Device-independent randomness generation with sublinear shared quantum resources

Cédric Bamps, Serge Massar, and Stefano Pironio

Laboratoire d'Information Quantique, CP 224, Université libre de Bruxelles (ULB), 1050 Brussels, Belgium

June 16, 2018

In quantum cryptography, device-independent (DI) protocols can be certified secure without requiring assumptions about the inner workings of the devices used to perform the protocol. In order to display nonlocality, which is an essential feature in DI protocols, the device must consist of at least two separate components sharing entanglement. This raises a fundamental question: how much entanglement is needed to run such DI protocols? We present a two-device protocol for DI random number generation (DIRNG) which produces approximately n bits of randomness starting from n pairs of arbitrarily weakly entangled qubits. We also consider a variant of the protocol where m singlet states are diluted into n partially entangled states before performing the first protocol, and show that the number m of singlet states need only scale sublinearly with the number n of random bits produced. Operationally, this leads to a DIRNG protocol between distant laboratories that requires only a sublinear amount of quantum communication to prepare the devices.

1 Introduction

A quantum random number generation (RNG) protocol is device-independent (DI) if its output can be guaranteed to be random with respect to any adversary on the sole basis of certain minimal assumptions, such as the validity of quantum physics and the existence of secure physical locations [1]. The internal workings of the devices, however, do not need to be trusted.

Device-independence is made possible by exploiting the violation of a Bell inequality [2], which certifies the random nature of quantum measurement outcomes. As a result, DIRNG protocols necessarily consume two fundamental resources: entangled states shared across separated devices and an initial public random seed that is uncorrelated to the devices and used to determine the random measurements performed on the entangled states. Out of these two resources, a DIRNG protocol produces n private random bits.

The initial random seed that is consumed can be of extremely low quantity or quality. Indeed, n private random bits can be produced starting from an initial string of uniform bits whose required length has gradually been reduced in a series of works [3–6], culminating in the result that only a constant, i.e., independent of the output length n, amount of initial uniform random bits are required [5, 6]. Furthermore, the initial seed does not necessarily need to consist of uniform random bits, as it possible to design DIRNG protocols consuming an arbitrarily weak random seed characterized only by its total min-entropy [7].

What about entanglement, the second fundamental resource that is consumed in any DIRNG protocol? This quantum resource usually consists of m copies $|\psi\rangle^{\otimes m}$ of some bipartite entangled state $|\psi\rangle$ shared between two separated devices A and B that can be prevented at will from interacting with one another. Though DIRNG protocols involve a single user, it is useful for exposition purposes to view these two devices as being operated by two agents, Alice and Bob, in two remote sublaboratories. The m copies $|\psi\rangle^{\otimes m}$ can either be stored prior to the start of the protocol inside quantum memories in Alice's and Bob's sublaboratories, or each copy $|\psi\rangle$ can be produced individually during each execution round of the protocol, say by a source located between Alice and Bob.

All existing protocols consume at best a linear amount $m = \Omega(n)$ of such shared entangled states $|\psi\rangle$, as they operate by separately measuring (in sequence or in parallel) each of these m copies, with each separate measurement yielding at most a constant amount of random bits. Furthermore, the states $|\psi\rangle$ are typically highly entangled states—the prototypical example of a DIRNG protocol involves the measurement of n maximally entangled two-qubit states $|\phi^+\rangle$, from each of which roughly 1 bit of randomness can be certified using the CHSH inequality [3].

We will show that the consumption of entangled resources can be dramatically improved qualitatively and quantitatively. First, we show—by analogy with the fact that the initial random seed does not need to consist of uniform bits—that highly entangled states are not necessary for DIRNG: instead of using n copies of maximally entangled two-qubit pairs $|\phi^+\rangle$, n random bits can be produced from n copies of any partially entangled two-qubit pair $|\psi_\theta\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$ with $0 < \theta \leq \pi/4$ (see Theorem 3 and Corol-

lary 4).

We then turn this statement concerning the quality of the shared entangled resources into a quantitative statement about the amount of entanglement that needs to be consumed in a DIRNG protocol. The n copies of the partially entangled state $|\psi_\theta\rangle$ correspond to a total of $nS(\theta)$ ebits where $S(\theta) = h_2(\sin^2\theta)$ is the entropy of entanglement of $|\psi_\theta\rangle$ expressed in terms of the binary entropy h_2. Since $S(\theta)$ can be made arbitrarily low by considering sufficiently small values of θ, the above result seems to suggest that the total amount $nS(\theta)$ of entanglement consumed can also be made arbitrarily small as a function of n by considering sufficiently fast decreasing values for $\theta = \theta(n)$. However, if it is true that for any given θ, one can produce n random bits from n copies of $|\psi_\theta\rangle$ for any n sufficiently large, the dependency between θ and n cannot be chosen arbitrarily. This essentially originates from the fact that as $\theta \to 0$ the robustness to noise of the corresponding states $|\psi_\theta\rangle$, which become less and less entangled, decreases and must be compensated by increasing the number n of copies of the states $|\psi_\theta\rangle$ to improve the estimation phase of the protocol. There is thus a tradeoff between θ and n, which we show can nevertheless result in a total amount of entanglement $nS(\theta) = \Omega(n^k \log n)$ with $7/8 < k < 1$ (see Corollary 5). This amount of entanglement is *sublinear* in the number n of output random bits, fundamentally improving over existing protocols for which the entanglement consumption is at best linear.

Though the protocol that we introduce consumes a sublinear amount of entanglement, it still requires a linear number of shared quantum resources in the form of n copies of the two-qubit entangled states $|\psi_\theta\rangle$. These shared entangled states must be established through some quantum communication between Alice's and Bob's sublaboratories, either during the protocol itself or prior to the protocol, and will thus require the exchange of n qubits. Since this quantum communication will typically be costly (for instance because of high losses in the communication channel), it represents a measure of the use of shared quantum resources which is more operational and better motivated than the entropy of entanglement. From this perspective, however, our first protocol is not fundamentally different from existing protocols that also involve the exchange of n qubits to produce n random bits.

This leads us to consider a slight modification of our protocol in which Alice and Bob initially share m maximally entangled two-qubit states $|\phi^+\rangle$, which can be established through the exchange of m qubits. These singlets are then transformed by entanglement dilution [8] into roughly $n = S(\theta)/m$ copies of $|\psi_\theta\rangle$ states through local operations and classical communication (LOCC), which are then used in our regular protocol.

However entanglement dilution is only noiseless asymptotically, in the limit of an infinite number of copies $m \to \infty$. For finite m, entanglement dilution is inherently noisy. As our protocol is increasingly sensitive to noise as the degree of entanglement of the states θ tends to 0, it is not a priori obvious that combining randomness generation with entanglement dilution will work.

Nevertheless we show that such a two-step protocol works even though the entanglement dilution slightly degrades the tradeoff between θ and n. Specifically we exhibit a protocol that can get n output random bits starting from a sublinear number $m = nS(\theta) = \Omega(n^{k'} \log n)$ of initial copies of $|\phi^+\rangle$ states, with $7/8 < k' < 1$. This represents a quantitative improvement of the use of quantum resources with respect to all existing protocols, analogous to the fact that a DIRNG protocol needs only a sublinear amount of uniform random bits.

The starting point of our work is the work [9] wherein a family of variants of the CHSH inequality, the tilted-CHSH inequalities, are introduced, which seem particularly suited to generate randomness from weakly entangled qubit states. Indeed, it was shown in [9] that maximal violation of a tilted CHSH inequality certifies one bit of randomness and can be achieved by entangled two-dimensional systems with arbitrarily little entanglement[1] This was later extended to show that by using sequential measurements, a single pair of entangled qubits in a pure state could certify an arbitrary amount of randomness [11]. However neither of these works presented a protocol, including an estimation phase and security analysis taking into account non-maximal violation, for device independent randomness generation. In fact the results of [9, 11] do not by themselves imply the existence of such a protocol.

We now recall the tilted-CHSH expressions of [9], whose properties of randomness certification in weakly entangled states will play a central role in our protocol.

Tilted-CHSH game

The tilted-CHSH expressions I_1^β are a family of Bell expressions introduced in [9] and parameterized by a tilting parameter $\beta \in [0, 2)$. We start by reformulating I_1^β as a nonlocal game, expressed in terms of a predicate function $V \in \{0, 1\}$. This will put us in the right conditions to apply the entropy accumulation theorem of [12] following [13]. In this reformulation, Alice is given input $x \in \{0, 1\}$ and Bob input

[1] Note that not all weakly entangled states can be used for device-independent randomness generation: for instance there is a regime of visibility in which noisy singlet states (so-called Werner states) are entangled but incapable of displaying nonlocality, and hence also incapable of displaying randomness [10].

