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

(Dated: August 29, 2017)

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.

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. 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 $|\psi\rangle^m$ 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 $|\psi\rangle^m$ 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 $|\psi_0\rangle = \cos \theta |00\rangle + \sin \theta |11\rangle$ with $0 < \theta \leq \pi/4$ (see Theorem 3 and Corollary 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 $n$ partially entangled states $|\psi_0\rangle$ correspond to a total of $nS(\theta)$ ebits where $S(\theta) = h_2(\sin^2 \theta)$ is the entropy of entanglement of $|\psi_0\rangle$ expressed in terms of the binary entropy $h_2$. Since $S(\theta)$ can be made arbitrarily low by considering sufficiently small values of $\theta$, 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 $\theta$, one can produce $n$ random bits from $n$ copies of $|\psi_0\rangle$ for any $n$ sufficiently large, the dependency between $\theta$ 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 $|\psi_0\rangle$, which become less and less entangled, decreases and must be compensated by increasing the number $n$ of copies of the states $|\psi_0\rangle$ to improve the estimation phase of the protocol. There is thus a tradeoff between $\theta$ and $n$, which we show can nevertheless result in a total amount of entanglement $nS(\theta) = \Omega(n^2 \log n)$ 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.
We start by recalling these results. It is easily seen that the expected winning probability respectively, and the game is won if the following predicate distribution is satisfied by a tilting parameter family of Bell expressions introduced in [9] and parameterized to suitably certify randomness from weakly entangled qubit states. CHSH inequalities, are introduced, which seem particularly suited to certify randomness from weakly entangled qubit states. Of the use of quantum resources with respect to all existing slightly degrades the tradeoff between \( \theta \) and Bob’s outputs. The winning probability \( \omega \) goes up to \( 1/2 + (2 + \beta)/(8 + 2\beta) \) for classical devices, and \( 1/2 + \sqrt{8 + 2\beta^2}/(8 + 2\beta) = \omega_q \) for quantum devices [9]. This quantum value \( \omega_q \) is uniquely achieved (up to local transformations) 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 \( \beta \). In the following we will sometimes use \( \theta \) as the game parameter instead of \( \beta \); 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 \( \omega_q \) 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 \( \omega < \omega_q \). 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) [10, 11] and the self-testing properties of the tilted-CHSH inequalities introduced in [12].

**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 \( \gamma \);
   - A success threshold \( \omega_q - \varepsilon \).
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 \( C_i = \pm \).
3. Repeat step 2 for \( i = 2, \ldots, n \).
4. Finally, if \( \sum_{i} C_{i} \leq 1 \geq n\gamma(\omega q - \xi) \), the protocol succeeds. Otherwise, it aborts.

An immediate application of Hoeffding’s inequality [13] produces an upper bound on the \textit{completeness error} for this protocol, that is, the probability that the ideal devices fail the protocol:

**Theorem 3.** Using the reference devices in the \( n \) rounds, the completeness error for the protocol is bounded by

\[
\epsilon_{c} = \exp\left(-2n(\gamma\xi)^2\right) .
\]

\[
(4)
\]

**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) [10] following closely its application to DIRNG in [11].

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 \( \omega \) 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 [12], provides a bound on the conditional min-entropy, which in turn bounds the conditional von Neumann entropy:

**Lemma 2.** Let \( \omega \) be the expected winning probability for the tilted-CHSH game with parameter \( \beta \) 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

\[
H_{\text{min}}^{c}(A \mid E; X = 1) \geq 1 - \kappa\theta^{-4}\sqrt{\omega_{q} - \omega}
\]

(5)

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

The behavior of the bound with respect to \( \omega_{q} - \omega \) is optimal [14], while numerical results suggest that the optimal dependency in \( \theta \) is \( O(\theta^{-2}) \) [15]. This will not, however, significantly affect our conclusions.

Using this bound in the EAT along the lines of [11] (see Appendix B) yields the following theorem:

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

\[
H_{\text{min}}^{c}(AB \mid XYTE; S) \geq vtn ,
\]

where \( v = 1 - \gamma(2 + \beta)/(4 + \beta) \).

\[
(6)
\]

and \( H_{\text{min}}^{c}(AB \mid XYTE; S) \) is the \( \epsilon_{c} \)-smooth min-entropy of the output \((A, B)\) given \( X, Y, T, E \) and conditioned on the event \( S \).

