


Randomness Expansion Secured by Quantum Contextuality

Mark Um,¹ Qi Zhao,¹ Junhua Zhang,^{1,2} Pengfei Wang^{1b},¹ Ye Wang^{1b},¹ Mu Qiao,¹ Hongyi Zhou,¹ Xiongfeng Ma,^{1,†} and Kihwan Kim^{1,*}

¹*Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, People's Republic of China*

²*Shenzhen Institute for Quantum Science and Engineering, and Department of Physics, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China*

 (Received 1 March 2019; revised manuscript received 13 January 2020; accepted 2 March 2020; published 31 March 2020)

The output randomness from a random number generator can be certified by observing the violation of quantum contextuality inequalities based on the Kochen-Specker theorem. Contextuality can be tested in a single quantum system, which significantly simplifies the experimental requirements to observe the violation comparing to the ones based on nonlocality tests. However, it is not yet resolved as to how to ensure compatibilities for sequential measurements that is required in contextuality tests. Here, we employ a modified Klyachko-Can-Binicioğlu-Shumovsky contextuality inequality, which can ease the strict compatibility requirement on measurements. On a trapped single $^{138}\text{Ba}^+$ ion system, we experimentally demonstrate violation of the contextuality inequality and realize quantum random number expansion by closing detection loopholes. We perform 1.29×10^8 trials of experiments and extract a randomness of 5.28×10^5 bits with a speed of 270 bits s^{-1} . Our demonstration paves the way for practical high-speed spot-checking quantum random number expansion and other secure information processing applications.

DOI: [10.1103/PhysRevApplied.13.034077](https://doi.org/10.1103/PhysRevApplied.13.034077)

I. INTRODUCTION

Randomness is a critical resource for information processing with applications ranging from computer simulations [1] to cryptography [2]. For cryptographic purposes, in particular, streams of random numbers should have good statistical behavior and unpredictability against adversaries [3,4]. In reality, random numbers produced by an algorithm or a classical chaotic process are intrinsically deterministic, thereby in principle allowing an adversary with the information of the device to find a pattern. On the other hand, the nature of quantum mechanics is fundamentally random, which, in this sense, provides a foundation for genuine randomness. Due to the unpredictable behavior of quantum mechanics, various quantum random number generators have been proposed and implemented [5–7]. In practice, however, security can be jeopardized if an adversary partially manipulates the devices or the devices are exposed to imperfection or malfunction. In order to address this realistic issue, device-independent protocols have been proposed to guarantee generated randomness without relying on detailed knowledge of uncharacterized devices [8–17].

The essence of device-independent randomness expansion lies in the fact that any violation of nonlocality

inequalities [18] shows unpredictability of measurement results. Recent security proofs show that randomness can be certified under the device-independent scenario by a class of Bell inequalities [8–14]. On the experimental side, loophole-free violations of Bell's inequality have been demonstrated [19–21], which have been applied to generate random numbers [22,23]. However, the randomness certification by the loophole-free Bell test suffers from a low generation rate and requires high-fidelity entanglement sources. Moreover, it requires a large space separation between two detection sites to rule out the locality loophole, meaning that it is almost impossible to make the whole system compact. In Refs. [22] and [23], the output randomness is less than the input randomness. Until now, a strict and practical randomness expansion, i.e., more output randomness than input randomness and practical randomness expansion, based on loophole-free Bell tests still has not been demonstrated and remains as an experimental challenge.

Similar to the Bell theorem, the Kochen-Specker theorem [24,25] states that quantum mechanics cannot be fully explained by noncontextual hidden variable models that have definite predetermined values for measurement outcomes. Contextuality can be tested with a single system without entanglement by using the Klyachko-Can-Binicioğlu-Shumovsky (KCBS) inequality [26], which can significantly reduce the experimental requirements compared to the nonlocality test. Inequalities based on

*kimkihwan@mail.tsinghua.edu.cn

†xma@tsinghua.edu.cn

the Kochen-Specker theorem can provide alternatives for randomness certification, which has been studied in both theory and experiment [14,27,28]. A contextuality test contains a set of contexts that are composed of a certain number of compatible, i.e., commuting in quantum mechanics, measurements. Note that the measurements in the nonlocality Bell test can also be regarded as compatible measurements. The randomness certification has been proven for the case with perfectly compatible measurements [14]. In reality, when the contextuality test is performed on a single party, it is difficult to establish perfect compatibility between sequential measurements. Although a couple of experimental demonstrations of randomness certification with the KCBS inequality have been reported [27,28], the security of the scheme has not been fully resolved.

In this work, first, we experimentally demonstrate the violation of a modified KCBS inequality [29,30], which reveals quantum correlations without the requirement of perfect compatibility on sequential measurements. Then we employ it for a spot-checking protocol of randomness expansion with exponential gain [14]. Our scheme is not a fully device-independent protocol, since it requires a few assumptions on hidden variable models, which are shown in Sec. III.

In this scenario, we can expand the randomness from the generated strings merely based on the experimental observed data that violate the modified KCBS inequality, which is a self-testing manner [6,31]. We implement the protocol with a single trapped $^{138}\text{Ba}^+$ ion instead of a $^{171}\text{Yb}^+$ ion that was used for a previous demonstration [28] in order to fully address the experimental requirements of a modified KCBS inequality [29,30]. The $^{138}\text{Ba}^+$ ion has long-lived states that can be used for the coherent shelving of a quantum state during sequential measurements. We develop a narrow-line laser system that is stabilized to a high-finesse cavity to precisely manipulate the long-lived states and observe a sufficient amount of violation for randomness expansion with a large enough number of trials. We perform 1.29×10^8 trials of experiments and extract a randomness of 5.28×10^5 bits with a speed of 270 bits s^{-1} .

II. MODIFIED KCBS INEQUALITY

In order to test contextuality, various inequalities have been proposed [26,32] and demonstrated in diverse physical systems, including trapped ion system [33–35], photonic system [36,37], and superconducting system [38]. Among the contextuality inequalities, the KCBS inequality, which uses five observables A_i taken ± 1 , shows that noncontextual hidden variable models [26] satisfy

$$\langle A_1 A_2 \rangle + \langle A_3 A_2 \rangle + \langle A_3 A_4 \rangle + \langle A_5 A_4 \rangle + \langle A_5 A_1 \rangle \geq -3. \quad (1)$$

In quantum mechanics, on the other hand, the inequality can be violated for a specific state with properly arranged observables A_i . In the case of $d = 3$, denote the basis states by $|1\rangle$, $|2\rangle$, and $|3\rangle$. Design the observable $A_i = 1 - 2|v_i\rangle\langle v_i|$ to be the projector along the axis of $|v_i\rangle$. The maximal violation of the inequality (1) is achieved when five state vectors, $\{|v_i\rangle\}$, form a regular pentagram, and the initial state vector passes through the center of the pentagram, as shown in Fig. 1. In this case, the sum of all the terms in (1) achieves $5 - 4\sqrt{5} \approx -3.944$. The assumption behind the above contextuality inequality is that the observables A_i and A_{i+1} (let $A_6 \equiv A_1$) are compatible. However, in an actual experiment using sequential measurements, the compatibility is difficult to verify, which leads to opening of the compatibility loophole. The issues of the compatibility in sequential measurements have been

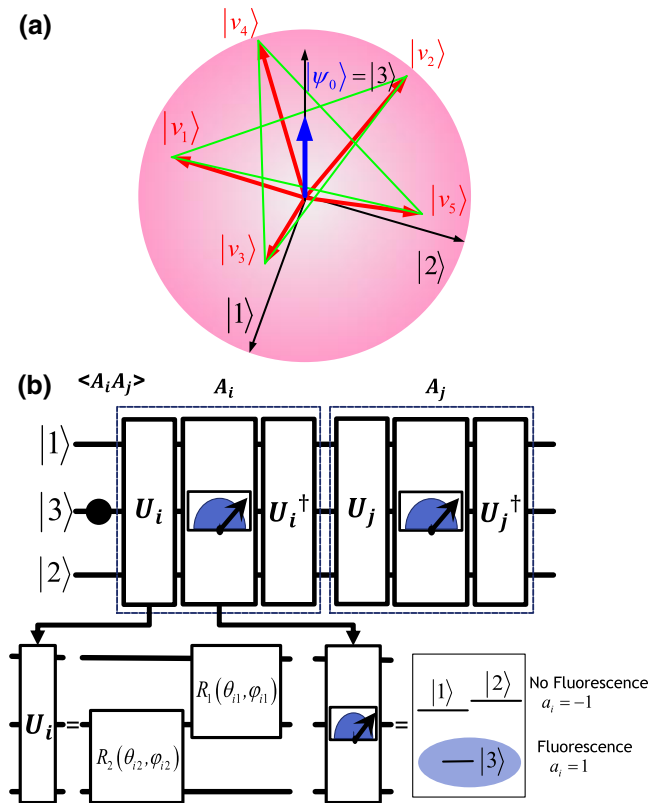


FIG. 1. KCBS pentagram and experimental procedure. (a) Initial state and five axes that form a pentagram in $d = 3$ space. The five observables A_1, A_2, \dots, A_5 are the projectors on the five axes, respectively. The connected axes $|v_i\rangle$ and $|v_{i+1}\rangle$ are orthogonal, representing compatibility of the corresponding observables A_i and A_{i+1} . (b) Initially, we prepare $|3\rangle$ state, then perform two sequential measurements of A_i and A_j . Each sequential measurement contains a unitary rotation U_i , projective measurement, and an inverse unitary rotation U_i^\dagger . Each unitary rotation U_i is comprised of first $R_2(\theta_{2i}, \phi_{2i})$ then $R_1(\theta_{1i}, \phi_{1i})$. In projective measurement, we assign $a_i = 1$ (-1) if fluorescence is (is not) detected.

addressed by modifying the KCBS inequality [29,30] (see also Appendix E).