$y \in \{0,1,2\}$ according to the joint distribution

$$p(x,y) = \begin{cases} \frac{1}{4+\beta} & (x,y) \in \{0,1\}^2\,, \\ \frac{\beta}{4+\beta} & (x,y) = (0,2)\,, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Alice and Bob then provide one answer each, $(a,b) \in \{0,1\}^2$ respectively, and the game is won if the following predicate function $V(a,b,x,y) \in \{0,1\}$ returns 1:

$$V(a,b,x,y) = \begin{cases} 1 & (x,y) \in \{0,1\}^2 \text{ and } a \oplus b = xy\,, \\ 1 & (x,y) = (0,2) \text{ and } a = 0\,, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Note that in our reformulation of the tilted-CHSH expression as a game, we have introduced for convenience a third setting for Bob ($y = 2$) that is absent in the original tilted-CHSH expression. This game can be understood as a convex combination between the CHSH game and a "trivial" game: the former's success criterion is $a \oplus b = xy$ with input probabilities $p(x,y) = 1/4$ for $(x,y) \in \{0,1\}^2$, while the latter's success criterion is $a = 0$ with a deterministic input $(x,y) = (0,2)$. While Bob can tell the two games apart from his input y thanks to the introduction of the third setting $y = 2$, from Alice's point of view they are not distinguishable. This makes the mixture of the CHSH game with the trivial game nontrivial.

Given the predicate function (2), it can easily be verified that the expected winning probability ω for the tilted-CHSH game is linked to the expectation value $\bar{I}_1^\beta$ of the tilted-CHSH expression through

$$\omega = \sum_{a,b,x,y} V(a,b,x,y)p(x,y)p(a,b \mid x,y) \tag{3}$$

$$= \frac{1}{2} + \frac{1}{8+2\beta}\bar{I}_1^\beta\,, \tag{4}$$

where $p(a,b \mid x,y)$ are the probabilities characterizing Alice and Bob's outputs.

Note that when $y = 2$, Bob's output does not affect the outcome of the game, and Bob is free to provide any output. We expect that it should be possible to reformulate our results without introducing Bob's third setting, but we have found it simplest to proceed as above in order to follow closely the results of [13], where non-local games are used rather than Bell inequalities.

From the relation (4) between the tilted-CHSH game and the tilted-CHSH expression, it follows from the results of [9] that the winning probability ω goes up to $1/2 + (2+\beta)/(8+2\beta)$ for classical devices, and $1/2 + \sqrt{8+2\beta^2}/(8+2\beta) = \omega_{\mathrm{q}}$ for quantum devices. This quantum value ω_{q} is uniquely achieved (up to local transformations and up to Bob's measurement operator for $y = 2$) by a pair of devices implementing certain local measurements on a two-qubit partially entangled state $|\psi_\theta\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$ with $\tan(2\theta) = \sqrt{2/\beta^2 - 1/2}$ [9]. We call this optimal pair of devices the *reference devices* for the tilted-CHSH game of tilting parameter β. In the following we will sometimes use θ as the game parameter instead of β; it is always understood that they are linked by the above relation.

One important feature of the reference devices, as highlighted in [9], is that, for any $0 < \theta \leq \pi/4$, Alice's measurement when $x = 1$ returns a uniformly distributed outcome $a \in \{0,1\}$ uncorrelated with the environment, i.e., one bit of ideal randomness. Thus by separately measuring n copies of the partially entangled state $|\psi_\theta\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$ according to the reference measurements, one could in principle generate n bits of randomness for any $0 < \theta \leq \pi/4$.

However, the results of [9] do not immediately imply this claim because they only apply to a single use of a quantum system that is known to achieve the maximal winning probability ω_{q} of the tilted-CHSH game. Thus one should first embed the tilted-CHSH game in a proper DIRNG protocol in which no assumptions are made beforehand about the quantum systems, but where the amount of randomness generated is instead estimated from their observed behavior. This requires in particular a robust version of the results of [9], i.e., an assessment of the randomness produced by quantum devices achieving a suboptimal winning probability $\omega < \omega_{\mathrm{q}}$. Indeed, even ideal devices are not expected to achieve the quantum maximum when they are used a finite number n of times because of inherent statistical noise. We now address this by introducing an explicit DIRNG protocol based on the tilted-CHSH inequalities and a robust security analysis based on the entropy accumulation theorem (EAT) [12–14] and the self-testing properties of the tilted-CHSH inequalities introduced in [15].

2 DIRNG protocol based on the tilted-CHSH game

Our protocol consists of the following steps:

1. Select values for the following parameters:
   - The game parameter $\beta \in [0,2)$;
   - The number of measurement rounds n;
   - The expected fraction of test rounds γ;
   - A success threshold $\omega_{\mathrm{q}} - \xi$.
2. Let $i = 1$. Choose $T_i \in \{0,1\}$ independently at random such that $\Pr[T_i = 1] = \gamma$. If $T_i = 1$, perform a game round: measure the devices with settings (X_i, Y_i), selected at random according to the distribution given in (1), record the output (A_i, B_i) and compute $C_i = V(A_i, B_i, X_i, Y_i)$ according to (2). If $T_i = 0$, perform a generation round: measure the devices with $(X_i, Y_i) = (1,0)$, record the output (A_i, B_i) and let $C_i = \perp$.
3. Repeat step 2 for $i = 2, \ldots, n$.

4. Finally, if $\sum_{i:C_i=1} 1 \geq n\gamma(\omega_{\mathrm{q}} - \xi)$, the protocol succeeds. Otherwise, it aborts.

An immediate application of Hoeffding's inequality [16] produces an upper bound on the *completeness error* for this protocol, that is, the probability that the ideal devices fail the protocol:

Lemma 1. *Using the reference devices in the n rounds, the completeness error for the protocol is bounded by*

$$\epsilon_{\mathrm{c}} = \exp\left(-2n(\gamma\xi)^2\right). \tag{5}$$

Soundness of the protocol

We now establish the soundness of our protocol, that is, its ability to produce a positive amount of randomness with high probability given that the protocol did not abort. The security of this protocol rests on three standard assumptions in the DI setting: that the devices and their environment obey the laws of quantum mechanics, that the random seed used to select inputs is independent from the devices, and that the two devices are unable to communicate during each round of the protocol. Our analysis is based on the entropy accumulation theorem (EAT) [12] following closely its application to DIRNG in [13].

The EAT, as its name indicates, provides an estimate of the smooth min-entropy accumulated throughout a sequence of measurements. It implies that the smooth min-entropy of the joint measurement outcomes of our protocol scales linearly with the number of rounds, with each round providing on average an amount of min-entropy roughly equivalent to the von Neumann entropy of a single round's outcome. In order to use the EAT, it is first necessary to bound this single-round von Neumann entropy as a function of the expected probability of success ω in the tilted-CHSH game. The following Lemma, which we derive in Appendix A from the robust self-testing bounds for the tilted-CHSH inequality [15], provides a bound on the conditional min-entropy, which in turn bounds the conditional von Neumann entropy:

Lemma 2. *Let ω be the expected winning probability for the tilted-CHSH game with parameter β of a pair of quantum devices, whose internal degrees of freedom can be entangled with the environment E. Then the conditional min-entropy of the measurement outcome A for input $X = 1$ is bounded as*

$$H_{\min}(A \mid E; X = 1) \geq 1 - \kappa\theta^{-4}\sqrt{\omega_{\mathrm{q}} - \omega} \equiv g(\omega) \tag{6}$$

with $\kappa \leq 4\sqrt{4+\beta}\,(4\sqrt{2}+61)/\ln 2 \leq 385\sqrt{4+\beta}$.

The behavior of the bound with respect to $\omega_{\mathrm{q}} - \omega$ is optimal [17], while numerical results suggest that the optimal dependency in θ is $O(\theta^{-2})$ [11]. This will not, however, significantly affect our conclusions.

