B}^m}$ on $\mathcal{B}$. Alice* sequentially *applies a measurement with Krauss operators $\{M_{x_i}^{a_i}\}_{a_i\in\mathbf{A}}$ on $\mathcal{A}$, for $i = 1,\ldots,m$. Let $D \subseteq (\mathbf{X}\times\mathbf{A}\times\mathbf{B})^m$ be a set of probability $\Pr(D) = \varepsilon$. For $i \in [m]$, let ρ_i be the state of the system $\mathcal{AB}X_i$ after $i-1$ measurements have been performed by Alice, conditioned on $(x_{<i}, a_{<i}, b_{<i}) \in D_{<i}$:*

$$\rho_i \propto \sum_{(x,a,b):(x_{<i},a_{<i},b_{<i})\in D_{<i}} p_x \left(\left(\prod_{j<i} M_{x_j}^{a_j}\right)\otimes N_x^b\right)\rho\left(\left(\prod_{j<i}\left(M_{x_j}^{a_j}\right)^\dagger\right)\otimes\left(B_x^b\right)^\dagger\right),$$

and ρ_i is normalized. Then the following bound holds:

$$\sum_i I(\mathcal{A}:X_i|D_{<i})_{\rho_i} \leq \log(1/\varepsilon).$$

Proof. We prove the lemma using standard techniques from quantum information theory; specifically the proof of the Holevo-Schumacher-Westmoreland theorem [Hol98, Sch96]. We assume that the reader is familiar with the coding and decoding strategies employed in that result, and in particular the notion of typical subspace (see e.g. [Chapters 14 and 19][Wil11], and more specifically the proof of Theorem 19.3.1). We prove the claim by describing an experiment by which Bob transmits $H(X)$ bits of information to Alice using only $H(X) + \log(1/\varepsilon) - \sum_i I(\mathcal{A}:X_i)_{\rho_i}$ bits of communication from him to Alice. This implies the claimed inequality: if it did not hold Alice could guess Bob's $H(X)$ bits with success larger than $2^{-H(X)}$ simply by running the protocol by herself, and guessing Bob's messages.

Suppose Alice and Bob share an infinite number of copies of ρ. For each $i \in [m]$, Alice and Bob also agree on a random code $\mathcal{C}_i \subseteq \mathcal{X}^K$, where K is a large integer, such that $|\mathcal{C}_i| = 2^{KI(\mathcal{A}:X_i|D_{<i})_{\rho_i}}$. By the properties of typical subspaces, with high probability over the choice of $\mathcal{C}_i$ the collection of states $\otimes_{j=1}^K \rho_i(x'_j)$ for $(x'_1,\ldots,x'_K) \in \mathcal{C}$, where $\rho_i(x'_j)$ is the reduced density of ρ_i on $\mathcal{A}$ conditioned on $X_i = x'_j$, are almost perfectly distinguishable.[8]

The experiment proceeds as follows. The copies of ρ are grouped in groups of K. For each group, Bob selects a random $x = (x_i^j)_{1\leq i\leq m, 1\leq j\leq K} \in (\mathcal{X}^m)^K$ and applies the measurements $\{N_{x^j}\}$ in the j-th copy of ρ in that group, obtaining an outcome $b^j \in \mathbf{B}^m$. For each $i \in [m]$, Alice does the following, independently for each group. She guesses whether Bob's choice of $(x_i^1,\ldots,x_i^K)$ is in $\mathcal{C}_i$ (the probability with which she guesses this should be so is equal to the probability that $x_i \in \mathcal{C}_i$, i.e. $2^{K(I(\mathcal{A}:X_i|D_{<i})_{\rho_i} - H(X_i))}$). If so, she performs the decoding measurement to recover x_i. If not, she guesses $(x_i^1,\ldots,x_i^K)$ according to $p^{\times K}$. She then applies the measurements $\{M_{x_i^j}^{a_i^j}\}$ corresponding to the guessed (x_i^j). At the end of the m repetitions, Alice sends all her guesses, and her outcomes, to Bob.

Finally, Bob finds the first group of K states in which Alice's guesses were all correct, and $(x^j, a^j, b^j) \in D$ (for each $1 \leq j \leq K$). In any group, the probability that this event happens is $2^{-K(H(X)-\sum_i I(\mathcal{A}:X_i|D_{<i})_{\rho_i})}\varepsilon^K$. Moreover, note that Alice's probability of correctly guessing Bob's choice of (x_i^j) is independent of (x_i^j). Hence Bob can indicate to Alice the index of the first group of states on which she was correct by transmitting $O(K\log(1/\varepsilon) + K(H(X) - \sum_i I(\mathcal{A}:X_i|D_{<i})_{\rho_i}))$ bits. Alice then knows all $KH(X)$ bits of information about Bob's choices of x in the m rounds on the group of K states. □

[8] Precisely, there exists a distinguishing measurement whose success probability can be made arbitrarily close to 1 by taking K large enough.