We combine two modifications of the KCBS inequality to relax the condition of the perfect compatibility, which introduce additional terms of ϵ 's [29] and $\langle A_1 A_1 \rangle$ [30]:

$$\begin{aligned} \langle \chi_{\text{KCBS}} \rangle &= \langle A_1 A_2 \rangle + \langle A_3 A_2 \rangle + \langle A_3 A_4 \rangle + \langle A_5 A_4 \rangle \\ &\quad + \langle A_5 A_1 \rangle - \langle A_1 A_1 \rangle \\ &\geq -4 - (\epsilon_{12} + \epsilon_{32} + \epsilon_{34} + \epsilon_{54} + \epsilon_{51} + \epsilon_{11}). \end{aligned} \quad (2)$$

Here, $\langle A_i A_j \rangle$ denotes the expectation value of the measurement results in the time order of $A_i A_j$ for the sequential measurements and $\langle A_j | A_i A_j \rangle$ denotes the expectation value of observable $\langle A_j \rangle$ in these $A_i A_j$ sequential measurements. The terms ϵ_{ij} describe the difference between a same pair of observables A_i and A_j in different time orders, $A_i A_j$ and $A_j A_i$, which can be regarded as the bound of incompatibility between these sequential measurements [29], $\epsilon_{ij} = |\langle A_j | A_j A_i \rangle - \langle A_j | A_i A_j \rangle|$.

The term $\langle A_1 A_1 \rangle$ is later introduced to address different types of incompatibility, which cannot be excluded with the terms ϵ_{ij} [30]. The derivation of this inequality is shown in Appendix E. In our work, we include both of the modifications that address all types of incompatibility discussed in Refs. [29] and [30].

A. Randomness expansion protocol

The violation of the KCBS inequality implies the existence of quantum randomness that cannot be imitated by classical variables, which is not only fundamentally interesting but also has value for practical applications. The noncontextuality inequalities provide an alternative way of generating secure randomness. Similar to the Bell inequality, in each trial, certain bits of randomness are consumed. Thus in order to efficiently expand the randomness from small input randomness, the idea of spot checking is necessary in our scheme. Recently, a robust (error-tolerant) randomness expansion scheme has been proposed [14], which is a spot-checking protocol that achieves exponential expansion. The denotation and protocol are shown as follows, with our experimental settings.

Denotation

(1) G : KCBS game with 11 random inputs $\{1, 2\}, \{2, 1\}, \{2, 3\}, \{3, 2\}, \{3, 4\}, \{4, 3\}, \{4, 5\}, \{5, 4\}, \{5, 1\}, \{1, 5\}, \{1, 1\}$ for the game rounds, and the input $\{1, 2\}$ is also for the generation rounds.

(2) D : a quantum device compatible with G .

(3) Output length N : $N_{\text{exp}} = 1.29 \times 10^8$ in experiment.

(4) Test probability $q \in (0, 1)$: $q_{\text{exp}} = 10^{-4}$ in experiment.

(5) Score threshold $\chi_g \in (0, 1)$: $\chi_g = 2/3$ in this KCBS game.

Protocol R_{gen}

(1) Choose a bit $t \in \{0, 1\}$ according to the binomial distribution $(1 - q, q)$.

(2) If $t = 1$ (“game round”), game G is played with D and the output is recorded. Outputs of game rounds are additionally collected for checking.

(3) If $t = 0$ (“generation round”), $\{1, 2\}$ is given to D and the output is recorded.

(4) Steps 1–3 are repeated N times.

(5) Calculate the average score g_{KCBS} from all game round outputs. If $g_{\text{KCBS}} < \chi_g$, then abort. Otherwise, move to randomness extraction.

According to the definition of Ref. [14], the score of the KCBS game is given by $g \in \{0, 1\}$. Thus, (2) can be rewritten in the form KCBS game G as

$$\begin{aligned} g_{\text{KCBS}} &= -\frac{1}{6}(\langle A_1 A_2 \rangle + \langle A_3 A_2 \rangle + \langle A_3 A_4 \rangle + \langle A_5 A_4 \rangle + \langle A_5 A_1 \rangle \\ &\quad - \langle A_1 A_1 \rangle) + \epsilon_{12} + \epsilon_{32} + \epsilon_{34} + \epsilon_{54} + \epsilon_{51} + \epsilon_{11}. \end{aligned} \quad (3)$$

The classical winning probability is $\chi_g = 2/3$ (see Appendix F for details) and the achievable maximal quantum winning probability is $\chi'_g = (4\sqrt{5} - 4)/6 \approx 0.824$. The gap between χ_g and χ'_g enables randomness expansion.

III. ASSUMPTIONS OF OUR PROTOCOL

In our scheme, the amount of randomness quantified by the min-entropy is related to the violation of the modified KCBS inequality (see Appendix A). The min-entropy is the minimum of the Rényi entropies. If the device obtains a superclassical average score, then it must exhibit certain quantumness, which implies unpredictable randomness. The violation is only based on the observation of experimental data, and can be independent of the sources of prepared states and other device specifications. In our scheme, there are three underlying main assumptions: (1) the input is chosen from an independent random distribution uncorrelated with the system; (2) the measurement outcomes cannot be leaked directly to adversaries; and (3) an adversary's behaviors [hidden variable (HV) strategies] are characterized contextual such that HV models can be characterized by the degree of incompatibility ϵ_{ij} and $\langle A_1 A_1 \rangle$ in Eq. (2). Assumptions (1) and (2) are widely used in other self-testing tasks, such as device-independent quantum random number generators [11,14,17]. Assumption (3) is related to the validity of the quantum contextuality test, which is similar to all the other experimental contextuality tests with sequential measurements.

In a fully device-independent scheme, spacelike separation can guarantee that two measurements are compatible. However, in our scheme, as well as other quantum

contextuality tests, two measurements in a context are performed on a single system. In principle, we cannot prevent incompatible sequential measurements. A malicious manufacturer can use contextual HV models by registering the setting and results of the first measurements and using them for the second measurements. Thus, we need additional assumption (3) for the measurement devices. Note that assumption (3) is more applicable than the case of indirectly restricting an adversary to noncontextuality HV, which is most widely used in contextuality tests and consistent with practical implementation with imperfect compatible measurements. Here it helps us to avoid a malicious adversary with contextual HV models. We require the underlying probability distributions generated by the measurement or HV models to have the same properties as all accessible distributions, which is the same argument as in Ref. [29]. Then, we could characterize HV or restrict an adversary by the testable incompatibility ϵ_{ij} and $\langle A_1 A_1 \rangle$. For simplicity, we call this assumption the “characterized contextual” assumption. The detailed derivation is shown in Appendix E.

Therefore, our protocol cannot be viewed as a fully device-independent scenario, but a source-independent scenario where we do not need any assumptions of the source [39–42]. We assume that the adversary’s behaviors can be characterized by additional incompatible terms, but it is fine to have imperfections in the realization and disturbance from classical or quantum noisy environments since the amounts of introduced incompatibilities are quantified. In the standard source-independent scenarios, the measurements are assumed to be fully characterized without any imperfections, which is much stronger than our assumption. Our protocol is well fitted to a scenario of trusted but error-susceptible devices. Given these assumptions, the generated randomness is certified by only experimental statistics. Thus our security level is between device independent and source independent.

IV. QUTRIT AND EXPERIMENTAL PROCEDURE

Randomness expansion based on the experimental violations of the KCBS inequality using a single trapped $^{171}\text{Y b}^+$ ion has been demonstrated [28]. In the demonstration, however, it is not possible to test the modified KCBS inequality (2), due to lack of ability to obtain all correlations. For example, when we observe fluorescence in the first measurement, the second measurement does not provide any useful information [28]. Instead, we develop a single $^{138}\text{Ba}^+$ ion system [43,44] with which we can obtain full-correlation results from sequential measurements by using long-lived shelving states in the $^5D_{5/2}$ manifold similar to a $^{40}\text{Ca}^+$ ion [35]. We choose two Zeeman sublevels ($|m_j = +1/2\rangle \equiv |1\rangle$, $|m_j = +3/2\rangle \equiv |2\rangle$) in the $^5D_{5/2}$ manifold and one Zeeman sublevel ($|m_j = +1/2\rangle \equiv |3\rangle$) in the $^6S_{1/2}$ manifold to represent the qutrit system

as shown Fig. 2(a). In the projective measurement, we observe fluorescence when the state is projected to $|3\rangle$ and no fluorescence for all the other projections on the subspace that consists of $|1\rangle$ and $|2\rangle$ basis while conserving coherence. Different from the $^{171}\text{Y b}^+$ ion realization, since the coherence is not destroyed even when we observe fluorescence in the first measurement, we can get meaningful outcomes in the second measurement. The transitions between $^6S_{1/2}$ and $^5D_{5/2}$ are coherently manipulated by a