Using this bound in the EAT along the lines of [13] (see Appendix B) yields the following theorem:

Theorem 3. *Let $\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{Y}, \mathbf{T}, \mathbf{C}$ be the classical random variables output by the protocol, and E the quantum side information of a potential adversary. Let $\mathcal{S} = \mathcal{S}(\mathbf{C})$ be the success event for the protocol. Let ϵ', ϵ_{s} be two positive error parameters. Then, for any given pair of devices used in the protocol, either $\Pr[\mathcal{S}] \leq \epsilon'$ or*

$$H_{\min}^{\epsilon_{\mathrm{s}}}(\mathbf{AB} \mid \mathbf{XYT}E; \mathcal{S}) \geq \nu\tau n\,, \tag{7}$$

where $\nu = 1 - \gamma(2+\beta)/(4+\beta)$,

$$\tau = 1 - \kappa\theta^{-4}\sqrt{\xi + \frac{2}{\gamma\sqrt{n}}\sqrt{1 - 2\log_2(\epsilon_{\mathrm{s}}\epsilon')}} - \frac{2\log_2 26}{\sqrt{n}}\sqrt{1 - 2\log_2(\epsilon_{\mathrm{s}}\epsilon')}\,, \tag{8}$$

and $H_{\min}^{\epsilon_{\mathrm{s}}}(\mathbf{AB} \mid \mathbf{XYT}E; \mathcal{S})$ is the ϵ_{s}-smooth min-entropy of the output $(\mathbf{A}, \mathbf{B})$ given $\mathbf{X}, \mathbf{Y}, \mathbf{T}, E$ and conditioned on the event $\mathcal{S}$.

Given such a bound on the smooth min-entropy, there exist efficient procedures to extract from the raw outputs of the protocol a string of close-to-uniform random bits whose length is of the order of $H_{\min}^{\epsilon_{\mathrm{s}}}$, with the smoothing parameter ϵ_{s} characterizing the closeness to the uniform distribution.

Random bits from any partially entangled two-qubit state

Theorem 3 directly implies the following corollary, which shows the possibility of generating one bit of randomness per arbitrarily weakly entangled qubit pair:

Corollary 4. *For any constant values of the protocol parameters θ, ξ, and γ such that $\kappa\theta^{-4}\sqrt{\xi} < 1$ and for sufficiently large n, the protocol has vanishing completeness error and it generates $\Omega(n)$ bits of randomness from n partially entangled states $|\psi_\theta\rangle$. For ξ and γ approaching 0, the production of randomness in the protocol is asymptotically equal to n.* □

Sublinear entanglement consumption

We now consider how, in an ideal implementation of our protocol, the amount of shared entanglement consumed is related to the amount of randomness produced. For given n, the entanglement consumption obviously decreases with smaller values of θ. According to (7) and (8), the randomness produced, however, also decreases with smaller θ, unless this decrease is compensated by a suitable choice of the parameters γ and ξ. Indeed, for small θ, γ should be made larger to increase the fraction of game rounds and better test the devices. Similarly, ξ should be smaller (i.e., the threshold for the protocol's success should be set higher) in order for Lemma 2 to certify a nontrivial

amount of min-entropy. But the parameters γ and ξ also appear in the completeness error (15) and thus cannot be set completely freely if this error is to remain small: setting γ too low makes the estimation of the success rate at step 4 of the protocol more uncertain,[2] and setting ξ too low makes the threshold harder to reach. In the following corollary to Theorem 3, we show that there exists a choice for the parameters θ, ξ, and γ, expressed as functions of n, such that the consumption of ebits m is sublinear in the number of rounds n:

Corollary 5. *Let λ_ξ, λ_γ, λ_θ be positive scaling parameters such that*

$$\lambda_\theta < 2\lambda_\xi \,. \tag{9}$$

$$\lambda_\xi + \lambda_\gamma < 1/2 \,, \tag{10}$$

Let $\theta = n^{-\lambda_\theta/16}$, $\xi = n^{-\lambda_\xi}$, $\gamma = n^{-\lambda_\gamma}$, and constant ϵ_{s} and ϵ'. Then, for $n \to \infty$, the entropy bound of Theorem 3 is asymptotically equal to n, the completeness error vanishes, and the amount of entanglement consumed is sublinear:

$$m = nS(\theta) \sim n\theta^2 \log_2 \theta^{-2} = \frac{\lambda_\theta}{8} n^k \log_2 n \,, \tag{11}$$

with $k = 1 - \lambda_\theta/8 \in (7/8, 1)$. □

The constants λ_θ, λ_ξ, λ_γ give the rate at which the parameters θ, ξ and γ tend to zero with increasing n. The condition (9) expresses the fact that when the entanglement is small, the success threshold must be close to the maximum in order for the min-tradeoff function to take a nontrivial value (see Appendix B). The condition (10) expresses a tradeoff in the completeness error between how close the success threshold is to the maximum and the number of rounds that must be devoted to testing the correlations. A larger fraction of game rounds (i.e., a larger γ) makes the success criterion fluctuate less, which allows for a higher threshold (i.e., a smaller ξ).

3 Using diluted singlets

As mentioned in the introduction, the use of partially entangled states for randomness expansion enables us to reduce the amount of qubits exchanged between the devices when preparing their shared entanglement. We reach this goal by applying our protocol to the outcome of an *entanglement dilution* procedure, which transforms m singlet states $|\phi^+\rangle$ to $n \simeq m/S(\theta)$ partially entangled states $|\psi_\theta\rangle$. Thus, only m qubits need to be transferred between the devices in order to prepare the initial state $|\phi^+\rangle^{\otimes m}$.

[2] From the perspective of randomness generation, a small value of γ is desirable as it increases the factor ν in (7) by increasing the rate of generation rounds. However, game rounds also contribute to the final randomness, which makes the choice of $\gamma = 1$ possible.

We use the procedure of Bennett et al. [8], in which Alice prepares the n pairs locally, processes Bob's share with Schumacher compression then teleports them to Bob using the m singlets, who expands them back to n qubits. Since Schumacher compression is a lossy operation, the resulting state shared by Alice and Bob, which we denote as $\mathcal{D}_{\theta,\delta}(|\phi^+\rangle\langle\phi^+|^{\otimes m})$ is not exactly $|\psi_\theta\rangle^{\otimes n}$, but it is close in trace distance (with $\|\rho\|_1 = \mathrm{Tr}|\rho|$):

Lemma 6. *Using perfect devices, the dilution channel $\mathcal{D}_{\theta,\delta}$ maps m copies of the singlet $|\phi^+\rangle$ into n copies of the partially entangled qubit state $|\psi_\theta\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$ with $m = (S(\theta) + \delta)n$, up to error terms bounded by*

$$\left\| \mathcal{D}_{\theta,\delta}(|\phi^+\rangle\langle\phi^+|^{\otimes m}) - |\psi_\theta\rangle\langle\psi_\theta|^{\otimes n} \right\|_1 \leq 2\sqrt{\epsilon_\pi} + \epsilon_\pi \equiv \epsilon_{\mathrm{prep}} \,, \tag{12}$$

with

$$\epsilon_\pi = 2\exp(-2n\delta^2/\Delta^2) \,, \tag{13}$$

$$\Delta = -\log_2 \tan^2\theta \,. \tag{14}$$

This lemma mostly follows from [8, 18]; we prove it in Appendix C.

It follows from Lemma 6 that even a perfectly implemented dilution procedure introduces some noise in the protocol. We thus need to derive a new statement for the completeness error, the probability that perfect devices fail the protocol. Using the indistinguishability interpretation of the trace distance, if the reference state $|\psi_\theta\rangle^{\otimes n}$ passes the threshold of the protocol with probability $1 - \epsilon$, the diluted state $\mathcal{D}_{\theta,\delta}(|\phi^+\rangle\langle\phi^+|^{\otimes m})$, which is ϵ_{prep}-close to the reference, will pass the same threshold with probability at least $1 - \epsilon - \frac{1}{2}\epsilon_{\mathrm{prep}}$ [18]. Using the value of ϵ given in Lemma 1 immediately implies the following:

Lemma 7. *Starting from m perfect singlets, the composition of the dilution procedure $\mathcal{D}_{\theta,\delta}$ and the randomness generation protocol has its completeness error bounded by*

$$\epsilon_{\mathrm{c}} = \tfrac{1}{2}\epsilon_{\mathrm{prep}} + \exp\left(-2n(\gamma\xi)^2\right) . \tag{15}$$

The following analogue of Corollary 5 applies to the composition of entanglement dilution and randomness expansion; it immediately follows from the chosen parameterization:

Corollary 8. *Let $\delta = S(\theta)c$ with $c = n^{\lambda_c/8}$ for some real parameter λ_c. Let the parameters of the protocol be set as in Corollary 5, with the additional constraint that*

$$0 < \lambda_c < \lambda_\theta \,. \tag{16}$$

Starting from m singlets, the composition of entanglement dilution with parameters δ and θ with the randomness expansion protocol yields an entropy bound in Theorem 3 which is asymptotically equal to n for

$n \to \infty$, with a vanishing completeness error, and a sublinear consumption of entanglement:

$$m = n\,(S(\theta) + \delta) \sim n^{1+\lambda_c/8}\theta^2 \log_2 \theta^{-2} = \frac{\lambda_\theta}{8} n^{k'} \log_2 n\,, \quad (17)$$

with $k' = 1 - (\lambda_\theta - \lambda_c)/8 \in (7/8, 1)$. □

In addition to the tradeoffs (9) and (10), which we discussed after Corollary 5, the bound $\lambda_c > 0$ ensures that the completeness error vanishes (which requires that ϵ_π, defined in (13), also vanishes). The upper bound $\lambda_c < \lambda_\theta$ ensures that the dilution process increases the number of states, i.e., $n > m$.