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 \( H_{\text{min}}^{c} \), with the smoothing parameter \( \epsilon_{c} \) 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 \( \theta, \xi, \gamma \) 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 \( |\psi_{0}\rangle \). For \( \xi \) and \( \gamma \) approaching \( 0 \), the production of randomness in the protocol is asymptotically equal to \( n \). \( \square \)

**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 \( \theta \). According to (6) and (7), the randomness produced, however, also decreases with smaller \( \theta \), unless this decrease is compensated by a suitable choice of the parameters \( \xi \) and \( \gamma \). But these latter parameters also appear in the completeness error (11) and thus cannot be set completely freely if this error is to remain small. In the following corollary to Theorem 3, we show that there exists a choice for the parameters \( \theta, \xi, \gamma \), expressed as functions of \( n \), such that the consumption of ebits \( m \) is sublinear in the number of rounds \( n \):

**Corollary 5.** Let \( \lambda_{\xi}, \lambda_{\gamma}, \lambda_{\theta} \) be positive scaling parameters such that \( \lambda_{\xi} + \lambda_{\gamma} < 1/2 \) and \( \lambda_{\theta} < 2\lambda_{\xi} \). Let \( \theta = n^{-\lambda_{\theta}/16}, \xi = n^{-\lambda_{\xi}}, \gamma = n^{-\lambda_{\gamma}}, \) and constant \( \epsilon \) and \( \epsilon' \). 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:

\[
m = nS(\theta) - n\theta^{2}\log_{2}\theta^{-2} = \frac{\lambda_{\theta}}{8} n^{k} \log_{2} n ,
\]

with \( k = 1 - \lambda_{\theta}/8 \in (7/8, 1) \). \( \square \)
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 \( |\psi^+\rangle \) to \( m = nS(\theta) \) partially entangled states \( |\phi_\theta\rangle \). Thus, only \( m \) qubits need to be transferred between the devices in order to prepare the initial state \( |\phi^+\rangle^{\otimes m} \).

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 \( D_{0,\delta}(|\phi^+\rangle^{\otimes m}) \), is not exactly \( |\phi_\theta\rangle^{\otimes n} \), but it is close in trace distance (with \( \|\rho\|_1 = \operatorname{Tr}(|\rho|) \)):

**Lemma 6.** Using perfect devices,

\[
\left\| D_{0,\delta}(|\phi^+\rangle^{\otimes m}) - |\phi_\theta\rangle^{\otimes m} \right\| \leq 2\sqrt{\epsilon_\delta} + \epsilon_\sigma \equiv \epsilon_{\text{prep}},
\]

with

\[
\epsilon_\sigma = 2 \exp(-2n\delta^2/\Delta^2) , \quad \Delta = -\log_2 \tan^2 \theta .
\]

This lemma mostly follows from [8, 16]; we prove it in Appendix C.

It follows from Lemma 6 that even a perfectly implemented protocol has its completeness error bounded by

\[
\epsilon_{\text{c}} = \exp\left(-2n\gamma^2(\xi - \zeta)^2\right).
\]

**Noise robustness.** 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. (4) is evaluated assuming quantum devices with an expected winning probability equal to the quantum maximum \( \omega_q \). If the quantum devices are noisy, for example due to faulty measurements, and have instead a suboptimal winning probability \( \omega = \omega_q - \zeta \), then the completeness error becomes

\[
\epsilon_{\text{c}} = \exp\left(-2n\gamma^2(\xi - \zeta)^2\right).
\]
are both sublinear. Interestingly, it appears that our approach involves a tradeoff between entanglement and seed consumption, given the constraints placed on $\lambda_\gamma$ and $\lambda_\theta$ in Corollaries 5 and 8. In Corollary 5, if $\lambda_\theta$ is close to 1 (corresponding to a small consumption of entanglement), $\lambda_\gamma$ must be close to 0, which indicates a high proportion $\gamma$ of test rounds and, as a result, high consumption of random seed. Likewise, if $\lambda_\gamma$ is close to $1/2$, $\lambda_\theta$ 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 trade-off 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.