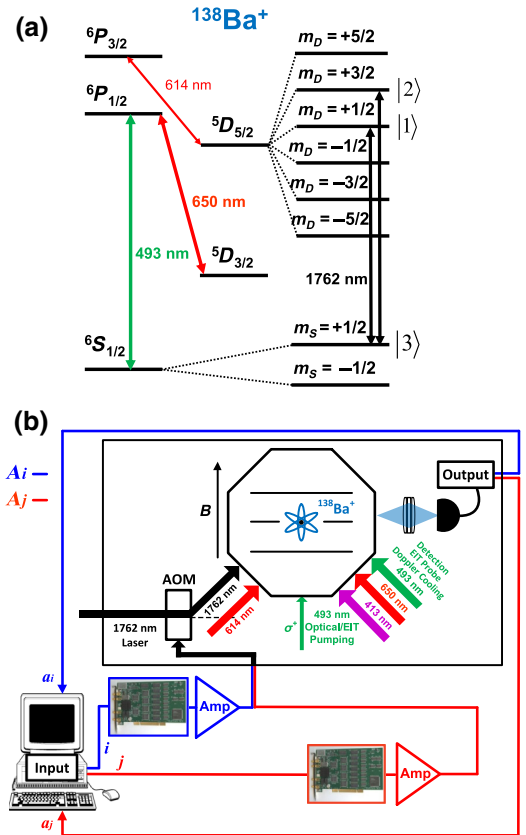


FIG. 2. Experimental setup of the $^{138}\text{Ba}^+$ ion system. (a) The energy level diagram of a $^{138}\text{Ba}^+$ ion for a qutrit system, which is represented by two Zeeman sublevels ($|m_D = +1/2\rangle \equiv |1\rangle$, $|m_D = +3/2\rangle \equiv |2\rangle$) in the $^5D_{5/2}$ manifold and one sublevel ($|m_S = +1/2\rangle \equiv |3\rangle$) in the $^6S_{1/2}$ manifold. The quadrupole transitions between $^6S_{1/2}$ and $^5D_{5/2}$ are coherently manipulated using a narrow-line 1762-nm laser that is stabilized to a high-finesse cavity. Lasers of 493 and 650 nm are used for Doppler cooling, EIT cooling, optical pumping, and detection. A 614-nm laser is used for depopulation of the $^5D_{5/2}$ level to the $^6S_{1/2}$ level. (b) The experimental setup of a trapped $^{138}\text{Ba}^+$ ion for testing the KCBS inequality and for spot checking random number expansion. One of 11 measurement configurations $\{A_i, A_j\}$ is randomly selected. Their pulse sequences are independently generated by their own direct digital synthesizer and amplifiers, sent to an AOM through independent paths, and finally applied to the ion at different time order. Fluorescence is observed by a photomultiplier tube at different time order and the values of the observables are assigned accordingly.

narrow-line laser with a wavelength of 1762 nm, which is stabilized to a high-finesse optical cavity. The coherent rotations R_1 (θ_1, ϕ_1) between $|1\rangle$ and $|3\rangle$ and R_2 (θ_2, ϕ_2) between $|2\rangle$ and $|3\rangle$ (see Appendix B for details) are realized by applying the 1762-nm laser beam, where θ and ϕ are controlled by the duration and the phase of the laser beam, respectively, using an acousto-optic modulator (AOM).

The procedure for the experimental test of the KCBS inequality consists of Doppler and electromagnetically induced transparency (EIT) cooling [45–47], initialization, the first projective measurement of observable A_i , and the second projective measurement of A_j . Initialization to state $|3\rangle$ is performed by applying an optical pumping beam of 493 nm with σ^+ polarization shown in Fig. 2(b). The first measurement of the observable A_i is realized by the rotation U_i , the projective measurement, and the reverse of the rotation U_i^\dagger (see Appendix C). Rotation U_i maps the axis $|v_i\rangle$ to the axis $|3\rangle$ and the projective measurement can be described as the projector $M_{|3\rangle} = 2|3\rangle\langle 3| - 1$ (see Appendix B). Thus A_i is assigned to value $a_i = 1$ when fluorescence is observed and $a_i = -1$ when no fluorescence is observed. The projective measurement consists of the state-dependent fluorescence detection and the optical pumping sequence (see Appendix C). The second measurement of the observable A_j is realized using the same scheme as that for the first measurement. Unitary rotations of A_i (Alice) and A_j (Bob) are realized by different signal generators and amplifiers; their results are also collected independently. In this way, we can eliminate the potential correlations between two measurements caused by a possible memory effect on their signals when the same signal generator and amplifier are used.

V. RESULTS

To test the modified KCBS inequality (2), we need to measure the 11 combinations of sequential measurements, which include five terms explicitly shown in inequality (2) and the other five terms with reverse order that are necessary to observe ϵ terms. The detailed experimental results of the measurements are summarized in Table I.

From the security proof [Eq. (A4) of Appendix A], we can see that when the violation is small, the total number of rounds N is a critical parameter. A positive generation rate requires a sufficiently large N . Thus we give the minimum required rounds for different violations, which is instructive for experiments. Figure 3(a) shows the minimum total rounds N_{\min} to obtain net randomness depending on the KCBS game score g_{KCBS} . In order to gain net randomness at our experimentally observed $g_{\text{KCBS}} = 0.790$, we perform $N_{\text{exp}} = 1.29 \times 10^8$ rounds, which is sufficiently larger than $N_{\min} = 6.2 \times 10^7$. Under our experimental condition of N_{exp} , Fig. 3(b) shows the generation rate of net

TABLE I. Experimental results for different observables and compatibility terms for the KCBS inequality (2). Total game rounds are 1.2×10^4 . The standard deviations of the final result are 0.015 and 0.023 for the single observables and correlations, respectively, and 10^{-3} order for the compatibility terms, all as shown in parentheses. The standard deviation for the violation σ is 0.101 and our experimental data show the violation of the extended inequality (2) with 7σ . The bold fonts indicate the terms shown in the KCBS game Eq. (3). All the terms are used for ϵ_{ij} .

| $\{i, j\}$ | $\langle A_i A_j \rangle$ | $\langle A_i \rangle$ | $\langle A_j \rangle$ | ϵ_{ij} |
|--|---------------------------|-----------------------|-----------------------|------------------|
| {1,2} | -0.768(23) | 0.082(15) | 0.091(15) | 0.005(21) |
| {2, 1} | -0.783(23) | 0.096(15) | 0.065(15) | 0.017(21) |
| {2, 3} | -0.767(22) | 0.098(14) | 0.088(14) | 0.019(21) |
| {3,2} | -0.750(23) | 0.107(15) | 0.098(15) | 0.000(21) |
| {3,4} | -0.773(23) | 0.084(15) | 0.082(15) | 0.040(20) |
| {4, 3} | -0.762(22) | 0.122(14) | 0.068(14) | 0.016(21) |
| {4, 5} | -0.782(23) | 0.095(15) | 0.075(15) | 0.019(21) |
| {5,4} | -0.789(22) | 0.056(15) | 0.094(15) | 0.002(21) |
| {5,1} | -0.773(22) | 0.100(14) | 0.069(14) | 0.041(20) |
| {1, 5} | -0.767(23) | 0.109(15) | 0.066(15) | 0.033(20) |
| {1,1} | 0.977(21) | 0.106(15) | 0.108(15) | 0.001(21) |
| $g_{\text{KCBS}} = 4.742(101)/6 = 0.790(17)$ | | | | |

randomness depending on g_{KCBS} and test probability q . If $g_{\text{KCBS}} \leq 0.77$, we cannot observe net randomness regardless of q . When $g_{\text{KCBS}} > 0.77$, there exist optimal q values. In our experiment, we choose $q_{\text{exp}} = 10^{-4}$ as shown by the red circle in Fig. 3(b).

We observe for the left-hand side of inequality (2) that $\langle \chi_{\text{KCBS}} \rangle = -4.831$, and for the right-hand side that $-4 - (\epsilon_{12} + \epsilon_{32} + \epsilon_{34} + \epsilon_{54} + \epsilon_{51} + \epsilon_{11}) = -4.088$. The obtained final score of the KCBS game is $g_{\text{KCBS}} = 4.742(101)/6 = 0.790(17)$, which violates inequality (2) by 11 standard deviations. Our test probability is $q_{\text{exp}} = 10^{-4} \sim O[(\log^3 N_{\text{exp}})/N_{\text{exp}}]$, and the required amount of initial random seed is $O(\log^4 N_{\text{exp}})$ bits (see Appendixes G and H for details). The min-entropy of final randomness is 4.4×10^{-3} per bit, which is $\Theta(N_{\text{exp}})$, achieving exponential randomness expansion. In real numbers, we get 5.73×10^5 bits of min-entropy, which exceeds the 2.35×10^5 bits of input randomness, resulting in 3.38×10^5 net random bits, the expansion rate per round being 2.6×10^{-3} .