4 Robustness to noise

While we have shown above that the inherent noise associated to dilution is tolerated by our protocol, we implicitly assumed that the quantum devices themselves are noise-free. Indeed the completeness error given by Eq. (5) is evaluated assuming quantum devices with an expected winning probability equal to the quantum maximum ω_q. If the quantum devices are noisy, for example due to faulty measurements, and have instead a suboptimal winning probability $\omega = \omega_q - \zeta$ with $\zeta < \xi$, then the completeness error becomes

$$\epsilon_c = \exp\bigl(-2n\gamma^2(\xi - \zeta)^2\bigr)\,. \quad (18)$$

It is easy to see that a sublinear entanglement consumption remains possible with such noisy devices, provided the noise parameter ζ decreases with n as $\zeta = n^{-\lambda_\zeta}$ for some suitable scaling parameter λ_ζ. However, realistic devices will be subject to a constant amount of noise, rather than an asymptotically vanishing one. In this case, the protocol as described so far breaks down. This is most easily seen from Lemma 2: it is clear that θ can only be taken as low as values of the order of $(\omega_q - \omega)^{1/8} = \zeta^{1/8}$ to get a non-trivial bound on the min-entropy. Nevertheless, given a small enough finite upper bound on the amount of noise ζ, partially entangled qubit pairs with an appropriate value of $\theta < \pi/4$ can still be used to produce a linear amount of randomness with a yield per ebit higher than 1 according to Theorem 3, thus improving what can directly be achieved using maximally entangled states.

A sublinear consumption of entanglement using diluted singlets can be recovered even with devices whose components fail with constant probability if our protocol is combined with error correction and fault-tolerant quantum computation. Indeed, according to the threshold theorem for fault-tolerant quantum computation, an arbitrary quantum circuit containing $G(n)$ gates may be simulated with probability of error $e(n)$ on hardware whose components fail with constant probability at most p, provided p is below some threshold, through an encoding that increases the local dimension of each qubit by a factor $\mathrm{poly}\log G(n)/e(n)$ [19]. In our case, the number $G(n)$ of gates needed to perform entanglement dilution [20] and the subsequent bipartite measurements is polynomial in n. On the other hand, aiming for a probability of error $e(n)$ that decreases polynomially in n for the simulating circuit yields a completeness error that vanishes asymptotically. The number of ebits needed in such a fault-tolerant version of our protocol is then multiplied only by a factor $\mathrm{poly}\log G(n)/e(n) = \mathrm{poly}\log n$ resulting in a total number of ebits that is still sublinear in n, i.e., $m \sim n^k \,\mathrm{poly}\log n$.

Discussion

In summary, earlier work [9] showed that a pair of entangled qubits with arbitrarily little entanglement could be used to certify one bit of randomness. Here we have carried out the further step of transforming the intuition of [9] into DIRNG protocols in which a sublinear amount of entanglement is consumed. This shows that the consumption of entanglement resources in DIRNG can be dramatically improved qualitatively and quantitatively with respect to existing protocols. These results about entanglement are analogous to those concerning the initial random seed, the other fundamental resource required in DIRNG. Interestingly, the recent work [11] suggests that one could devise DIRNG protocols that consume a constant amount of entanglement. Whether the intuition of [11] can be transformed into such a DIRNG protocol is an interesting open question.

In the present work, we did not attempt to minimize simultaneously the entanglement and the size of the initial seed. (Note that in our protocol, the size of the initial random seed is determined by the parameter γ specifying the proportion of test rounds.) Nevertheless, in the parameter regimes of Corollaries 5 and 8, the entanglement and the initial seed are both sublinear. Interestingly, it appears that our approach involves a tradeoff between entanglement and seed consumption, given the constraints placed on λ_γ and λ_θ in Corollaries 5 and 8. Indeed, equations (9) and (10) imply that $\lambda_\theta + 2\lambda_\gamma < 1$. Thus if λ_θ is close to 1 (corresponding to a small consumption of entanglement), λ_γ must be close to 0, which indicates a high proportion γ of test rounds and, as a result, high consumption of random seed. Likewise, if λ_γ is close to 1/2, λ_θ must be close to 0, and the protocol requires high entanglement and low random seed. We leave as open questions whether there is some kind of fundamental tradeoff between the required amounts of random seed and shared quantum resources, and whether the amount of quantum resources in our protocol is optimal or can be further decreased.

Acknowledgments. This work is supported by the Fondation Wiener-Anspach, the Interuniversity Attraction Poles program of the Belgian Science Policy Office under the grant IAP P7-35 photonics@be. S. P. is a Research Associate of the Fonds de la Recherche Scientifique (F.R.S.-FNRS). C. B. acknowledges funding from the F.R.S.-FNRS through a Research Fellowship.

References

[1] A. Acín and Ll. Masanes, "Certified randomness in quantum physics", Nature **540**, 213–219 (2016) DOI: 10.1038/nature20119.

[2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality", Rev. Mod. Phys. **86**, 419–478 (2014) DOI: 10.1103/RevModPhys.86.419.

[3] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, "Random numbers certified by Bell's theorem", Nature **464**, 1021 (2010) DOI: 10.1038/nature09008.

[4] U. Vazirani and T. Vidick, "Certifiable quantum dice", Phil. Trans. R. Soc. A **370**, 3432–3448 (2012) DOI: 10.1098/rsta.2011.0336.

[5] M. Coudron and H. Yuen, "Infinite randomness expansion and amplification with a constant number of devices", (2013), arXiv:1310.6755 [quant-ph].

[6] C. A. Miller and Y. Shi, "Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices", J. ACM **63**, 33:1–33:63 (2016) DOI: 10.1145/2885493.

[7] K.-M. Chung, Y. Shi, and X. Wu, "Physical randomness extractors: generating random numbers with minimal assumptions", (2014), arXiv:1402.4797 [quant-ph].

[8] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations", Phys. Rev. A **53**, 2046–2052 (1996) DOI: 10.1103/PhysRevA.53.2046.

[9] A. Acín, S. Massar, and S. Pironio, "Randomness versus nonlocality and entanglement", Phys. Rev. Lett. **108**, 100402 (2012) DOI: 10.1103/PhysRevLett.108.100402.

[10] R. F. Werner, "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model", Phys. Rev. A **40**, 4277–4281 (1989) DOI: 10.1103/PhysRevA.40.4277.

[11] F. J. Curchod, M. Johansson, R. Augusiak, M. J. Hoban, P. Wittek, and A. Acín, "Unbounded randomness certification using sequences of measurements", Phys. Rev. A **95**, 020102 (2017) DOI: 10.1103/PhysRevA.95.020102.

[12] F. Dupuis, O. Fawzi, and R. Renner, "Entropy accumulation", (2016), arXiv:1607.01796 [quant-ph].

[13] R. Arnon-Friedman, R. Renner, and T. Vidick, "Simple and tight device-independent security proofs", (2016), arXiv:1607.01797 [quant-ph].

[14] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, "Practical device-independent quantum cryptography via entropy accumulation", Nat. Commun. **9**, 459 (2018) DOI: 10.1038/s41467-017-02307-4.

[15] C. Bamps and S. Pironio, "Sum-of-squares decompositions for a family of Clauser-Horne-Shimony-Holt-like inequalities and their application to self-testing", Phys. Rev. A **91**, 052111 (2015) DOI: 10.1103/PhysRevA.91.052111.

[16] W. Hoeffding, "Probability inequalities for sums of bounded random variables", J. Am. Stat. Assoc. **58**, 13–30 (1963) DOI: 10.1080/01621459.1963.10500830.