Lemma 27 (Guessing lemma)**.** *Let $\delta, \nu, \eta > 0$. Suppose given six bipartite states $\rho^{xy}_{\mathcal{AB}}$, where $x \in \{0,1,2\}$, $y \in \{0,1\}$, such that the following hold:*

1. *Letting $\rho_{\mathcal{A}} = (1/6)\sum_{xy} \mathrm{Tr}_{\mathcal{B}}(\rho^{xy}_{\mathcal{AB}})$ and $\rho_{\mathcal{B}} = (1/6)\sum_{xy} \mathrm{Tr}_{\mathcal{A}}(\rho^{xy}_{\mathcal{AB}})$,*

$$\frac{1}{6}\sum_{x,y} \left\|\rho_{\mathcal{A}} - \rho^{xy}_{\mathcal{A}}\right\|_1 \leq \nu \qquad \text{and} \qquad \frac{1}{6}\sum_{x,y} \left\|\rho_{\mathcal{B}} - \rho^{xy}_{\mathcal{B}}\right\|_1 \leq \nu, \tag{15}$$

2. *There exists observables $A_x = A^0_x - A^1_x$, $B_y = B^0_y - B^1_y$ on $\mathcal{A}, \mathcal{B}$ respectively that satisfy*

$$\frac{1}{4}\Big(\mathrm{Tr}\big((A_0 \otimes B_0)\rho^{00}_{\mathcal{AB}}\big) + \mathrm{Tr}\big((A_0 \otimes B_1)\rho^{01}_{\mathcal{AB}}\big) + \mathrm{Tr}\big((A_1 \otimes B_0)\rho^{10}_{\mathcal{AB}}\big) - \mathrm{Tr}\big((A_1 \otimes B_1)\rho^{11}_{\mathcal{AB}}\big)\Big) \geq \frac{\sqrt{2}}{2} - \eta,$$

3. *Bob's B_1 measurement produces outcome $b_1 \in \{0,1\}$ with probability $1-\delta$, when performed on his share of $\rho^{21}_{\mathcal{AB}}$:* $\mathrm{Tr}((\mathrm{Id} \otimes B^{b_1}_1)\rho^{21}_{\mathcal{AB}}) \geq 1 - \delta$.

Then the condition

$$\delta \geq \Big(\frac{\sqrt{2}-1}{2} - \eta\Big) - 75\nu$$

must hold.

Proof. For every $(a,b,x,y) \in \{0,1\}^2 \times \{0,1,2\} \times \{0,1\}$ let $p(a,b|x,y) := \mathrm{Tr}((A^a_x \otimes B^b_y)\rho^{xy}_{\mathcal{AB}})$. Condition (15) implies that the distribution p is approximately no-signalling, in the following sense: on average over the choice of a uniformly random pair (x,y), the statistical distance

$$\begin{aligned}
\frac{1}{6}\sum_{x,y}\sum_a \Big|\sum_b p(a,b|x,y) - \frac{1}{2}\sum_{y'}\Big(\sum_b p(a,b|x,y')\Big)\Big| &\leq \frac{1}{6}\sum_{x,y}\sum_a \left|\mathrm{Tr}\big((A^a_x \otimes \mathrm{Id})(\rho^{xy}_{\mathcal{AB}} - \rho^{x}_{\mathcal{AB}})\big)\right| \\
&\leq \frac{1}{6}\sum_{x,y} \left\|\rho^{xy}_{\mathcal{AB}} - \rho^{x}_{\mathcal{AB}}\right\|_1 \\
&\leq 2\nu,
\end{aligned}$$

and a similar bound holds for the marginals on $\mathcal{B}$. Lemma 9.5 in [Hol09] implies that there exists a distribution $q(a,b|x,y)$ such that q is (perfectly) no-signalling, and moreover, on average over (x,y) the statistical distance $\|p(\cdot,\cdot|x,y) - q(\cdot,\cdot|x,y)\|_1 \leq 10\nu$. In particular, the second assumption in the lemma implies that the distribution q must violate the CHSH inequality by at least $\sqrt{2}/2 - \eta - 15\nu$, and the third assumption implies that $\sum_a q(a,1|2,1) \geq 1 - \delta - 60\nu$. Applying the bound (A.11) derived in the supplementary information to [PAM+10] with $I/4 = \sqrt{2}/2 - \eta - 15\nu$ we obtain the inequality claimed in the lemma. □

References

[ABG+07] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, 2007.

[AGM06] A. Acín, N. Gisin, and L. Masanes. From Bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.*, 97:120405, 2006.

[AMP06] A. Acín, S. Massar, and S. Pironio. Efficient quantum key distribution secure against no-signalling eavesdroppers. *New Journal of Physics*, 8(8):126, 2006.

[BB84] C. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing*, pages 175–179. 1984.

[BBB+92] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin. Experimental quantum cryptography. *J. Cryptol.*, 5(1):3–28, January 1992.

[BC90] S. L. Braunstein and C. M. Caves. Wringing out better Bell inequalities. *Annals of Physics*, 202(1):22 – 56, 1990.