Meanwhile, we also apply an improved randomness rate referred to as the Huang-Shi (HS) bound to the experimental data and get larger min-entropy and expansion rate. Reducing the security failure probability δ to 10^{-4} , we obtain randomness of 4.1×10^{-3} per round, and the expansion rate is 2.3×10^{-3} . In total we get 2.92×10^5 net random bits. Note that the optimal q for the HS bound is different from that for the Miller-Shi (MS) bound, but our q_{exp} is still good enough to generate net randomness as shown in Fig. 3(d). If we use an optimized q based on

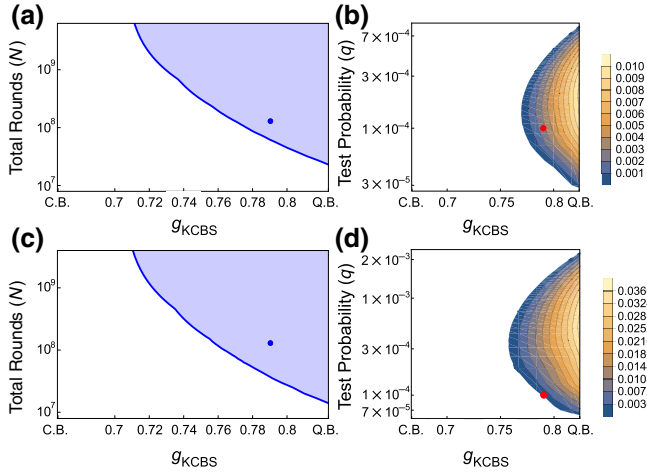


FIG. 3. (a),(b) For the Miller-Shi bound the relation of the score of KCBS game g_{KCBS} , number of total rounds N , test probability q , and randomness expansion rate with smoothing parameter $\delta = 10^{-2}$ in Eq. (A4) of Appendix A. (a) The minimum number of rounds to have net randomness depending on the score g_{KCBS} . The minimum N decreases as g_{KCBS} increases. We can get net randomness only within the shaded area. Our experimental $g_{\text{KCBS}} = 0.790$ and $N_{\text{exp}} = 1.29 \times 10^8$ are shown as the green circle. (b) Randomness expansion rate at different g_{KCBS} and q for our N_{exp} . Only with the combination of large enough g_{KCBS} and appropriate q can we obtain net randomness. Our experimental $g_{\text{KCBS}} = 0.790$ and $q_{\text{exp}} = 0.0001$ are shown as the red circle, the resulting expansion rate being 2.6×10^{-3} per bit. (c),(d) For the Huang-Shi bound the relation of the score of KCBS game g_{KCBS} , number of total rounds N , test probability q , and randomness expansion rate with smoothing parameter $\delta = 10^{-4}$ in Eq. (A4) of Appendix A. (c) The minimum number of rounds to have net randomness depending on the score g_{KCBS} . Our experimental condition is shown as the green circle. (d) Randomness expansion rate at different g_{KCBS} and q for our N_{exp} . Our experimental $g_{\text{KCBS}} = 0.790$ and $q_{\text{exp}} = 0.0001$ are shown as the red circle, the resulting expansion rate being 2.3×10^{-3} per bit, although our q_{exp} is not optimal for this case.

the calculation using the HS bound, we can get even larger min-entropy and expansion rate.

VI. CONCLUSION AND DISCUSSION

In this work, we achieve an exponential randomness expansion secured by quantum contextuality. Regardless of imperfections and experimental noises, the observed violation of the modified KCBS inequality, Eq. (2), verifies the generated randomness. In our protocol, we can guarantee the randomness without the independent identically distributed assumption even when imperfections or noises may originate from quantum mechanics, which would be a quantum adversary.

Because of the advantage of using contextuality for randomness certification, our current generation speed is 270 bits s^{-1} , which is faster than that using Bell's

inequality [9,22]. We believe that we can achieve orders-of-magnitude higher generation speed using several improvements in terms of duration of cooling, optical pumping, and detection, coherence time of qutrit, and coherent operation time (see Appendix I for details). From the theoretical aspect, although the generation rate used in our scheme is robust and noise-tolerable, a large number of trials are still required, needing much effort. An improved generation rate based on general contextuality inequality is still an open problem. Furthermore, the security proof in Ref. [14] only considers the perfect compatible case. Here we characterize the imperfections and modify the score of the KCBS game. We assume the imperfections in experiments do not affect the adversary and security proof in Ref. [14] and only lead to a modified classical bound. The rigorous proof of a self-testing random number generator with limited compatibility is an interesting open problem and we will leave it for future theoretical work.

Moreover, quantum contextuality can also provide an alternative means for randomness amplification. In principle, we can individually manipulate multiple ions and use them to generate random numbers simultaneously, which could lead to orders-of-magnitude faster generation speed. Such kinds of multiple ion systems can be applied to realize a randomness amplification protocol [15], which generates true randomness out of weak randomness input. The protocol can be implemented by the combination of our developed randomness expansion systems and the exclusive-OR of their outputs.

ACKNOWLEDGMENTS

We thank Yaoyun Shi, Carl Miller, Kai-Min Chung, Cupjin Huang, and Xiao Yuan for helpful discussions. This work is supported by National Key Research and Development Program of China under Grants No. 2016YFA0301900, No. 2016YFA0301901, No. 2017YFA0303900, and No. 2017YFA0304004, and by National Natural Science Foundation of China Grants No. 11574002, No. 11674193, No. 11875173, and No. 11970407.

M.U. and Q.Z. contributed equally to this work.

Note added.—Recently, another similar paper [48] based on Bell tests was posted.

APPENDIX A: RANDOMNESS GENERATION RATE

Here, we consider the case that the average probability of measurement setting choice is unbiased, $p(a) = 1/11$, $a \in \{(i, i+1), (i+1, i), (1, 1)\}$ ($i = 1, 2, \dots, 5$). The violation of the inequality in Eq. (2) of the article indicates the presence of genuine quantum randomness in the measurement outcomes. The amount of secure randomness can be quantified by the smooth min-entropy $H_{\text{min}}^{\delta}(X|AE)$, which

is bounded by

$$H_{\min}^{\delta}(X|AE) \geq NR_{\text{gen}}(g_{\text{KCBS}}, q, \epsilon, N, \delta), \quad (\text{A1})$$

where X and A denote the output and input sequences, respectively; E denotes the system of a quantum adversary; δ is the smoothing parameter representing the security failure probability; g_{KCBS} is the KCBS game score; N is the total number of experiment trials; q is the probability of choosing game round; ϵ is the parameter of Schatten norm, in the security analysis, $(1 + \epsilon)$ -Schatten norm is applied; and R_{gen} is the lower bound of randomness generation on average for each trial. In order to achieve the maximal randomness expansion, we also need to consider the input randomness for each trial,

$$R_{\text{In}} = q \log 11 + H(q), \quad (\text{A2})$$

and the randomness expansion rate can be expressed as $R_{\text{exp}} = R_{\text{gen}} - R_{\text{In}}$. The output randomness rate R_{gen} is given by

$$R_{\text{gen}} = \pi(\chi) - \Delta, \quad (\text{A3})$$

where

$$\begin{aligned} \chi &= g_{\text{KCBS}} - \chi_g, \\ \pi(\chi) &= 2 \frac{\log(e)\chi^2}{r-1}, \\ \Delta &= \frac{\epsilon}{q} \frac{8 \log(e)\chi^2}{(r-1)^2} + \frac{\log(2/\delta^2)}{N\epsilon} + 2rq + O\left[\left(\frac{\epsilon}{q}\right)^2\right]. \end{aligned} \quad (\text{A4})$$

Here, all the log terms are base 2 throughout the paper and r is the output alphabet size, which is $r = 4$ in our KCBS game. The explicit form of $O[(\epsilon/q)^2]$ and derivation of (A4) are shown in Appendixes C and D. Denote the above bound as the MS bound [14] and afterwards a tighter bound is obtained, referred to as the HS bound without the dependence of r [49]. For the experiment, we perform the parameter optimization of q and ϵ to achieve the maximal randomness expansion rate R_{exp} with the MS bound and also show the final randomness rate for two different bounds. However, in the real experiment, we need an estimation of g_{KCBS} and N in order to perform optimization. Firstly, we need to roughly estimate a score g_{KCBS} based on a small amount of demonstration data. Then we have to carefully estimate an appropriate total number N , which should be large enough to have enough randomness expansion, but also in a practical time-consuming period. In realization, there exists a difference in both g_{KCBS} and N between the estimated and real experimental value. Thus, $q_{\text{exp}} = 0.0001$ is an optimized probability based on our estimated parameters, which is slightly different from the precisely maximal value for final experimental parameters as shown in Fig. 3(b).

TABLE II. Unitary rotations U_i .

| U | Rotation |
|-------|--|
| U_1 | $R_1(0.531\pi, \pi)R_2(0.066\pi, 0)$ |
| U_2 | $R_1(0.442\pi, 0)R_2(0.328\pi, 0)$ |
| U_3 | $R_1(0.191\pi, \pi)R_2(0.506\pi, \pi)$ |
| U_4 | $R_1(0.104\pi, \pi)R_2(0.526\pi, 0)$ |
| U_5 | $R_1(0.377\pi, 0)R_2(0.404\pi, \pi)$ |

APPENDIX B: UNITARY ROTATIONS

Here, $R_1(\theta_1, \phi_1)$ and $R_2(\theta_2, \phi_2)$ are defined as

$$R_1(\theta_1, \phi_1) = \begin{pmatrix} \cos \frac{\theta_1}{2} & 0 & -ie^{i(\phi_1 + \pi/2)} \sin \frac{\theta_1}{2} \\ 0 & 1 & 0 \\ -ie^{-i(\phi_1 + \pi/2)} \sin \frac{\theta_1}{2} & 0 & \cos \frac{\theta_1}{2} \end{pmatrix},$$

$$R_2(\theta_2, \phi_2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \frac{\theta_2}{2} & -ie^{-i(\phi_2 + \pi/2)} \sin \frac{\theta_2}{2} \\ 0 & -ie^{i(\phi_2 + \pi/2)} \sin \frac{\theta_2}{2} & \cos \frac{\theta_2}{2} \end{pmatrix}.$$

The unitary rotations U_i in the measurement configurations shown in Fig. 1(b) are realized by corresponding $R_2(\theta_{2i}, \phi_{2i})$ then $R_1(\theta_{1i}, \phi_{1i})$, while U_i^\dagger are composed of $R_1(\theta_{1i}, \pi - \phi_{1i})$ then $R_2(\theta_{2i}, \pi - \phi_{2i})$, where the specific U_i are listed in Table II.