[17] B. W. Reichardt, F. Unger, and U. Vazirani, "A classical leash for a quantum system: command of quantum systems via rigidity of CHSH games", (2012), arXiv:1209.0448 [quant-ph].

[18] M. Wilde, *Quantum information theory* (Cambridge University Press, Apr. 2013).

[19] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, Oct. 2000).

[20] R. Cleve and D. P. DiVincenzo, "Schumacher's quantum data compression as a quantum computation", Phys. Rev. A **54**, 2636–2650 (1996) DOI: 10.1103/PhysRevA.54.2636.

[21] R. König, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy", IEEE Trans. Inf. Theory **55**, 4337–4347 (2009) DOI: 10.1109/TIT.2009.2025545.

[22] M. Tomamichel, "A framework for non-asymptotic quantum information theory", PhD thesis (ETH Zurich, Mar. 2012), arXiv:1203.2142 [quant-ph].

[23] B. Schumacher, "Quantum coding", Phys. Rev. A **51**, 2738–2747 (1995) DOI: 10.1103/PhysRevA.51.2738.

[24] T. M. Cover and J. A. Thomas, *Elements of information theory*, second (John Wiley & Sons, Nov. 2012).

[25] A. Winter, "Coding theorem and strong converse for quantum channels", IEEE Trans. Inf. Theory **45**, 2481–2485 (1999) DOI: 10.1109/18.796385.

[26] T. Ogawa and H. Nagaoka, "A new proof of the channel coding theorem via hypothesis testing in quantum information theory", in 2002 IEEE international symposium on information theory, 2002. proceedings (2002), p. 73, DOI: 10.1109/ISIT.2002.1023345.

A Proof of Lemma 2

Lemma 2 gives a lower bound on the conditional min-entropy in the outcome of Alice's measurement $x = 1$ for devices which achieve a certain success probability ω in the tilted-CHSH game of parameter β. We restate it here:

Lemma 2. *Let ω be the expected winning probability for the tilted-CHSH game with parameter β of a pair of quantum devices, whose internal degrees of freedom can be entangled with the environment E. Then the conditional min-entropy of the measurement outcome A for input $X = 1$ is bounded as*

$$H_{\min}(A \mid E; X = 1) \geq 1 - \kappa\theta^{-4}\sqrt{\omega_{\mathrm{q}} - \omega} \equiv g(\omega) \tag{6}$$

with $\kappa \leq 4\sqrt{4+\beta}\,(4\sqrt{2}+61)/\ln 2 \leq 385\sqrt{4+\beta}$.

Proof. To derive this bound, we use our self-testing result for the tilted-CHSH inequalities [15]. The robustness bounds for the self-test in [15] are rather unwieldy, so we will instead provide a crude upper bound that retains the same asymptotic behavior in β and ω and greatly simplifies the use of the bound.

We will lower-bound $H_{\min}(A \mid E; X = 1)$ as a function of the expected violation of the tilted-CHSH inequality, $I = I_{\mathrm{q}} - \epsilon$, where $I_{\mathrm{q}} = \sqrt{8 + 2\beta^2} = 4/\sqrt{1 + \sin^2 2\theta}$ is the maximal quantum value of the expression. This min-entropy is equivalent to the guessing probability for the measurement $X = 1$, which is defined as

$$2^{-H_{\min}(A|E;X=1)} = p_{\mathrm{guess}}(A \mid E; X = 1) = \max_{\{M_g\}} \Pr[A = G \mid X = 1]\,, \tag{19}$$

where $\{M_g\}$ is a POVM on the subsystem E, which an adversary would use to measure the side information contained in E to formulate a guess G for A [21]. Formulated in terms of a given physical state $|\tilde{\psi}\rangle_{\mathrm{ABE}}$ and observables $\tilde{A}_x \equiv \tilde{A}_x \otimes I_{\mathrm{B}} \otimes I_{\mathrm{E}}$ and $\tilde{B}_y \equiv I_{\mathrm{A}} \otimes \tilde{B}_y \otimes I_{\mathrm{E}}$, for a given adversary POVM $\{M_g\}$ we have

$$\Pr[A = G \mid X = 1] = \sum_{a\in\{0,1\}} \langle\tilde{\psi}| \frac{I_{\mathrm{A}} + (-1)^a \tilde{A}_1}{2} \otimes I_{\mathrm{B}} \otimes M_a\, |\tilde{\psi}\rangle \tag{20}$$

$$= \tfrac{1}{2} + \tfrac{1}{2}\,\langle\tilde{\psi}|\,\tilde{A}_1 \otimes I_{\mathrm{B}} \otimes (M_0 - M_1)\,|\tilde{\psi}\rangle \equiv \tfrac{1}{2} + \tfrac{1}{2}\langle\tilde{A}_1 C\rangle\,, \tag{21}$$

letting $C = M_0 - M_1$, which is bounded as $\|C\|_\infty \leq 1$. (From here on we will sometimes use a shorter notation where instead of e.g. $I_{\mathrm{A}} \otimes I_{\mathrm{B}} \otimes C$, we simply write C.) We will now relate this expression to the reference system using self-testing.

Using the notation of [15], the self-testing result shows that any pure state $|\tilde{\psi}\rangle_{\mathrm{ABE}}$ measured by $\tilde{A}_x$ and $\tilde{B}_y$ in such a way that the tilted-CHSH inequality for these observables is violated up to $I_{\mathrm{q}} - \epsilon$ obeys[3]

$$\|\Phi(|\tilde{\psi}\rangle_{\mathrm{ABE}}) - |\psi_\theta\rangle_{\mathrm{A'B'}} \otimes |\mathrm{junk}\rangle_{\mathrm{ABE}}\| \leq 2\bar{\delta}\,, \tag{22a}$$

$$\|\Phi(\tilde{A}_1\,|\tilde{\psi}\rangle_{\mathrm{ABE}}) - A_1\,|\psi_\theta\rangle_{\mathrm{A'B'}} \otimes |\mathrm{junk}\rangle_{\mathrm{ABE}}\| \leq 2\bar{\delta} + 2\delta_{\mathrm{a}}^{\mathrm{A}}\,, \tag{22b}$$

where $\Phi = \Phi_{\mathrm{A}} \otimes \Phi_{\mathrm{B}} \otimes I_{\mathrm{E}}$ is a local isometry acting on Alice and Bob's subsystems which introduces and transforms ancillary qubits A′ and B′, $|\psi_\theta\rangle$ is the reference state (see main text), A_1 is the reference observable that yields one bit of randomness, and the error bound parameters $\bar{\delta}, \delta_{\mathrm{a}}^{\mathrm{A}} = O(\sqrt{\epsilon}\theta^{-4})$ are explicitly defined in [15]. Effectively, this isometric transformation extracts a state onto A′B′ which is close to a copy of the reference state and almost decorrelated from the initial system ABE. Likewise, the isometry approximately maps the physical observables' action on the physical state in AB to ideal actions on the reference state in the ancillary registers A′B′.

From this result, we see that the guessing probability with respect to $|\tilde{\psi}\rangle$ (which is by isometry identical to the guessing probability for $\Phi(|\tilde{\psi}\rangle)$) is close to the guessing probability with respect to the reference state, for which $p_{\mathrm{guess}} = \max_a \Pr[A = a \mid X = 1] = 1/2$ since the side information E is decorrelated from the devices A and B.

To show this approximate equality of guessing probabilities, we rewrite the last term of (21) as

$$\langle\tilde{A}_1 C\rangle = \Phi^\dagger(\langle\tilde{\psi}|)\, C\, \Phi(\tilde{A}_1\,|\tilde{\psi}\rangle) \tag{23}$$

using that Φ is an isometry which acts like the identity on E. We then use (22) and the triangle inequality to transform this into $\langle\psi_\theta|\,A_1\,|\psi_\theta\rangle\,\langle\mathrm{junk}|C|\mathrm{junk}\rangle$, with additional error terms. First, we replace $\Phi(\tilde{A}_1\,|\tilde{\psi}\rangle)$ with

[3] In [15] the subsystem E is implicitly included in A and/or B as a purifying subsystem for the mixed state held and measured by the devices. The statement can easily be modified to separate it without changing the proofs; the black-box measurement operators $\tilde{A}_x$ and $\tilde{B}_y$ then act as the identity on E.