[1] A. Acín and Ll. Masanes, “Certified randomness in quantum physics,” Nature 540, 213–219 (2016).
[2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, “Bell nonlocality,” Rev. Mod. Phys. 86, 419–478 (2014).
[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).
[4] U. Vazirani and T. Vidick, “Certifiable quantum dice,” Phil. Trans. R. Soc. A 370, 3432–3448 (2012).
[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).
[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).
[9] A. Acín, S. Massar, and S. Pironio, “Randomness versus nonlocality and entanglement,” Phys. Rev. Lett. 108, 100402 (2012).
[10] F. Dupuis, O. Fawzi, and R. Renner, “Entropy accumulation,” (2016), arXiv:1607.01796 [quant-ph].
[11] R. Arnon-Friedman, R. Renner, and T. Vidick, “Simple and tight device-independent security proofs,” (2016), arXiv:1607.01797 [quant-ph].
[12] 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).
[13] W. Hoeffding, “Probability inequalities for sums of bounded random variables,” J. Am. Stat. Assoc. 58, 13–30 (1963).
[14] 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].
[15] F. J. Curcho, M. Johansson, R. Augusiak, M. J. Hoban, P. Wittek, and A. Acín, “Unbounded randomness certification using sequences of measurements,” (2015), arXiv:1510.03394 [quant-ph].
[16] M. Wilde, Quantum Information Theory (Cambridge University Press, 2013).
[17] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, 2000).
[18] R. Cleve and D. P. DiVincenzo, “Schumacher’s quantum data compression as a quantum computation,” Phys. Rev. A 54, 2636–2650 (1996).
[19] R. König, R. Renner, and C. Schaffner, “The operational meaning of min- and max-entropy,” IEEE Trans. Inf. Theory 55, 4337–4347 (2009).
[20] M. Tomamichel, A framework for non-asymptotic quantum information theory, Ph.D. thesis, ETH Zurich (2012), arXiv:1203.2142 [quant-ph].
[21] B. Schumacher, “Quantum coding,” Phys. Rev. A 51, 2738–2747 (1995).
[22] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. (John Wiley & Sons, 2012).
[23] A. Winter, “Coding theorem and strong converse for quantum channels,” IEEE Trans. Inf. Theory 45, 2481–2485 (1999).
[24] 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.
Appendix 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 \( \omega \) in the tilted-CHSH game of parameter \( \beta \).

To derive this bound, we use our self-testing result for the tilted-CHSH inequalities [12]. The robustness bounds for the self-test in [12] are rather unwieldy, so we will instead provide a crude upper bound that retains the same asymptotic behavior in \( \beta \) and \( \omega \) and greatly simplifies the use of the bound.

We will lower-bound \( H_{\text{min}}(A \mid E; X = 1) \) as a function of the expected violation of the tilted-CHSH inequality, \( I = I_q - \epsilon \), where \( I_q = \sqrt{8 + 2\beta^2} = 4/\sqrt{1 + \sin^2\theta} \) 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

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

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 [19]. Formulated in terms of a given physical state \( |\tilde{\psi}\rangle_{ABE} \) and observables \( \tilde{A}_x \equiv \tilde{A}_x \otimes I_B \otimes I_E \) and \( \tilde{B}_y \equiv I_A \otimes \tilde{B}_y \otimes I_E \), for a given adversary POVM \( \{M_g\} \) we have

\[
\Pr[A = G \mid X = 1] = \sum_{\alpha \in \{0,1\}} \langle \tilde{\psi} | I_A + (-1)^\alpha \tilde{A}_1 \rangle \otimes I_B \otimes M_{g,\alpha} |\tilde{\psi}\rangle
\]

\[
= \frac{1}{2} + \frac{1}{2} \langle \tilde{\psi} | \tilde{A}_1 \otimes I_B \otimes (M_0 - M_1) |\tilde{\psi}\rangle \equiv \frac{1}{2} + \frac{1}{2} \langle \tilde{A}_1 \rangle,
\]

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. \( I_A \otimes I_B \otimes C \), we simply write \( C \).) We will now relate this expression to the reference system using self-testing.