APPENDIX C: EXPERIMENTAL SEQUENCE

Each round comprises Doppler cooling, EIT cooling, optical pumping, rotation (U_i), the first projective measurement, inverse rotation (U_i^\dagger), rotation (U_j), the second projective measurement, and inverse rotation (U_j^\dagger). The $^{138}\text{Ba}^+$ ion is first cooled with 500 μs of Doppler cooling and 1000 μs of EIT cooling. Optical pumping initializes the internal state of the ion to $|m_S = +1/2\rangle$ by carefully adjusting the polarization of a 493-nm laser beam. We manipulate the states between $|1\rangle$ and $|3\rangle$, and between $|2\rangle$ and $|3\rangle$ by applying a 1762-nm laser with different frequencies and amplitudes controlled by an AOM. The 1762-nm fiber laser is stabilized with a high-finesse cavity to achieve a linewidth below 1 Hz using the Pound-Drever-Hall technique. The cavity is made of ultralow-expansion material and is mounted in a vacuum cavity with active temperature stabilization to maximize the stability of its length. Frequency and amplitude of the rf signal for AOM

inputs are generated by two independent pairs of direct digital synthesizers (AD9910) for A_i and A_j measurements, which represent Alice and Bob, ensuring they are compatible without communication. The 2π time for both Rabi oscillations is adjusted to $37 \mu\text{s}$, that is $\Omega = (2\pi) 27 \text{ kHz}$. Every rotation U_j is performed with the same duration of no longer than $16 \mu\text{s}$.

EIT cooling implements the asymmetry profile of the absorption spectrum to cancel the heating effect caused by carrier transition meanwhile strengthening the red-sideband transition to hold the cooling function [45–47]. EIT cooling only needs three levels; however, there are four Zeeman states of the $^{138}\text{Ba}^+$ ion. Although with only Doppler cooling and EIT cooling the ion is not perfectly cooled to the ground state without sideband cooling (average phonon number $\langle \bar{n} = 0.1 \rangle$), the carrier transition operated by the stabilized 1762-nm laser has enough fidelity due to the small Lamb-Dicke parameter $\eta = 0.07$.

Our projective measurement includes state discrimination and state re-preparation. We differentiate one state versus the other two states of a qutrit using the standard fluorescence detection method. For the $|3\rangle$ state, an average of 32 photons at 493 nm can be detected during $600 \mu\text{s}$ and no photons for the $|1\rangle$ state or the $|2\rangle$ state. In experiment, perfect state detection fidelity is achieved for $|3\rangle$, while the error for $|1\rangle$ and $|2\rangle$ is 1.3%. The duration of the first projective measurement is set to $600 \mu\text{s}$ with discrimination $n_{\text{ph}} = 3$ while the second projective measurement is $300 \mu\text{s}$ and $n_{\text{ph}} = 1$. Fluorescence detection duration is longer than the coherence time between $|1\rangle$ and $|2\rangle$, which is around $200 \mu\text{s}$. Therefore we add spin echo pulses during the fluorescence detection to keep the coherence until the second measurement is done. Thus, the first detection is divided into the following sequences: $150 \mu\text{s}$ (1/4 of normal detection period) detection, a π -pulse between $|1\rangle$ and $|3\rangle$, a π -pulse between $|2\rangle$ and $|3\rangle$, a π -pulse between $|1\rangle$ and $|3\rangle$, $150 \mu\text{s}$ (1/2 of normal detection period) detection, a π -pulse between $|1\rangle$ and $|3\rangle$, a π -pulse between $|2\rangle$ and $|3\rangle$, a π -pulse between $|1\rangle$ and $|3\rangle$, and $150 \mu\text{s}$ (1/4 of normal detection period) detection. We add up all the collected photons during these three detections ($150 \mu\text{s}$, $300 \mu\text{s}$, $150 \mu\text{s}$) as the first fluorescence detection result. Re-preparation to $|3\rangle$ state takes $30 \mu\text{s}$; it is realized by optical pumping without the 614-nm laser, keeping the coherence between $|1\rangle$ and $|2\rangle$ in the $^5D_{5/2}$ manifold. Since the second projective measurement is the end of the experiment without further operations, we do not apply spin echo pulses and state re-preparation, which results in shorter duration.

APPENDIX D: EXTRACTOR AND RANDOM TEST

A random number extractor is a hashing function transforming a nonperfect random number string $\{0, 1\}^N$ to a nearly perfect one $\{0, 1\}^m$. In our experiment, the length

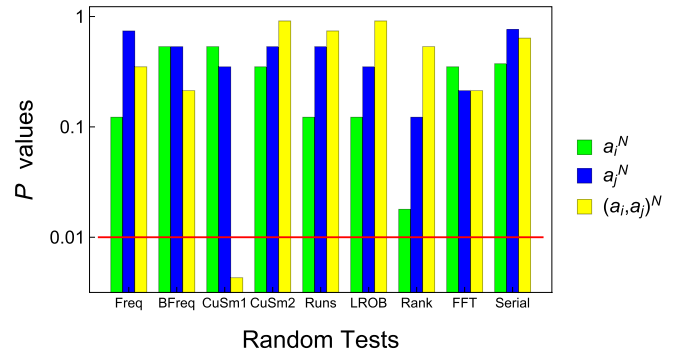


FIG. 4. The results for random tests [53] of the outputs of the first measurement a_i and the second measurement a_j , and both measurement $a_i a_j$. Outputs of a_i^N and a_j^N pass the listed tests since all p values exceed the threshold of 0.01, while the outputs of $(a_i a_j)^N$ fail to pass the first test of “Cumulative Sums (CuSm).”

of the input string is $N_{\text{exp}} = 1.29 \times 10^8$ and $H_{\min}(X|IE) = 6.2 \times 10^{-3}$ per bit. According to leftover hash lemma [50]

$$m \leq NH_{\min}(X|IE) - 2 \log \frac{1}{\epsilon_h}, \quad (\text{D1})$$

we set the security parameter ϵ_h to be a typical value $\epsilon_h = 2^{-100}$, and the length of the output string is $m = 8.06 \times 10^5$. Here we apply a random $m \times N_{\text{exp}}$ Toeplitz matrix [51] as the hashing function. The input random seed $\{0, 1\}^s$ ($s = m + N_{\text{exp}} - 1$) is from Ref. [52].

We apply the random test [53] to the extracted data. The tests include “Frequency,” “Block Frequency (BFreq),” two “Cumulative Sums (CuSm)” tests, “Runs,” “Longest-Run-of-Ones in a Block (LROB),” “Rank,” “Fast Fourier Transform (FFT),” and “Serial.” The p values are distributed in the interval $(0, 1)$, which show the probabilities that an ideal random number generator would produce a less random sequence than the tested one. If the p value is taken as 0, it means the tested data are fully nonrandom, while 1 means completely random. The threshold we set for accepting the data as random is 0.01. As shown in Fig. 4, the output strings a_i^N and a_j^N pass all tests. However, as expected, the combined outputs $(a_i a_j)^N$ do not pass all tests because the measurement outputs of two observables are correlated and thus are not independent random variables.

APPENDIX E: MODIFIED NONCONTEXTUAL INEQUALITY

Among the Kochen-Specker inequalities, the KCBS inequality, which uses five observables A_i taking ± 1 , shows that with noncontextual hidden variables, the left-hand side of the inequality is no less than -3 [26] as shown in Eq. (1). In practice, the observables $\langle A_i A_j \rangle$ have to be implemented in a sequential measurement. We denote the

observable A_i with superscript m , A_i^m , as the measurement of A_i at position m in the sequence. For example, $A_1^1 A_2^2$ denotes the sequence of measuring A_1 first, then A_2 .

Noncontextual HV model requires that the outcomes of any observable A_i do not depend on other compatible jointly measured observables with A_i . To be more specific, we take A_1 as an example. It is compatible with A_2 and A_5 . We denote the obtained value as v , then have $v(A_1^1) = v(A_1^2 | A_2^1 A_2^2)$ and $v(A_1^1) = v(A_1^2 | A_5^1 A_5^2)$.

The assumption behind the above contextuality inequality is that the observables A_i and A_{i+1} (let $A_6 \equiv A_1$) are compatible. However, in an actual experiment using sequential measurements, the compatibility is not perfect, which leads to the compatibility loophole.

In Ref. [29], this imperfection can be quantified by

$$p^{\text{flip}}[A_1 A_2] = p[(A_2^1(+)|A_2^1) \text{ and } (A_2^2(-)|A_1^1 A_2^2)] \\ + p[(A_2^1(-)|A_2^1) \text{ and } (A_2^2(+)|A_1^1 A_2^2)]. \quad (\text{E1})$$

Here $+$, $-$ denote the obtained value and this probability can be understood as A_1 flips the predetermined value of A_2 . Then using the fact that $\langle A_1 A_2 \rangle \leq \langle A_1^1 A_2^2 \rangle + 2p^{\text{flip}}[A_1 A_2]$, the inequality can be modified as

$$\langle A_1^1 A_2^2 \rangle + \langle A_3^1 A_2^2 \rangle + \langle A_3^1 A_4^2 \rangle + \langle A_5^1 A_4^2 \rangle + \langle A_5^1 A_2^2 \rangle \\ \geq -3 - 2(p^{\text{flip}}[A_1 A_2] + p^{\text{flip}}[A_3 A_2] + p^{\text{flip}}[A_3 A_4] \\ + p^{\text{flip}}[A_5 A_4] + p^{\text{flip}}[A_5 A_1]). \quad (\text{E2})$$

Note that this inequality holds for any HV models. In the experiment, p^{flip} is not achievable and different approaches are proposed for estimating with different assumptions. Here we use $\epsilon_{ij} = |\langle A_j | A_j A_i \rangle - \langle A_j | A_i A_j \rangle|$ to quantify the difference between a same pair of observables A_i and A_j in different time order, $A_i A_j$ and $A_j A_i$, which can be regarded as the bound of incompatibility of these sequential measurements.