$A_1 |\psi_\theta\rangle \otimes |\text{junk}\rangle$ with an additional error term $2\bar{\delta} + 2\delta_{\text{a}}^{\text{A}}$ since $\|\Phi^\dagger(\langle\tilde{\psi}|)\, C\| \leq 1$, then we replace $\Phi^\dagger(\langle\tilde{\psi}|)$, again using $\|A_1\|_\infty, \|C\|_\infty \leq 1$:

$$\langle \tilde{A}_1 C\rangle \leq \Phi^\dagger(\langle\tilde{\psi}|)\left(A_1 |\psi_\theta\rangle \otimes C\,|\text{junk}\rangle\right) + 2\bar{\delta} + 2\delta_{\text{a}}^{\text{A}} \tag{24}$$

$$\leq \langle\psi_\theta|\, A_1\, |\psi_\theta\rangle\, \langle\text{junk}|\, C\, |\text{junk}\rangle + 4\bar{\delta} + 2\delta_{\text{a}}^{\text{A}} \tag{25}$$

$$\leq |\langle\psi_\theta|\, A_1\, |\psi_\theta\rangle| + 4\bar{\delta} + 2\delta_{\text{a}}^{\text{A}}\,. \tag{26}$$

Since for the reference system $\langle\psi_\theta|\, A_1\, |\psi_\theta\rangle = 0$, we find that

$$p_{\text{guess}} \leq \tfrac{1}{2} + 2\bar{\delta} + \delta_{\text{a}}^{\text{A}} \tag{27}$$

We now proceed to find a simple expression of ϵ and θ that upper-bounds $2\bar{\delta} + \delta_{\text{a}}^{\text{A}}$. After some careful manipulation of the rather long expressions for $\bar{\delta}$ and $\delta_{\text{a}}^{\text{A}}$, we find

$$2\bar{\delta} + \delta_{\text{a}}^{\text{A}} = \sqrt{2I_{\text{q}}}\sqrt{\epsilon}\left[\frac{\sqrt{1+s^2}}{2s^2}\left(1 + c + \sqrt{1+s^2}\right) + \frac{\sqrt{1+s^2}}{4s}\left(2 - c + \sqrt{1+s^2}\right)\right.$$
$$\left. + \frac{c + \sqrt{1+s^2}}{2s^2}(1+c)\left(8 + 2\frac{1+\sqrt{1+s^2}}{s^2} + 3\tan\theta\right)\right], \tag{28}$$

with $c = \cos(2\theta)$, $s = \sin(2\theta)$, $I_{\text{q}} = 4/\sqrt{1+s^2}$. The dominating term in this bound for small θ comes from the term in s^{-4}, namely $2\sqrt{2I_{\text{q}}}\sqrt{\epsilon}(1+c)(c+\sqrt{1+s^2})(1+\sqrt{1+s^2})s^{-4} = O(\sqrt{\epsilon}\theta^{-4})$.

A crude upper bound on (28) is obtained by taking a s^{-4} factor out of the square brackets and giving rough numerical bounds on the bounded function that remains. For instance, the factor of $s^{-4}\sqrt{2I_{\text{q}}}\sqrt{\epsilon}$ in the first term becomes $s^2\sqrt{1+s^2}(1+c+\sqrt{1+s^2})/2 \leq (3\sqrt{2}/2)s^2$ because $c + \sqrt{1+s^2} = \sqrt{1-s^2} + \sqrt{1+s^2} \leq 2$ for $s^2 \in [0,1]$, and $\sqrt{1+s^2} \leq \sqrt{2}$. We obtain the following bound:

$$2\bar{\delta} + \delta_{\text{a}}^{\text{A}} \leq 2\sqrt{2}\left[\frac{3\sqrt{2}}{2}s^2 + \frac{1+\sqrt{2}}{2}s^3 + 16s^2 + 8 + 6s^2\tan\theta\right]\frac{\sqrt{\epsilon}}{s^4}\,, \tag{29}$$

where we have also bounded $I_{\text{q}} \leq 4$ and used the following tight bounds after expanding the third term in the square brackets of (28):

$$(c + \sqrt{1+s^2})(1+c) \leq 4\,, \tag{30}$$

$$(c + \sqrt{1+s^2})(1+c)(1+\sqrt{1+s^2}) = (1 + c\sqrt{1+s^2})(2 + c + \sqrt{1+s^2}) \leq 8\,. \tag{31}$$

The factor $\tan\theta$ in the last term is simply bounded by 1. The bound we reach is the following:

$$2\bar{\delta} + \delta_{\text{a}}^{\text{A}} \leq 2\sqrt{2}\left[\frac{1+\sqrt{2}}{2}s^3 + \frac{3\sqrt{2}+44}{2}s^2 + 8\right]\frac{\sqrt{\epsilon}}{s^4}\,. \tag{32}$$

Finally, the polynomial in s in square brackets is bounded by its maximum at $s = 1$. Eq. (27) then becomes

$$p_{\text{guess}} - \tfrac{1}{2} \leq 2\bar{\delta} + \delta_{\text{a}}^{\text{A}} \leq (8 + 61\sqrt{2})\frac{\sqrt{\epsilon}}{s^4}\,. \tag{33}$$

Further bounding

$$s = \sin 2\theta \geq \frac{2\theta}{\pi/2} \geq \theta \tag{34}$$

by concavity of the sine function on $[0, \pi/2]$ and substituting $\epsilon = (8+2\beta)(\omega_{\text{q}} - \omega)$, we reach our final bound for the guessing probability,

$$p_{\text{guess}} \leq \tfrac{1}{2} + \sqrt{8+2\beta}\,(8 + 61\sqrt{2})\,\theta^{-4}\sqrt{\omega_{\text{q}} - \omega}\,. \tag{35}$$

Putting this together with (19), we find, using $\ln(1+x) \leq x$,

$$H_{\min}(A \mid E; X=1) \geq 1 - \log_2\left(1 + 2\sqrt{8+2\beta}\,(8+61\sqrt{2})\,\theta^{-4}\sqrt{\omega_{\text{q}} - \omega}\right) \tag{36}$$

$$\geq 1 - \kappa\theta^{-4}\sqrt{\omega_{\text{q}} - \omega} \tag{37}$$

with $\kappa = 4\sqrt{4+\beta}(4\sqrt{2} + 61)/\ln 2$. □

A numerical maximization of the factor of $\sqrt{\epsilon}$ in (28) shows that a tighter numerical factor of 45.13 could replace the numerical factor $8 + 61\sqrt{2} = 94.27$ in (33), or less if the range of θ is limited.

B Proof of Theorem 3

It is shown in [12, 13] that obtaining a bound on the smooth min-entropy produced by a generic protocol of the type that we consider here reduces to finding a *min-tradeoff function*, a certain function that bounds the randomness produced in an average round of the protocol. This function is specific to the particular game used in the protocol and obtaining it for the tilted-CHSH game is the only part of the general analysis of [13] that we need to tailor to our situation.

The min-tradeoff function is defined as follows. Any protocol round (i.e., step 2 in the protocol; see Section 2) can be thought of as a quantum channel $\mathcal{G}_i$ mapping the state $\rho = \rho_{DE}$ of the pair of devices D and the adversary information E before that round to the resulting state $\mathcal{G}_i(\rho) = \rho' = \rho'_{A_iB_iX_iY_iT_iC_iD'E}$ after the protocol round, which also includes explicitly the classical data that was produced in that round. In particular, the channel $\mathcal{G}_i$ and the initial state ρ determine the probability distribution $p \equiv (p_0, p_1, p_\perp)$ for the classical random variable $C_i \in \{0, 1, \perp\}$. This probability distribution is related to the randomness produced in the protocol round: from Lemma 2 we expect that a pair of devices which succeeds at any round ($C_i = 1$) with higher probability produces more entropy in its outputs. The min-tradeoff function is a function $f_{\min}(p)$ that bounds the randomness produced in the protocol round solely on the basis of the probability distribution p. Formally, a function $f_{\min}(p)$ is a min-tradeoff function if it satisfies

$$f_{\min}(p) \leq H(A_iB_i \mid X_iY_iT_iE)_{\mathcal{G}_i(\rho)}\,, \tag{38}$$

where $H(A_iB_i \mid X_iY_iT_iE)_{\mathcal{G}_i(\rho)}$ is the von Neumann entropy of the joint outputs conditioned on the classical side information produced in the round and on the quantum information of the adversary E. This inequality should hold for all channels $\mathcal{G}_i$ that are compatible with the protocol and for all initial states ρ such that the variable C_i in $\mathcal{G}_i(\rho)$ is distributed as p.