[BCK12a] J. Barrett, R. Colbeck, and A. Kent. Prisoners of their own device: Trojan attacks on device-independent quantum cryptography, 2012. Technical report arXiv:1201.4407.

[BCK12b] J. Barrett, R. Colbeck, and A. Kent. Unconditionally secure device-independent quantum key distribution with only two devices, 2012. Technical report arXiv:1209.0435.

[BHK05] J. Barrett, L. Hardy, and A. Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95:010503, 2005.

[BLM+05] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*, 71:022101, Feb 2005.

[CHSH69] 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.

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

[DPVR12] A. De, C. Portmann, T. Vidick, and R. Renner. Trevisan's extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41(4):915–940, 2012.

[DV10] A. De and T. Vidick. Near-optimal extractors against quantum storage. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 161–170. ACM, New York, NY, USA, 2010. ISBN 978-1-4503-0050-6.

[Eke91] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67:661–663, 1991.

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

[GHSZ02] V. Guruswami, J. Håstad, M. Sudan, and D. Zuckerman. Combinatorial bounds for list decoding. *IEEE Transactions on Information Theory*, 48(5):1021–1034, 2002.

[Hol98] A. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269 –273, 1998.

[Hol09] T. Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009.

[HR10] E. Hänggi and R. Renner. Device-independent quantum key distribution with commuting measurements, 2010. Technical report arXiv:1009.1833.

[HRW10] E. Hänggi, R. Renner, and S. Wolf. Efficient device-independent quantum key distribution. In *Proceedings of the 29th Annual international conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'10, pages 216–234. Springer-Verlag, Berlin, Heidelberg, 2010.

[KR11] R. König and R. Renner. Sampling of min-entropy relative to quantum knowledge. *IEEE Transactions on Information Theory*, 57(7):4760 –4787, 2011.

[KRS09] 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.

[KT08] 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.

[Mas09] L. Masanes. Universally composable privacy amplification from causality constraints. *Phys. Rev. Lett.*, 102:140501, 2009.

[May01] D. Mayers. Unconditional security in quantum cryptography. *J. ACM*, 48(3):351–406, May 2001.

[MHH+97] A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin. "Plug and play" systems for quantum cryptography. *Applied Physics Letters*, 70(7):793–795, 1997.

[MMMO06] F. Magniez, D. Mayers, M. Mosca, and H. Ollivier. Self-testing of quantum circuits. In *Proceedings of 33rd International Colloquium on Automata, Languages and Programming*, pages 72–83. LNCS 4051, 2006.

[MPA11] L. Masanes, S. Pironio, and A. Acín. Secure device-independent quantum key distribution with causally independent measurement devices. *Nature Communications*, 2(238):7, 2011.

[MRC+09] L. Masanes, R. Renner, M. Christandl, A. Winter, and J. Barrett. Unconditional security of key distribution from causality constraints, 2009. Technical report arXiv:quant-ph/0606049v4.

[MY98] D. Mayers and A. Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, page 503. IEEE Computer Society, Washington, DC, USA, 1998.

[MYS12] M. McKague, T. H. Yang, and V. Scarani. Robust self-testing of the singlet, 2012. Technical report arXiv:1203.2976.

[NC00] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[NPA08] M. Navascus, S. Pironio, and A. Acn. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[PAB+09] S. Pironio, A. Acn, N. Brunner, N. Gisin, S. Massar, and V. Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009.

[PAM+10] 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, 2010.

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

[Ren05] R. Renner. *Security of Quantum Key Distribution*. Ph.D. thesis, Swiss Federal Institute of Technology Zurich, September 2005.

[RRV02] R. Raz, O. Reingold, and S. Vadhan. Extracting all the randomness and reducing the error in Trevisan's extractors. *Journal of Computer and System Sciences*, 65(1):97–128, 2002.

[RUV12] B. Reichardt, F. Unger, and U. Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games, 2012. Technical report arXiv:1209.0448.

[Sch96] B. Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, 54:2614–2628, Oct 1996.

[SGB+06] V. Scarani, N. Gisin, N. Brunner, L. Masanes, S. Pino, and A. Acín. Secrecy extraction from no-signaling correlations. *Phys. Rev. A*, 74:042339, Oct 2006.

[SK09] V. Scarani and C. Kurtsiefer. The black paper of quantum cryptography: real implementation problems, 2009. Technical report arXiv:0906.4547.

[SP00] P. W. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85:441–444, Jul 2000.

[Tre01] L. Trevisan. Extractors and pseudorandom generators. *J. ACM*, 48:860–879, July 2001.

[VDTR12] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner. Chain rules for smooth min- and max-entropies, 2012. Technical report arXiv:1205.5231.

[VV12] U. Vazirani and T. Vidick. Certifiable quantum dice: or, true random number generation secure against quantum adversaries. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 61–76. ACM, 2012.

[Wil11] M. Wilde. From classical to quantum Shannon theory, 2011. Technical report arXiv:1106.1445, to be published by Cambridge University Press.