For experimentally accessible distributions,

$$|p(A_i = a | A_i A_{i+1}) - p(A_i = a | A_{i+1} A_i)| \leq \epsilon_{ij}/2, \quad (\text{E3})$$

where $a \in \{+, -\}$. We assume that the underlying probability distributions have the same properties as all accessible distributions. Then $p^{\text{flip}}[A_1 A_2]$ can be bounded by $\epsilon_{12}/2$, which is obtained in the experiments: $p^{\text{flip}}[A_1 A_2] \leq \epsilon_{12}/2$. However, the probability distributions of a general HV model may not belong to the set of experimentally accessible probability distributions. We assume that this difference is negligible and that the properties verified in accessible experiments hold also for some HV models.

Combining another modification in Ref. [30], we apply an extended version of KCBS inequality (2), where for

simplicity we omit the time order superscript and $\langle A_i A_j \rangle$ denotes the expectation value of the measurement results in the time order of $A_i A_j$ for the sequential measurements.

APPENDIX F: MILLER AND SHI'S SECURITY PROOF AND ITS FEASIBILITY IN PRACTICAL CASES

In this appendix, we mainly focus on the work of Miller and Shi [14] and overview their security proof.

The min-entropy is used for evaluating the randomness. Given the output X , conditioned on input A and adversary system E , the smooth min-entropy $H_{\min}^{\delta}(X|AE)$ is defined as

$$H_{\min}^{\delta}(X|AE) = \max_{\|\Gamma' - \Gamma_{AEX}\| \leq \delta} H_{\min}(X|AE)_{\Gamma'}. \quad (\text{F1})$$

The direct estimation of min-entropy is generally difficult; thus the MS security proof applied Renyi entropy to give the lower bound of min-entropy. For a quantum state ρ , its smooth min-entropies satisfy

$$H_{\min}^{\delta}(\rho) = H_{1+\epsilon}(\rho) - \frac{\log(1/\delta)}{\epsilon}, \quad (\text{F2})$$

where $H_{1+\epsilon}(\rho) = -(1/\epsilon) \log \text{Tr}[\rho^{1+\epsilon}]$. The randomness in its output is quantified by this $(1+\epsilon)$ -randomness. The main tool proposed in this proof is a $(1+\epsilon)$ -uncertain relation. After a projective measurement, the amount of randomness [$(1+\epsilon)$ -randomness] obtained from a measurement is related to the degree of disturbance caused by the measurement, shown in Proposition 4.4. For a given fixed input, the device has a classically predictable output and the achievable maximal score is w . Then if the device obtains a score higher than this threshold w , there must be unpredictable randomness in the output of this device. The rate curve is achieved in Corollary 6.11. This security proof is general not only for nonlocal games but also for contextuality. The uncertain relation is only relevant to the size of output alphabet and the measurement in contextuality can fit this proposition. For different schemes, the major difference is the classically predictable bound w . Note that this bound w is the maximal score for devices that have classically predictable outputs on an input. It is different from the classical strategy bound by hidden variable C_G in general. Although different in the definition, the value can be the same for some specific cases; for example, nonlocal game with binary input in each party and contextuality shown in Appendix D of Ref. [14]. However, in practical cases, the measurements in contextuality are not compatible. Although the uncertain relation in Proposition 4.4 still holds, the remaining problem is to calculate w and check whether it equals the classical bound achieved by an approximately contextual hidden variable. We express this

KCBS game as

$$G(A_1, A_2, A_3, A_4, A_5) = -\frac{1}{6}(A_1^1 A_2^2 + A_3^1 A_2^2 + A_3^1 A_4^2 + A_5^1 A_4^2 + A_5^1 A_1^2 - A_1^1 A_1^2 + \epsilon_{12} + \epsilon_{32} + \epsilon_{34} + \epsilon_{54} + \epsilon_{51} + \epsilon_{11}). \quad (\text{F3})$$

Proposition 1. *Let G be the game given above, $w = 2/3$.*

Proof: With the approximately noncontextual hidden variable, the maximal score is $C_G = 2/3$. This strategy is classically predictable; thus the maximal score w with a classically predictable input should not be less than C_G , i.e., $w \geq C_G$. We suppose that there is a device D (can be quantum) applied in the KCBS game that outputs a score above $2/3$, and that gives a deterministic output on input 1:

$$-4 \geq \langle \chi_{\text{KCBS}} \rangle, \quad (\text{F4})$$

where $\langle \chi_{\text{KCBS}} \rangle = \langle A_1^1 A_2^2 \rangle + \langle A_3^1 A_2^2 \rangle + \langle A_3^1 A_4^2 \rangle + \langle A_5^1 A_4^2 \rangle + \langle A_5^1 A_1^2 \rangle - \langle A_1^1 A_1^2 \rangle + \epsilon_{12} + \epsilon_{32} + \epsilon_{34} + \epsilon_{54} + \epsilon_{51} + \epsilon_{11}$ is the practical mean value with sequential measurements. Because $\langle A_i A_j \rangle \geq -1 + |\langle A_i \rangle + \langle A_j \rangle|$, $\langle A_i A_j \rangle \leq \langle A_i^1 A_j^2 \rangle + 2p^{\text{flip}}[A_i A_j]$, and $p^{\text{flip}}[A_i A_j] \leq \epsilon_{ij}$, we have

$$\begin{aligned} \langle \chi_{\text{KCBS}} \rangle &\geq -6 + |\langle A_1 \rangle + \langle A_2 \rangle| + |\langle A_3 \rangle + \langle A_2 \rangle| \\ &\quad + |\langle A_3 \rangle + \langle A_4 \rangle| + |\langle A_5 \rangle + \langle A_4 \rangle| + |\langle A_5 \rangle + \langle A_1 \rangle| \\ &\geq -6 + |\langle A_1 \rangle + \langle A_2 \rangle| + |-\langle A_2 \rangle - \langle A_3 \rangle| \\ &\quad + |\langle A_3 \rangle - \langle -A_4 \rangle| + |-\langle A_4 \rangle - \langle A_5 \rangle| \\ &\quad + |\langle A_5 \rangle - \langle -A_1 \rangle|. \end{aligned} \quad (\text{F5})$$

Therefore, with the triangle inequality,

$$-4 \geq -6 + |\langle A_1 \rangle - \langle -A_1 \rangle|. \quad (\text{F6})$$

The fixed input 1 is deterministic, and thus $\langle A_1 \rangle = \pm 1$; this is a contradiction. Thus $w \leq C_G = 2/3$ and $w = 2/3$. ■

With this proposition, any score above w can be used to generate randomness, although the observables are approximately compatible.

APPENDIX G: RANDOMNESS GENERATION RATE

In this appendix, based on the work of Miller and Shi [14], we give an exact result for the randomness expansion rate. The min-entropy is used for evaluating the randomness. Combining Theorem 4.1 and Proposition 6.8

in Ref. [14] yields

$$H_{\min}^{\delta}(X|AE) \geq N \left\{ \pi(\chi) - O \left[q + \epsilon/q + \frac{\log(2/\delta^2)}{N\epsilon} \right] \right\}, \quad (\text{G1})$$

where $O[\log(2/\delta^2)/N\epsilon]$ and $O(q + \epsilon/q)$ come from Theorem 3.2 and Proposition 6.8, respectively. From Theorem 3.2, we can let $O[\log(2/\delta^2)/N\epsilon] = \log(2/\delta^2)/N\epsilon$. $O(q + \epsilon/q)$ comes from Proposition 6.5, a combination of Propositions 6.3 and 6.4. In the proof of Proposition 6.4, from Eqs. (6.25) to (6.26) it is equivalent to

$$\frac{\sum_x \langle \rho_{\bar{a}}^x \rangle_{1+\epsilon}}{\langle \rho \rangle_{1+\epsilon}} \geq 1 - O(\epsilon), \quad (\text{G2})$$

where x is the output with output alphabet size r and \bar{a} is the input. According to Propositions B.2 and B.3 in Ref. [14], we apply the induction $\sum_x \langle \rho_{\bar{a}}^x \rangle_{1+\epsilon} \geq (1-\epsilon)^r \langle \sum_x \rho_{\bar{a}}^x \rangle_{1+\epsilon}$ and $\langle \sum_x \rho_{\bar{a}}^x \rangle_{1+\epsilon} \geq (1-\epsilon)^r \langle \rho \rangle_{1+\epsilon}$. Thus $\sum_x \langle \rho_{\bar{a}}^x \rangle_{1+\epsilon} / \langle \rho \rangle_{1+\epsilon} \geq (1-\epsilon)^{2r} \geq 1 - 2r\epsilon$ and $O(\epsilon) = 2r\epsilon$. Consequently, the term in Proposition 6.4 $O(q) = 2rq$.