Using the notation of [12], the self-testing result shows that any pure state \( |\tilde{\psi}\rangle_{ABE} \) 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 \( I_q - \epsilon \) obeys\(^1\)

\[
||\Phi(|\tilde{\psi}\rangle_{ABE}) - |\psi_\theta\rangle_{AB} \otimes (\text{junk})_{BE}|| \leq 2\delta,
\]

\[
||\Phi(\tilde{A}_1 |\tilde{\psi}\rangle_{ABE}) - A_1 |\psi_\theta\rangle_{AB} \otimes (\text{junk})_{BE}|| \leq 2\delta + 2\delta_A^\Lambda,
\]

where \( \Phi = \Phi_A \otimes \Phi_B \otimes I_E \) 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 \( \delta, \delta_A^\Lambda = O(\sqrt{\epsilon} \theta^{-3}) \) are explicitly defined in [12]. 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 \( p_{\text{guess}} = \max_\alpha \Pr[A = a \mid X = 1] = 1/2 \) 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 (A3) as

\[
\langle \tilde{A}_1 \rangle = \Phi^\dagger(|\tilde{\psi}\rangle \otimes C \otimes |\tilde{\psi}\rangle)
\]

using that \( \Phi \) is an isometry which acts like the identity on \( E \). We then use (A4) and the triangle inequality to transform this into \( \langle \psi_\theta | A_1 |\psi_\theta\rangle \otimes (\text{junk}|C\rangle |\text{junk}) \), with additional error terms. First, we replace \( \Phi(\tilde{A}_1 |\tilde{\psi}\rangle) \) with \( A_1 |\psi_\theta\rangle \otimes (\text{junk}) \) with an additional error term \( 2\delta + 2\delta_A^\Lambda \) since \( ||\Phi^\dagger(|\tilde{\psi}\rangle \otimes C || \leq 1 \), then we replace \( \Phi^\dagger(|\tilde{\psi}\rangle) \), again using \( ||A_1||_{\infty} ||C||_{\infty} \leq 1 \):

\[
\langle \tilde{A}_1 \rangle \leq \langle \tilde{\psi} | A_1 |\psi_\theta\rangle \otimes C \otimes (\text{junk}) + 2\delta + 2\delta_A^\Lambda
\]

\[
\leq \langle \psi_\theta | A_1 |\psi_\theta\rangle \otimes C \otimes (\text{junk}) + 4\delta + 4\delta_A^\Lambda
\]

\[
\leq |\langle \psi_\theta | A_1 |\psi_\theta\rangle| + 4\delta + 2\delta_A^\Lambda.
\]

This last inequality is generally saturated by one of \( C \in \{\pm I\} \), which corresponds to the adversary making a constant guess for the ideal system’s most likely outcome.\(^2\) Since for the reference system \( \langle \psi_\theta | A_1 |\psi_\theta\rangle = 0 \), we find that

\[
p_{\text{guess}} \leq \frac{1}{2} + 2\delta + \delta_A^\Lambda
\]

---

\(^1\) In [12] 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 \( E \).

\(^2\) Of course, this strategy for the adversary maximizes the guessing probability bound, and not necessarily the guessing probability itself.
We now proceed to find a simple expression of $\epsilon$ and $\theta$ that upper-bounds $2\bar{\delta} + \delta_a^A$. After some careful manipulation of the rather long expressions for $\bar{\delta}$ and $\delta_a^A$, we find

\[
2\bar{\delta} + \delta_a^A = \sqrt{2T_q}\sqrt{\epsilon}\left[\frac{\sqrt{1 + s^2}}{2s^2} \left(1 + c + \sqrt{1 + s^2}\right) + \frac{\sqrt{1 + s^2}}{4s} \left(2 - c + \sqrt{1 + s^2}\right) + \frac{c + \sqrt{1 + s^2}}{2s^2} (1 + c) \left(8 + 2 \frac{1 + \sqrt{1 + s^2}}{s^2} + 3 \tan \theta\right)\right], \quad (A10)
\]

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

A crude upper bound on $(A10)$ 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 $s^{-4}\sqrt{2T_q}\sqrt{\epsilon}$ in the first term becomes $s^2 \sqrt{1 + s^2} / (1 + \sqrt{1 + s^2})^2 \leq (3\sqrt{2}/2)s^2$ 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:

\[
2\bar{\delta} + \delta_a^A \leq 2\sqrt{2} \left[\frac{3\sqrt{2}}{2} s^2 + \frac{1 + \sqrt{2}}{2} s^3 + 16s^2 + 8 + 6s^2 \tan \theta\right] \frac{\sqrt{\epsilon}}{s^4}, \quad (A11)
\]

where we have also bounded $I_q \leq 4$ and used the following tight bounds after expanding the third term in the square brackets of $(A10)$:

\[
(c + \sqrt{1 + s^2})(1 + c) \leq 4, \quad (A12)
\]

\[
(c + \sqrt{1 + s^2})(1 + c)(1 + \sqrt{1 + s^2}) = (1 + c\sqrt{1 + s^2})(2 + c + \sqrt{1 + s^2}) \leq 8. \quad (A13)
\]