Theorem 3, which we restate here, follows from the entropy accumulation theorem [12] and its application to randomness generation protocols by Arnon-Friedman et al. [13]. Our proof follows that of [13], in which we substitute a min-tradeoff function adapted to our protocol.

Theorem 3. *Let $\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{Y}, \mathbf{T}, \mathbf{C}$ be the classical random variables output by the protocol, and E the quantum side information of a potential adversary. Let $\mathcal{S} = \mathcal{S}(\mathbf{C})$ be the success event for the protocol. Let ϵ', ϵ_{s} be two positive error parameters. Then, for any given pair of devices used in the protocol, either $\Pr[\mathcal{S}] \leq \epsilon'$ or*

$$H_{\min}^{\epsilon_{\mathrm{s}}}(\mathbf{AB} \mid \mathbf{XYT}E; \mathcal{S}) \geq \nu\tau n\,, \tag{7}$$

where $\nu = 1 - \gamma(2+\beta)/(4+\beta)$,

$$\tau = 1 - \kappa\theta^{-4}\sqrt{\xi + \frac{2}{\gamma\sqrt{n}}\sqrt{1 - 2\log_2(\epsilon_{\mathrm{s}}\epsilon')}} - \frac{2\log_2 26}{\sqrt{n}}\sqrt{1 - 2\log_2(\epsilon_{\mathrm{s}}\epsilon')}\,, \tag{8}$$

and $H_{\min}^{\epsilon_{\mathrm{s}}}(\mathbf{AB} \mid \mathbf{XYT}E; \mathcal{S})$ is the ϵ_{s}-smooth min-entropy of the output $(\mathbf{A}, \mathbf{B})$ given $\mathbf{X}, \mathbf{Y}, \mathbf{T}, E$ and conditioned on the event $\mathcal{S}$.

Proof of Theorem 3. We first note that $C_i = \perp$ happens if and only if the protocol round is a generation round, hence we always have $p_\perp = 1 - \gamma$ and $p_0 + p_1 = \gamma$. Thus, as noted in [13], we are free to define $f_{\min}(p)$ to arbitrary values when $p_0 + p_1 \neq \gamma$, since such a distribution for C_i is not compatible with our protocol anyway. On the other hand, when $p_0 + p_1 = \gamma$, the expected probability of succeeding at the nonlocal game in a game round is p_1/γ. In that case, we can use Lemma 2 with $\omega = p_1/\gamma$ to set the value of $f_{\min}$. Indeed,

$$\begin{aligned}
H(A_iB_i \mid X_iY_iT_iE) &\geq H(A_i \mid X_iY_iT_iE) && (39)\\
&= H(A_i \mid X_iE) && (40)\\
&\geq \Pr[X_i = 1]\, H(A_i \mid E; X_i = 1) && (41)\\
&\geq \nu\, g(\omega)\,, && (42)
\end{aligned}$$

with $\nu = \Pr[X_i = 1] = 1 - \gamma(2+\beta)/(4+\beta)$. In (39), we used the chain rule and the positivity of the conditional entropy of classical information. In (40), we used that Alice's output A_i is independent of Bob's measurement choice Y_i and of the round flag T_i. To get (42), we used Lemma 2 and the fact that the conditional min-entropy lower-bounds the conditional von Neumann entropy [22, Proposition 4.3].

We can thus define $f_{\min}(p) = \nu\, g(p_1/\gamma)$ when $p_0 + p_1 = \gamma$. For convenience, we set it to the same value when $p_0 + p_1 \neq \gamma$, since it can be freely chosen in that case. All in all,

$$f_{\min}(p) = \nu\, g(p_1/\gamma)\,. \tag{43}$$

The EAT [12] requires *affine* min-tradeoff functions. Since g is convex, we can simply obtain affine lower bounds of (43) by taking its tangent at any point. The tangent of $g(\omega)$ at the point $\omega = \omega_{\mathrm{t}}$ is given by

$$\bar{g}_{\omega_{\mathrm{t}}}(\omega) = 1 - \kappa\theta^{-4}\frac{2\omega_{\mathrm{q}} - \omega_{\mathrm{t}} - \omega}{2\sqrt{\omega_{\mathrm{q}} - \omega_{\mathrm{t}}}}\,, \tag{44}$$

hence the min-tradeoff function we finally use will be $f_{\min}(p) = \nu\,\bar{g}_{\omega_{\mathrm{t}}}(p_1/\gamma)$ for some appropriately chosen ω_{t}.

Given such a min-tradeoff function, Lemma 9 of [13] then states that for any given pair of devices, either the protocol succeeds with low probability $\Pr[\mathcal{S}] \leq \epsilon'$, or

$$H_{\min}^{\epsilon_{\mathrm{s}}}(\mathbf{AB} \mid \mathbf{XYT}E; \mathcal{S}) \geq n\nu\,\bar{g}_{\omega_{\mathrm{t}}}(\omega_{\mathrm{q}} - \xi) - \mu\sqrt{n} \tag{45}$$

with

$$\mu = 2(\log_2(13) + \lceil \|\nabla f_{\min}\|_\infty \rceil)\sqrt{1 - 2\log_2(\epsilon_{\mathrm{s}}\epsilon')}\,. \tag{46}$$

The gradient of the min-tradeoff function is simply the slope of $\nu\,\bar{g}_{\omega_{\mathrm{t}}}(p_1/\gamma)$:

$$\|\nabla f_{\min}\|_\infty = \gamma^{-1}\nu\frac{\kappa\theta^{-4}}{2\sqrt{\omega_{\mathrm{q}} - \omega_{\mathrm{t}}}}\,. \tag{47}$$

Bounding the ceiling function as $\lceil x \rceil \leq x + 1$ and optimizing over the point of tangency ω_{t} produces the final expression (7) for the min-entropy bound of the theorem. □

C Proof of Lemma 6

In Section 3, we sketched the entanglement dilution procedure of Bennett et al. [8], which defines a channel $\mathcal{D}_{\theta,\delta}$ that approximately dilutes $|\phi^+\rangle^{\otimes m}$ into $|\psi_\theta\rangle^{\otimes n}$, with $m < n \simeq m/S(\theta)$.

In this appendix, we describe this procedure in detail then prove Lemma 6, which bounds its inherent error terms. We restate the Lemma here:

Lemma 6. *Using perfect devices, the dilution channel $\mathcal{D}_{\theta,\delta}$ maps m copies of the singlet $|\phi^+\rangle$ into n copies of the partially entangled qubit state $|\psi_\theta\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$ with $m = (S(\theta) + \delta)n$, up to error terms bounded by*

$$\left\| \mathcal{D}_{\theta,\delta}(|\phi^+\rangle\langle\phi^+|^{\otimes m}) - |\psi_\theta\rangle\langle\psi_\theta|^{\otimes n} \right\|_1 \leq 2\sqrt{\epsilon_\pi} + \epsilon_\pi \equiv \epsilon_{\mathrm{prep}}\,, \tag{12}$$

with

$$\epsilon_\pi = 2\exp(-2n\delta^2/\Delta^2)\,, \tag{13}$$

$$\Delta = -\log_2\tan^2\theta\,. \tag{14}$$

As stated in the main text, dilution is enabled by the possibility of compressing a number of weakly entangled states into a smaller Hilbert space at the cost of a small error. One procedure that realizes this is known as Schumacher compression [23] and is explained in great detail in [18], from which we borrow the notation.

We will apply Schumacher compression to the second half of the global state of n weakly entangled states $|\psi_\theta\rangle_{\mathrm{AA}'}^{\otimes n} = (\cos\theta\,|00\rangle + \sin\theta\,|11\rangle)^{\otimes n}$. This allows us to use $m < n$ maximally entangled qubit pairs to transport that second half from box A to box B so that the initial m singlets are effectively transformed into n weakly entangled pairs after decompression.