The estimation in Proposition 6.3 comes from the second-order terms in the Taylor expansion in Eqs. (6.20) and (6.21). For a function $F(x)$, its Taylor expansion at a is as follows:

$$\begin{aligned} F(b) &= F(a) + F'(a)(b-a) + \frac{F''(a)}{2}(b-a)^2 \\ &\quad + \frac{F'''[a + \theta(b-a)]}{6}(b-a)^3, \theta \in (0, 1), \end{aligned} \quad (\text{G3})$$

where the fourth term is the third-order Taylor-Lagrange remainder. Here $F(b) = 2^{\epsilon s H(a,x)/q}$ and $a = 0$:

$$\begin{aligned} &2^{\epsilon s H(a,x)/q} - 1 \\ &= \epsilon s (\ln 2) H(a,x)/q + \frac{1}{2} \left[\frac{\epsilon s (\ln 2) H(a,x)}{q} \right]^2 + R_3, \\ R_3 &= \frac{1}{6} \left[\frac{\epsilon s (\ln 2) H(a,x)}{q} \right]^3 2^{\theta \epsilon s H(a,x)/q}, \theta \in (0, 1), \end{aligned} \quad (\text{G4})$$

where the term R_3 is the third-order Taylor-Lagrange remainder. Substituting this expression in Eq. (6.20), we

have

$$\begin{aligned}
& \sum_{a,x} p(a) \left\{ \frac{1}{2} \left[\frac{\epsilon s (\ln 2) H(a,x)}{q} \right]^2 + R_3 \right\} \langle \rho_a^x \rangle_{1+\epsilon} \\
& \leq \left\{ \frac{1}{2} \left[\frac{\epsilon s (\ln 2)}{q} \right]^2 + \frac{1}{6} \left[\frac{\epsilon s (\ln 2)}{q} \right]^3 2^{\epsilon s/q} \right\} \\
& \times \sum_{a,x} p(a) H(a,x) \langle \rho_a^x \rangle_{1+\epsilon} \\
& \leq \frac{1}{2} \left[\frac{\epsilon s (\ln 2)}{q} \right]^2 + \frac{1}{6} \left[\frac{\epsilon s (\ln 2)}{q} \right]^3 2^{\epsilon s/q}. \quad (\text{G5})
\end{aligned}$$

After applying the function $-(1/\epsilon)\log()$, we have a more precise result similar to Proposition 6.3. The difference is that we replace $O(\epsilon/q)$ by $(\epsilon/q)[(\ln 2)s^2/2] + (\epsilon/q)^2[(\ln 2)^2 s^3/6]2^{\epsilon s/q}$. In Theorem 6.7, we let the parameter s be $\pi'(\chi)$. In Theorem 5.8, we know that

$$\begin{aligned}
\pi(\chi) &= 2 \frac{\log(e)(\chi - w)^2}{r - 1}, \\
\pi'(\chi) &= 4 \frac{\log(e)(\chi - w)}{r - 1}. \quad (\text{G6})
\end{aligned}$$

Thus

$$\begin{aligned}
O(\epsilon/q) &= \frac{\epsilon}{q} \frac{8 \log(e)(\chi - w)^2}{(r - 1)^2} \\
&+ \left(\frac{\epsilon}{q} \right)^2 \frac{32 \log(e)(\chi - w)^3}{3(r - 1)^3} 2^{\epsilon 4[\log(e)(\chi - w)/(r - 1)q]}. \quad (\text{G7})
\end{aligned}$$

Result 1

$$\begin{aligned}
H_{\min}^{\delta}(X|AE) &\geq N[\pi(\chi) - \Delta], \\
\pi(\chi) &= 2 \frac{\log(e)(\chi - w)^2}{r - 1}, \\
\Delta &= \frac{\epsilon}{q} \frac{8 \log(e)(\chi - w)^2}{(r - 1)^2} + \left(\frac{\epsilon}{q} \right)^2 \frac{32 \log(e)(\chi - w)^3}{3(r - 1)^3} \\
&\times 2^{(\epsilon/q)[4 \log(e)(\chi - w)/(r - 1)]} + \frac{\log(2/\delta^2)}{N\epsilon} + 2rq, \quad (\text{G8})
\end{aligned}$$

where $\chi \in [0, 1]$ is the score obtained in experiments, w is the classical bound for a certain game, r is the number of total outputs, q is the probability for test round, N is the total round number, δ is the failure probability, and $\epsilon \in (0, 1]$. The randomness expansion, generation, and input

rate per round are

$$\begin{aligned}
R_{\text{exp}} &= R_{\text{gen}} - R_{\text{In}}, \\
R_{\text{gen}} &= \pi(\chi) - \Delta, \\
R_{\text{In}} &= q \log 11 + H(q). \quad (\text{G9})
\end{aligned}$$

If we focus on the randomness expansion instead of the generation randomness, we should consider the random seed $H(q) + q \log 11$ used for random inputs. Different target functions have different optimal results; the figures in the main text show the effect of optimization parameter. Note that from Result 1, the generated randomness is $O(N)$, and we take the probability $q \sim (\log^3 N)/N$. Then the initial random seed required is $q \log 11 + H(q)$. And because $\log N < N$, $q \log 11 + H(q) \sim O(q) + q \log [(\log^3 N)/N] < O(\log^4 N)$. Thus compared with the generated randomness $O(N)$, exponential randomness expansion is achieved.

APPENDIX H: IMPROVED RATE CURVE

The important uncertain relation is related to the output alphabet size r . A larger r will lead to a poor performance. This disadvantage is removed by an improved uncertain relation. A tighter bound of Proposition 4.4 proposed in Ref. [49] is as follows.

Lemma 1. *For any finite-dimensional Hilbert space V , any positive semidefinite operator $\tau : V \rightarrow V$, and any projective measurement $\{P_0, P_1, \dots, P_n\}$ on V , the following holds. Let $\tau' = \sum_i P_i \tau P_i$. Then*

$$\|\tau'\|_{1+\epsilon}^2 \leq \|\tau\|_{1+\epsilon}^2 - \epsilon \|\tau - \tau'\|_{1+\epsilon}^2 \quad (\text{H1})$$

for all $\epsilon \in (0, 1)$. Consequently,

$$\|\tau'\|_{1+\epsilon} \leq \|\tau\|_{1+\epsilon} - \epsilon/2 \|\tau - \tau'\|_{1+\epsilon}. \quad (\text{H2})$$

This result can be applied in Theorem 5.8 to obtain a new rate curve:

$$\pi(\chi) = 2 \log(e)(\chi - w)^2 \quad \text{if } \chi \geq w. \quad (\text{H3})$$

Consequently, we have $\pi'(\chi) = 4 \log(e)(\chi - w)$, and we let the parameter s be $\pi'(\chi)$ in $O(\epsilon/q)$ by $(\epsilon/q)(\ln 2)s^2/2 + (\epsilon/q)^2[(\ln 2)^2 s^3/6]2^{\epsilon s/q}$. Then

$$\begin{aligned}
O(\epsilon/q) &= \frac{\epsilon}{q} 8 \log(e)(\chi - w)^2 \\
&+ \left(\frac{\epsilon}{q} \right)^2 \frac{32 \log(e)(\chi - w)^3}{3} 2^{(\epsilon/q)4 \log(e)(\chi - w)}. \quad (\text{H4})
\end{aligned}$$

Result 2

$$\begin{aligned}
H_{\min}^{\delta}(X|AE) &\geq N[\pi(\chi) - \Delta], \\
\pi(\chi) &= 2 \log(e)(\chi - w)^2, \\
\Delta &= \frac{\epsilon}{q} 8 \log(e)(\chi - w)^2 \\
&\quad + \left(\frac{\epsilon}{q}\right)^2 \frac{32 \log(e)(\chi - w)^3}{3} 2^{\epsilon 4[\log(e)(\chi - w)/q]} \\
&\quad + \frac{\log(2/\delta^2)}{N\epsilon} + 2rq. \tag{H5}
\end{aligned}$$

APPENDIX I: IMPROVEMENT OF RANDOM NUMBER GENERATION SPEED

Currently, each round takes 3700 μs , which consists of a 1500- μs cooling process, two detection procedures of 900 μs in total, 140- μs spin echo pulses for the first detection, two optical pumping pulses of 60 μs in total, rotations of 60 μs in total, some short gaps between sequences to make sure they do not affect each other, and a communication time of around 1000 μs . However, there is room for technical improvement as follows. By extending the coherence time between qutrit, spin echo will not be required. Detection time could be reduced to around 100 μs by replacing a high numerical aperture lens from 0.2 to 0.6. By amplifying ten times the power of the 1762- μm laser, Rabi oscillations between $|1\rangle$ and $|3\rangle$, and between $|2\rangle$ and $|3\rangle$ can be at least three times faster, and hence also the rotation. Each optical pumping could be reduced to 1 μs by further optimization. Currently we apply a 1500- μs cooling process in each round, but it will be possible to apply only one cooling process per ten rounds after some improvements. With all the developments mentioned above, we can achieve a generation speed that is at least one order of magnitude faster.

-
- [1] P. D. Coddington, Analysis of random number generators using Monte Carlo simulation, Northeast Parallel Architecture Center, Paper 14 (1994).
 - [2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, *Rev. Mod. Phys.* **74**, 145 (2002).
 - [3] M. Fiorentino, C. Santori, S. M. Spillane, R. G. Beausoleil, and W. J. Munro, Secure self-calibrating quantum random-bit generator, *Phys. Rev. A* **75**, 032334 (2007).
 - [4] Oded Goldreich, *Foundations of Cryptography* (Cambridge University Press, Cambridge, UK, 2007).
 - [5] Xiongfeng Ma, Xiao Yuan, Zhu Cao, and Bing Qi, and Zhen Zhang, Quantum random number generation, *Npj Quantum Inf.* **2**, 16021 (2016).
 - [6] Miguel Herrero-Collantes and Juan Carlos Garcia-Escartin, Quantum random number generators, *Rev. Mod. Phys.* **89**, 015004 (2017).