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

\[
2\bar{\delta} + \delta_a^A \leq 2\sqrt{2} \left[\frac{1 + \sqrt{2}}{2} s^3 + \frac{3\sqrt{2} + 44}{2} s^2 + 8\right] \frac{\sqrt{\epsilon}}{s^4}. \quad (A14)
\]

Finally, the polynomial in $s$ in square brackets is bounded by its maximum at $s = 1$. Eq. (A9) then becomes

\[
p_{\text{guess}} - \frac{1}{2} \leq 2\bar{\delta} + \delta_a^A \leq (8 + 61\sqrt{2}) \frac{\sqrt{\epsilon}}{s^4}. \quad (A15)
\]

Further bounding

\[
s = \sin 2\theta \geq \frac{2\theta}{\pi/2} \geq \theta \quad (A16)
\]

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

\[
p_{\text{guess}} \leq \frac{1}{2} + \sqrt{8 + 2\beta} (8 + 61\sqrt{2}) \theta^{-4} \sqrt{\omega_q - \omega}. \quad (A17)
\]

Putting this together with $(A1)$, we find, using $\ln(1 + x) \leq x$,

\[
H_{\text{min}}(A \mid E; X = 1) \geq 1 - \log_2 \left(1 + 2\sqrt{8 + 2\beta} (8 + 61\sqrt{2}) \theta^{-4} \sqrt{\omega_q - \omega}\right) \quad (A18)
\]

\[
\geq 1 - \kappa \theta^{-4} \sqrt{\omega_q - \omega} \quad (A19)
\]

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

A numerical maximization of the factor of $\sqrt{\epsilon}$ in $(A10)$ shows that a tighter numerical factor of 45.13 could replace the numerical factor $8 + 61\sqrt{2} = 94.27$ in $(A15)$, or less if the range of $\theta$ is limited.
Appendix B: Proof of Theorem 3

It is shown in [10, 11] 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 [11] 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 main text) can be thought of as a quantum channel \( G_i \) mapping the state \( \rho = \rho_{DE} \) of the pair of devices \( D \) and the adversary information \( E \) before that round to the resulting state \( \rho_i(\rho) = \rho' = \rho_{A_iB_iX_iY_iT_iC_iD'E} \) after the protocol round, which also includes explicitly the classical data that was produced in that round. In particular, the channel \( G_i \) and the initial state \( \rho \) 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

\[
 f_{\min}(p) \leq H(A_i | X_i Y_i T_i E)_{G_i(\rho)}, \tag{B1}
\]

where \( H(A_i | X_i Y_i T_i E)_{G_i(\rho)} \) 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 \( G_i \) that are compatible with the protocol and for all initial states \( \rho \) such that the variable \( C_i \) in \( G_i(\rho) \) is distributed as \( p \).

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 [11], 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 \( p_1/\gamma \). In that case, we can use Lemma 2 with \( \omega = p_1/\gamma \) to set the value of \( f_{\min} \). Indeed,

\[
 H(A_i | X_i Y_i T_i E) \geq H(A_i | X_i Y_i T_i E) \tag{B2}
 = H(A_i | X_i E) \tag{B3}
 \geq \Pr[X_i = 1] H(A_i | E; X_i = 1) \tag{B4}
 \geq \nu g(\omega), \tag{B5}
\]

with \( \nu = \Pr[X_i = 1] = 1 - \gamma (2 + \beta)/(4 + \beta) \). In (B2), we used the chain rule and the positivity of the conditional entropy of classical information. In (B3), 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 (B5), we used Lemma 2 and the fact that the conditional min-entropy lower-bounds the conditional von Neumann entropy [20, 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,

\[
 f_{\min}(p) = \nu g(p_1/\gamma). \tag{B6}
\]

The EAT [10] requires affine min-tradeoff functions. Since \( g \) is convex, we can simply obtain affine lower bounds of (B6) by taking its tangent at any point. The tangent of \( g(\omega) \) at the point \( \omega = \omega_k \) is given by

\[
 \tilde{g}_{\omega_k}(\omega) = 1 - \kappa \theta^{-4} \frac{2\omega_q - \omega_k - \omega}{2\sqrt{\omega_q - \omega_k}}, \tag{B7}
\]

hence the min-tradeoff function we finally use will be \( f_{\min}(p) = \nu \tilde{g}_{\omega_k}(p_1/\gamma) \) for some appropriately chosen \( \omega_k \).