Schumacher compression works on a source of pure states $\{|\psi_i\rangle, q_i\}$ which outputs the states in $\{|\psi_i\rangle\}$ at random, with respective probabilities $\{q_i\}$. The goal is to pack the information output by n i.i.d. uses of the source into a smaller Hilbert space in a way that makes it recoverable later with high fidelity—a generalization of Shannon's source coding to the quantum setting. We interpret the reduced density operator $\sigma_{\mathrm{A}'} = \mathrm{Tr}_{\mathrm{A}}\,|\psi_\theta\rangle\langle\psi_\theta|_{\mathrm{AA}'}$ for the second half of one pair $|\psi_\theta\rangle_{\mathrm{AA}'}$ as describing a quantum source of mutually orthogonal states—namely, the eigenstates of σ—with probabilities given by the corresponding eigenvalues. The eigenstates of σ coincide with the computational basis: they are $|0\rangle$ and $|1\rangle$, with corresponding eigenvalues $q_0 = \cos^2\theta$ and $q_1 = \sin^2\theta = 1 - q_0$. Using this source n times gives us the mixed state $\rho_{\mathrm{A}'} = \sigma_{\mathrm{A}'}^{\otimes n} = \mathrm{Tr}_{\mathrm{A}}\,|\psi_\theta\rangle\langle\psi_\theta|^{\otimes n}$, i.e., the reduced state of Bob's share of the set of n entangled pairs prepared by Alice. The eigenstates of $\rho_{\mathrm{A}'}$ are the computational basis states for n qubits, which we write as $|y\rangle$ with $y \in \{0,1\}^n$, where $|y\rangle$ is the tensor product of n qubits $|y_1\rangle\cdots|y_n\rangle$. The eigenvalue associated with y is $\lambda_j = (\cos^2\theta)^j(\sin^2\theta)^{n-j}$ for $j = n(0 \mid y)$, which gives the number of zeros in the binary string y.

Compressing a source of orthogonal states is more or less equivalent to Shannon source coding. The idea is to consider the string y obtained after n uses of the source and only let it through when it is deemed "typical"

enough. According to the theory of typical sequences [24], there is a family of subsets of the 2^n strings y, called the δ-typical subsets, that each contain an exponentially small fraction of strings but nevertheless have an exponentially large probability weight. Thus, while most strings we obtain are typical, their number is considerably smaller than 2^n. The δ-typical subset is defined as

$$\mathcal{T}_\delta = \left\{ y \in \{0,1\}^n : S - \delta \leq \frac{-\log_2 P(y)}{n} \leq S + \delta \right\} , \tag{48}$$

where S is the entropy of the source (which is also the entropy of entanglement of a single pair $|\psi_\theta\rangle$),

$$S = h_2(q_0) = -q_0 \log_2(q_0) - q_1 \log_2(q_1) \tag{49}$$
$$= -\log_2(q_0) + q_1 \Delta , \tag{50}$$

with $\Delta = \log_2(q_0/q_1) = -\log_2 \tan^2\theta$. Thus, a sequence y is δ-typical if and only if its sample entropy $-(1/n)\log_2 P(y)$ is δ-close to S. Since our ideal source is i.i.d., each random variable Y_i is distributed independently according to the same Bernoulli distribution of probabilities $\{q_0, q_1\}$. Thus, the sample entropy for a given value y can be rewritten as

$$-\frac{1}{n}\log_2 P(y) = \frac{1}{n}\sum_{i=1}^{n} -\log_2 q_{y_i} \tag{51}$$
$$= -\frac{n(0 \mid y)}{n}\log_2 q_0 - \frac{n(1 \mid y)}{n}\log_2 q_1 \tag{52}$$
$$= -\log_2 q_0 + \frac{n(1 \mid y)}{n}\Delta . \tag{53}$$

Hence, the definition of the typical set (48) can be rewritten as

$$\mathcal{T}_\delta = \left\{ y \in \{0,1\}^n : q_1 - \frac{\delta}{\Delta} \leq \frac{1}{n}\sum_{i=1}^{n} y_i \leq q_1 + \frac{\delta}{\Delta} \right\} . \tag{54}$$

That is, a sequence y is δ-typical if and only if it has a frequency of 1's that is (δ/Δ)-close to the expected value q_1.

Properties of the typical set are easily derived from those two expressions of $\mathcal{T}_\delta$ [24]. First, applying Hoeffding's inequality [16] for the binomially-distributed $\sum_i Y_i$ shows that the typical set has a high probability weight

$$\Pr[\mathcal{T}_\delta] \geq 1 - \epsilon_\pi , \tag{55}$$

where we defined the *projection error*

$$\epsilon_\pi = 2\exp(-2n\delta^2/\Delta^2) , \tag{56}$$

which is an upper bound on the probability of a sequence being atypical.

Secondly, in contrast to this first property, the typical set has a relatively low cardinality: the number of typical sequences is exponentially small compared to the total number of sequences. Indeed, from the definition (48),

$$|\mathcal{T}_\delta| = 2^{n(S+\delta)} \sum_{y\in\mathcal{T}_\delta} 2^{-n(S+\delta)} \tag{57}$$
$$\leq 2^{n(S+\delta)} \sum_{y\in\mathcal{T}_\delta} P(y) \leq 2^{n(S+\delta)} , \tag{58}$$

which is much smaller than the total number of 2^n sequences if $S < 1 - \delta$ and n is high.

Source coding consists in discarding atypical sequences, which occur with low probability, and encoding typical sequences into smaller codewords. This encoding is possible because of the small cardinality of the typical set: a sequence can simply be encoded by its index within a given ordering of the elements of the typical set, which gives binary codewords a length of at most $\log_2|\mathcal{T}_\delta| \leq n(S+\delta) \equiv m$.

Schumacher compression applies this procedure to the quantum state $\rho_{A'} = \sum_{y\in\{0,1\}^n} \lambda_{n(0|y)}\, |y\rangle\langle y|$, which describes the output of a quantum source of pure states $|y\rangle$ with probabilities $\lambda_{n(0|y)}$. In order to identify an atypical state, a projective *typicality measurement* is performed, with projectors $\{\Pi_\delta, I - \Pi_\delta\}$ where

$$\Pi_\delta = \sum_{y\in\mathcal{T}_\delta} |y\rangle\langle y| . \tag{59}$$

If the typicality measurement succeeds, the state ends up in the typical subspace spanned by $\{|y\rangle : y \in \mathcal{T}_\delta\}$, and can be encoded in a 2^m-dimensional Hilbert space by an invertible isometric map V. If instead the measurement fails, a given typical state τ is substituted and encoded the same way. The resulting state is therefore $\mathcal{C}(\rho_{\mathrm{A}'}) = V[\Pi_\delta \rho_{\mathrm{A}'} \Pi_\delta + (1 - \mathrm{Tr}(\Pi_\delta \rho_{\mathrm{A}'}))\,\tau]V^\dagger$.

The two properties (55) and (58) of the typical set can be expressed in terms of the typical projector Π_δ:

$$\mathrm{Tr}(\Pi_\delta \rho_{\mathrm{A}'}) \geq 1 - \epsilon_\pi\,, \tag{60}$$

$$\mathrm{Tr}(\Pi_\delta) \leq 2^{n(S+\delta)}\,. \tag{61}$$

The first property can be used in the gentle operator lemma [25, 26] to show that a successful typicality measurement does not disturb the state by much [18]:

$$\|\Pi_\delta \rho_{\mathrm{A}'} \Pi_\delta - \rho_{\mathrm{A}'}\|_1 \leq 2\sqrt{\epsilon_\pi}\,. \tag{62}$$

Hence, the decompressed state is close in trace distance to the original [18]:

$$\left\|V^\dagger \mathcal{C}(\rho_{\mathrm{A}'}) V - \rho_{\mathrm{A}'}\right\|_1 \leq 2\sqrt{\epsilon_\pi} + \epsilon_\pi\,. \tag{63}$$

As Schumacher originally noted [23], this remains true when we consider the global state in AA′; the entanglement of the state is therefore not destroyed by compression:

$$\left\|(I \otimes V^\dagger)(\mathrm{Id} \otimes \mathcal{C})[|\psi_\theta\rangle\langle\psi_\theta|^{\otimes n}](I \otimes V) - |\psi_\theta\rangle\langle\psi_\theta|^{\otimes n}\right\|_1 \leq 2\sqrt{\epsilon_\pi} + \epsilon_\pi\,. \tag{64}$$

Bennett et al.'s dilution procedure simply results from the composition of this compression with quantum teleportation. The proof of Lemma 6 is therefore immediate:

Proof of Lemma 6. Defining $\mathcal{D}_{\delta,\theta}(|\phi^+\rangle\langle\phi^+|^{\otimes m}_{\mathrm{A''B}})$ to be the outcome of the composition of a local preparation of $|\psi_\theta\rangle\langle\psi_\theta|^{\otimes n}_{\mathrm{AA}'}$, followed by Schumacher compression over δ-typical sequences of A′, teleportation from A′ to B using $|\phi^+\rangle\langle\phi^+|^{\otimes m}_{\mathrm{AB}}$ and decompression on B, Lemma 6 follows. □