- [7] Yang Liu *et al.*, High-Speed Device-Independent Quantum Random Number Generation without a Detection Loophole, *Phys. Rev. Lett.* **120**, 010503 (2018).
- [8] Roger Colbeck, Ph.D. thesis, University of Cambridge, 2007.
- [9] S. Pironio, A. Acin, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Random numbers certified by Bell's theorem, *Nature* **464**, 1021 (2010).
- [10] Roger Colbeck and Adrian Kent, Private randomness expansion with untrusted devices, *J. Phys. A: Math. Theor.* **44**, 095305 (2011).
- [11] Umesh Vazirani and Thomas Vidick, Certifiable quantum dice, *Philos. Trans. R. Soc. A* **370**, 3432 (2012).
- [12] Stefano Pironio and Serge Massar, Security of practical private randomness generation, *Phys. Rev. A* **87**, 012336 (2013).
- [13] Matthew Coudron, Thomas Vidick, and Henry Yuen, in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (Springer, Berkeley, CA, 2013), p. 468.
- [14] Carl A. Miller and Yaoyun Shi, Universal security for randomness expansion from the spot-checking protocol, *SIAM J. Comput.* **46**, 1304 (2017).
- [15] Kai Min Chung, Yaoyun Shi, and Xiaodi Wu, [arXiv:1402.4797](https://arxiv.org/abs/1402.4797) (2014).
- [16] Antonio Acin and Lluís Masanes, Certified randomness in quantum physics, *Nature* **540**, 213 (2016).
- [17] Rotem Arnon-Friedman, Renato Renner, and Thomas Vidick, Simple and tight device-independent security proofs, *SIAM J. Comput.* **48**, 181 (2019).
- [18] John S. Bell, On the Einstein-Podolsky-Rosen paradox, *Phys. Physique Fizika* **1**, 195 (1964).
- [19] Bas Hensen, Hannes Bernien, Anaïs E. Dréau, Andreas Reiserer, Norbert Kalb, Machiel S. Blok, Just Ruitenberg, Raymond F. L. Vermeulen, Raymond N. Schouten, Carlos Abellán *et al.*, Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, *Nature* **526**, 682 (2015).
- [20] Lynden K. Shalm *et al.*, Strong Loophole-Free Test of Local Realism, *Phys. Rev. Lett.* **115**, 250402 (2015).
- [21] Marissa Giustina *et al.*, Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons, *Phys. Rev. Lett.* **115**, 250401 (2015).
- [22] Peter Bierhorst, Emanuel Knill, Scott Glancy, Yanbao Zhang, Alan Mink, Stephen Jordan, Andrea Rommal, Yi-Kai Liu, Bradley Christensen, Sae Woo Nam *et al.*, Experimentally generated randomness certified by the impossibility of superluminal signals, *Nature* **556**, 223 (2018).
- [23] Yang Liu, Qi Zhao, Ming-Han Li, Jian-Yu Guan, Yanbao Zhang, Bing Bai, Weijun Zhang, Wen-Zhao Liu, Cheng Wu Xiao Yuan *et al.*, Device-independent quantum randomness generation, *Nature* **562**, 548 (2018).
- [24] J. S. Bell, On the problem of hidden variables in quantum mechanics, *Rev. Mod. Phys.* **38**, 447 (1966).
- [25] S. Kochen and E. P. Specker, The problem of hidden variables in quantum mechanics, *J. Math. Mech.* **17**, 59 (1967).

- [26] Alexander A. Klyachko, M. Ali Can, Sinem Binicioğlu, and Alexander S. Shumovsky, Simple Test for Hidden Variables in Spin-1 Systems, *Phys. Rev. Lett.* **101**, 020403 (2008).
- [27] D. L. Deng, C. Zu, X. Y. Chang, P. Y. Hou, H. X. Yang, Y. X. Wang, and L. M. Duan, arXiv: 1301.5364 (2013).
- [28] Mark Um, Xiang Zhang, Junhua Zhang, Ye Wang, Shen Yangchao, D.-L. Deng, Lu-Ming Duan, and Kihwan Kim, Experimental certification of random numbers via quantum contextuality, *Sci. Rep.* **3**, 1627 (2013).
- [29] Otfried Gühne, Matthias Kleinmann, Adán Cabello, Jan-Åke Larsson, Gerhard Kirchmair, Florian Zähringer, Rene Gerritsma, and Christian F. Roos, Compatibility and non-contextuality for sequential measurements, *Phys. Rev. A* **81**, 022121 (2010).
- [30] Jochen Szangolies, Matthias Kleinmann, and Otfried Gühne, Tests against noncontextual models with measurement disturbances, *Phys. Rev. A* **87**, 050101 (2013).
- [31] Tommaso Lunghi, Jonatan Bohr Brask, Charles Ci Wen Lim, Quentin Lavigne, Joseph Bowles, Anthony Martin, Hugo Zbinden, and Nicolas Brunner, Self-Testing Quantum Random Number Generator, *Phys. Rev. Lett.* **114**, 150501 (2015).
- [32] A. Cabello, Experimentally Testable State-Independent Quantum Contextuality, *Phys. Rev. Lett.* **101**, 210401 (2008).
- [33] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, State-independent experimental test of quantum contextuality, *Nature* **460**, 494 (2009).
- [34] Xiang Zhang, Mark Um, Junhua Zhang, Shuoming An, Ye Wang, Dong-ling Deng, Chao Shen, Lu-Ming Duan, and Kihwan Kim, State-Independent Experimental Tests of Quantum Contextuality in a Three Dimensional System, *Phys. Rev. Lett.* **110**, 070401 (2013).
- [35] M. Malinowski, C. Zhang, F. M. Leupold, A. Cabello, J. Alonso, and J. P. Home, Probing the limits of correlations in an indivisible quantum system, *Phys. Rev. A* **98**, 050102 (2018).
- [36] Radek Lapkiewicz, Peizhe Li, Christoph Schaeff, Nathan K. Langford, Sven Ramelow, Marcin Wieśniak, and Anton Zeilinger, Experimental non-classicality of an indivisible quantum system, *Nature* **474**, 490 (2011).
- [37] Ya Xiao, Zhen-Peng Xu, Qiang Li, Jin-Shi Xu, Kai Sun, Jin-Ming Cui, Zong-Quan Zhou, Hong-Yi Su, Adán Cabello, Jing-Ling Chen *et al.*, Experimental observation of quantum state-independent contextuality under no-signaling conditions, *Opt. Express* **26**, 32 (2018).
- [38] Markus Jerger, Yarema Reshitnyk, Markus Oppliger, Anton Potočník, Mintu Mondal, Andreas Wallraff, Kenneth Goodenough, Stephanie Wehner, Kristinn Juliusson, Nathan K. Langford, and Arkady Fedorov, Contextuality without non-locality in a superconducting quantum system, *Nat. Commun.* **7**, 12930 (2016).
- [39] Giuseppe Vallone, Davide G. Marangon, Marco Tomasin, and Paolo Villoresi, Quantum randomness certified by the uncertainty principle, *Phys. Rev. A* **90**, 052327 (2014).
- [40] Zhu Cao, Hongyi Zhou, Xiao Yuan, and Xiongfeng Ma, Source-Independent Quantum Random Number Generation, *Phys. Rev. X* **6**, 011020 (2016).
- [41] Davide G. Marangon, Giuseppe Vallone, and Paolo Villoresi, Source-Device-Independent Ultrafast Quantum Random Number Generation, *Phys. Rev. Lett.* **118**, 060503 (2017).
- [42] Marco Avesani, Davide G. Marangon, Giuseppe Vallone, and Paolo Villoresi, Source-device-independent heterodyne-based quantum random number generator at 17 Gbps, *Nat. Commun.* **9**, 5365 (2018).
- [43] M. R. Dietrich, N. Kurz, T. Noel, G. Shu, and B. B. Blinov, Hyperfine and optical barium ion qubits, *Phys. Rev. A* **81**, 052328 (2010).
- [44] L. Slodička, G. Hétet, N. Röck, S. Gerber, P. Schindler, M. Kumph, M. Hennrich, and R. Blatt, Interferometric thermometry of a single sub-doppler-cooled atom, *Phys. Rev. A* **85**, 043401 (2012).
- [45] Giovanna Morigi, Jürgen Eschner, and Christoph H. Keitel, Ground State Laser Cooling Using Electromagnetically Induced Transparency, *Phys. Rev. Lett.* **85**, 4458 (2000).
- [46] Y. Lin, J. P. Gaebler, T. R. Tan, R. Bowler, J. D. Jost, D. Leibfried, and D. J. Wineland, Sympathetic Electromagnetically-Induced-Transparency Laser Cooling of Motional Modes in an Ion Chain, *Phys. Rev. Lett.* **110**, 153002 (2013).
- [47] Regina Lechner, Christine Maier, Cornelius Hempel, Petar Jurecic, Ben P. Lanyon, Thomas Monz, Michael Brownnutt, Rainer Blatt, and Christian F. Roos, Electromagnetically-induced-transparency ground-state cooling of long ion strings, *Phys. Rev. A* **93**, 053401 (2016).
- [48] Ming-Han Li, Xingjian Zhang, Wen-Zhao Liu, Si-Ran Zhao, Bing Bai, Yang Liu, Qi Zhao, Yuxiang Peng, Jun Zhang, Xiongfeng Ma *et al.*, arXiv:1902.07529 (2019).
- [49] Cupjin Huang and Yaoyun Shi, Private communications (2017).
- [50] Impagliazzo Russell, Levin Leonid A, and Luby Michael, in *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, 1989), p. 12.
- [51] Mark N. Wegman and J. Lawrence Carter, New hash functions and their use in authentication and set equality, *J. Comput. Syst. Sci.* **22**, 265 (1981).
- [52] You-Qi Nie, Leilei Huang, Yang Liu, Frank Payne, Jun Zhang, and Jian-Wei Pan, The generation of 68 Gbps quantum random number by measuring laser phase fluctuations, *Rev. Sci. Instrum.* **86**, 063105 (2015).
- [53] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, and S. Vo, A statistical test suite for random and pseudorandom number generators for cryptographic applications, NIST special publication 800-22, Rev. 1-a (2010).