Given such a min-tradeoff function, Lemma 9 of [11] then states that for any given pair of devices, either the protocol succeeds with low probability \( \Pr[S] \leq \epsilon' \), or

\[
 H_{\min}^A(AB | XYTE; S) \geq \nu \tilde{g}_{\omega_k}(\omega_q - \xi) - \mu \sqrt{m} \tag{B8}
\]

with

\[
 \mu = 2 \log_2(13) + \|\|\nabla f_{\min}\|\|_\infty) \sqrt{1 - 2 \log_2(\epsilon \epsilon')} \tag{B9}
\]

The gradient of the min-tradeoff function is simply the slope of \( \nu \tilde{g}_{\omega_k}(p_1/\gamma) \):

\[
 \|\nabla f_{\min}\|_\infty = \gamma^{-1} \nu \frac{\kappa \theta^{-4}}{2 \sqrt{\omega_q - \omega_k}}. \tag{B10}
\]

Bounding the ceiling function as \( \lceil x \rceil \leq x + 1 \) and optimizing over the point of tangency \( \omega_k \) produces the final expression (6) for the min-entropy bound of the theorem. \( \Box \)
Appendix C: Proof of Lemma 6

In this appendix, we detail the procedure of entanglement dilution used in our randomness generation protocol and prove Lemma 6. 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 [21] and is explained in great detail in [16], from which we borrow the notation.

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

Schumacher compression works on a source of pure states \(|\{\psi_i\}, p_i\rangle\) which outputs the states in \(|\{\psi_i\}\rangle\) at random, with respective probabilities \( \{p_i\} \). The goal is to pack the information output by 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 Schumacher compression as describing a quantum source of mutually orthogonal states, namely, the eigenstates of the reduced operator with probabilities equal to the corresponding eigenvalues. The eigenstates coincide with the computational basis: they are \(|\{y\}\rangle\) with \( y \in \{0, \ldots, 2^n - 1\} \), where \(|y\rangle\) is the tensor product of \( n \) qubits in the states described by the binary digits \( \{y_i\} \) of \( y \). For instance, if \( n = 4 \), \(|y = 2\rangle \equiv |0\rangle |0\rangle |1\rangle |0\rangle\). The eigenvalue associated with \( y \) is \( \lambda_y = (\cos^2 \theta)^j (\sin^2 \theta)^{n-j} \) for \( j = n(0 | y) \), which gives the number of zeros in the binary digits of \( y \).

Compressing a source of distinguishable states is more or less equivalent to Shannon source coding. The idea is to consider the distribution of probabilities \( q_1, \ldots, q_n \) that describe a quantum source of mutually orthogonal states. The typical set is defined as

\[
T_\delta = \left\{ y \in \{0, 1\}^n : S - \delta \leq -\frac{\log_2 P(y)}{n} \leq S + \delta \right\},
\]

where

\[
S = h_2(\sin^2 \theta) = -\sin^2 \theta \log_2(\sin^2 \theta) - \cos^2 \theta \log_2(\cos^2 \theta)
\]

is the entropy of a single pair \(|\psi_y\rangle\). Thus, a sequence \( y \) is \( \delta \)-typical if and only if its sample entropy \(- (1/n) \log_2 P(y)\) is \( \delta \)-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 \( \{q_0; q_1 = 1 - q_0\} \). Thus, the sample entropy for a given value \( y \) can be rewritten as

\[
-\frac{1}{n} \log_2 P(y) = \frac{1}{n} \sum_{i=1}^{n} -\log_2 q_{y_i}
\]

\[
= -\frac{n(0 | y)}{n} \log_2 q_0 - \frac{n(1 | y)}{n} \log_2 q_1
\]

\[
= -\log_2 q_0 + \frac{n(1 | y)}{n} \Delta \tag{C5}
\]

where \( \Delta = \log_2(q_0/q_1) = -\log_2 \tan^2 \theta \). Hence, the definition of the typical set \((C1)\) can be rewritten as

\[
T_\delta = \left\{ y \in \{0, 1\}^n : q_1 - \frac{\delta}{\Delta} \leq \frac{1}{n} \sum_{i=1}^{n} y_i \leq q_1 + \frac{\delta}{\Delta} \right\}. \tag{C6}
\]

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

Properties of the typical set are easily derived from these two forms [22]. First, applying Hoeffding’s inequality [13] for the binomially-distributed \( \sum_i Y_i \) shows that the typical set has a high probability weight

\[
\Pr[T_\delta] \geq 1 - \epsilon_\pi, \tag{C7}
\]

where we defined the projection error

\[
\epsilon_\pi = 2 \exp(-2n\delta^2/\Delta^2), \tag{C8}
\]

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 (C1),

\[ |\mathcal{T}_\delta| = 2^{n(S+\delta)} \sum_{y \in \mathcal{T}_\delta} 2^{-n(S+\delta)} \leq 2^{n(S+\delta)} \sum_{y \in \mathcal{T}_\delta} P(y) \leq 2^{n(S+\delta)}, \]

which is smaller than the total number of \(2^n\) sequences if \(S < 1 - \delta\).

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 codewords a length of at most \(\log_2|\mathcal{T}_\delta| \leq n(S + \delta) = m\).

Schumacher compression applies this procedure to the quantum state \(\rho_{\Lambda'} = \Sigma_{y=0}^{2^n-1} \lambda_{m(0|y)} |y\rangle \langle y|\), which describes the output of a quantum source of pure states \(|y\rangle\) with probabilities \(\lambda_{m(0|y)}\). In order to identify an atypical state, a projective typicality measurement is performed, with projectors \(\{\Pi_\delta, I - \Pi_\delta\}\) where

\[ \Pi_\delta = \sum_{y \in \mathcal{T}_\delta} |y\rangle \langle y|. \]

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 \(\sigma\) is substituted and encoded the same way. The resulting state is therefore \(S(\rho_{\Lambda'}) = V[\Pi_\delta \rho_{\Lambda'} \Pi_\delta] + (1 - \Tr(\Pi_\delta \rho_{\Lambda'})) \sigma) V^\dagger\).

The two properties (C7) and (C10) of the typical set can be expressed in terms of the typical projector \(\Pi_\delta\):

\[ \Tr(\Pi_\delta \rho_{\Lambda'}) \geq 1 - \epsilon_\pi, \quad \text{(C12)} \]

\[ \Tr(\Pi_\delta) \leq 2^{n(S+\delta)}. \quad \text{(C13)} \]

The first property can be used in the gentle operator lemma [23, 24] to show that a successful typicality measurement does not disturb the state by much [16]:

\[ \|\Pi_\delta \rho_{\Lambda'} \Pi_\delta - \rho_{\Lambda'}\|_1 \leq 2\sqrt{\epsilon_\pi}. \quad \text{(C14)} \]

Hence, the decompressed state is close in trace distance to the original [16]:

\[ \|V^\dagger S(\rho_{\Lambda'}) V - \rho_{\Lambda'}\|_1 \leq 2\sqrt{\epsilon_\pi} + \epsilon_\pi. \quad \text{(C15)} \]

As Schumacher originally noted [21], this remains true when we consider the global state in \(\Lambda A'\); the entanglement of the state is therefore not destroyed by compression:

\[ \|(I \otimes V^\dagger)(\Id \otimes S)[|\psi_\theta\rangle \langle \psi_\theta|^{0n}](I \otimes V) - |\psi_\theta\rangle \langle \psi_\theta|^{0n}\|_1 \leq 2\sqrt{\epsilon_\pi} + \epsilon_\pi. \quad \text{(C16)} \]

Thus, defining \(\mathcal{D}_{\delta,\theta}(|\phi^+\rangle \langle \phi^+|^{0m}_{A'\Lambda'})\) to be the outcome of the composition of a local preparation of \(|\psi_\theta\rangle \langle \psi_\theta|^{0n}_{\lambda_{\Lambda'}}\), followed by Schumacher compression over \(\delta\)-typical sequences of \(\Lambda'\), teleportation from \(\Lambda'\) to \(B\) using \(|\phi^+\rangle \langle \phi^+|^{0m}_{AB}\) and decompression on \(B\), Lemma 6 follows. □