Device-independent certification of an elementary quantum network link

Jean-Daniel Bancal, 1 Kai Redeker, 2 Pavel Sekatski, 1 Wenjamin Rosenfeld, 3, 2 and Nicolas Sangouard 1

1 Quantum Optics Theory Group, Universität Basel, CH-4056 Basel, Switzerland
2 Department für Physik, Ludwig-Maximilians-Universität, 80797 München, Germany
3 Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, 85748 Garching, Germany

(Dated: December 24, 2018)

Device-independent characterization, also known as self-testing, provides a certification of devices from the result of a Bell test that is suitable for a wide range of applications. We here show that self-testing can be used in an elementary link of a quantum network to certify the successful distribution of a Bell state over 398 meters. Being based on a Bell test free of detection and locality loopholes, our certification is fully device-independent, that is, it does not rely on a knowledge of how the devices work. This guarantees that our link can be integrated in a quantum network for performing long-distance quantum communications with security guarantees that are independent of the details of the actual implementation.

Introduction—The distribution of entanglement over long distances is a key challenge in extending the range of quantum communication [1]. The direct transmission of entangled states through optical fibers is a viable solution for short distances but is limited by transmission loss. Quantum repeaters have thus been proposed for entanglement distribution over long distances [2–4]. The basic idea is to divide the global distance into short elementary links. Entanglement is created in each link independently and successive entanglement swapping operations are used to combine links and extend entanglement. Impressive progress along this line now allows one to envision multipartite quantum networks [5–10] where quantum keys are distributed between arbitrary two parties with unprecedented security guarantees.

At the heart of quantum networks [11] lies the ability to distribute and certify an entangled state between two distant locations. Although entangled states have been produced between remote locations forming an elementary network link in multiple experiments [12–19], their suitability for general purposes including – but not limited to – quantum key distribution, remains unproven. Demonstrations based on the qubit assumption for instance, stating that all elements involved are of dimension two, are subject to side-channels which completely corrupt security guarantees [20].

More generally, the identification of a quantum state provides the most complete description of a system. But the trace left by a state in the measurement outcomes is as much influenced by the state as by the measurement itself. Consequently, it is extremely challenging to obtain an accurate description of a quantum state from observed statistics without presuming a detailed description of the measurement apparatus. Yet, characterizing univocally a quantum resource by identifying its quantum state constitutes a crucial step to set quantum technologies on a solid stand.

The recent advent [19, 21–23] of loophole-free Bell tests [24] opened a new perspective for certifying quantum states without assumption on the dimension of the Hilbert space and on the correct calibration of the measurements [25]. The possibility of device-independent state characterization was first realized in Ref. [26, 27], where it was noted that the only quantum states able to achieve a maximum violation of the Bell-CHSH inequality [28], are Bell states – two-qubit maximally entangled states. This property was later coined the term of self-testing [29]. Since then, numerous additional theoretical self-testing results have been obtained, addressing an increasing range of states, and with improving tolerance to noise [30–39]. Moreover, self-testing has also been extended to the characterization of quantum measurements and channels [29, 30–34]. In the case of Bell states, it is now known that self-testing based on the Bell-CHSH inequality is strongly resistant to noise [32–34] which led recently to the first experimental self-testing certifying a singlet made with two ions separated by about 340 µm within a trap [36].

We here report on a self-testing result certifying device-independently the distribution of an entangled state between two locations separated by 398 meters. In our experiment, each location holds a single neutral atom that can be excited to create a single photon such that the atomic spin and the photon polarization are entangled. The photons are sent to a central station where a probabilistic Bell state measurement is performed. A successful Bell state measurement results in the heralded creation of entanglement of the atomic spins. Fast and efficient measurements of the atomic spin states allow us to perform a loophole-free Bell-CHSH test. From the observed Bell-CHSH value only, we conclude about the successful distribution of an entangled state with a Bell states fidelity of 55.54% at a confidence level of 99%. This constitutes the first result where a statistically relevant bound on the average fidelity of the distributed state is obtained directly from
the Bell-CHSH value and the first device-independent certification of an elementary link of a quantum network.

Assumptions— The scenario we consider involves three protagonists, colloquially referred to as Alice, Bob and Charlie, see Fig. 1. Charlie holds a preparation device which indicates when the experiment is ready: it heralds the start of every measurement procedure. The two other parties each hold one measurement device and one random number generator device. Upon heralding, the random number generators are used by Alice and Bob to choose a measurement setting which is applied to their measurement devices. Measurement settings and outcomes are recorded locally for later analysis. The claim of self-testing for the state measured is based on a number of assumptions that we review now.

1. The experiment admits a quantum description. Essentially, the state of the system can be represented in terms of a density operator, and the measurements as operators acting on the same Hilbert space with the appropriate tensor structure.

2. All devices mentioned above are well identified in space and operate sequentially in time. In particular, the separation between the parties Alice and Bob is clear, as well as between the random number generators and the measurement devices of each party. Moreover, results are recorded before going to the next round, hence we know exactly when a round is going on (two rounds don’t happen simultaneously), when it is finished, and we can monitor how many rounds happened in a given time.

3. The random number generators are independent from all other devices and sample from a well characterized probability distribution. Hence, the questions used are chosen freely: the measured particles cannot influence this choice, nor vice versa. The random number devices can be correlated to each other, but not to the rest of the setup.

4. Finally, the classical and quantum communication between Alice and Bob is limited: no communication (whether direct or indirect) is allowed between the measurement boxes once the settings choices are received and until the measurement outcomes are produced. Moreover, the random number generators only provide the choice of measurement setting when required, and to their respective measurement device.

Apart from the first assumption, which has not been challenged by any experiment so far, note that all three remaining assumptions concern the relation between the various devices involved in the experiment rather than their internal working. This approach is thus often called “black-box” or “device-independent”. This setting and these assumptions are sufficient to test a Bell inequality and they have been used recently in Ref. [23] to perform a loophole-free violation of the Bell-CHSH inequality. We briefly present this experiment in the next section.

Event-ready CHSH-Bell test with neutral atoms— In our experiment, Alice and Bob’s stations are made each with a single $^{87}$Rb atom stored in an optical dipole trap, see Fig. 1. The two setups are independently operated, that is, they are equipped with their own laser and control systems. Two Zeeman states $|m_F = \pm 1\rangle$ of the ground state manifold $5^2S_{1/2}$ are used as 1/2-spin states. After an initial state preparation, the atoms are optically excited to emit a photon whose polarization is entangled with the atomic spin states, see Fig. 3(a). The photons are sent to a station close to Alice’s location where a Bell state measurement is implemented with a beamsplitter followed by a polarizing beamsplitter at each output port and four single photon detectors. The atom excitation procedure is synchronized on a timescale that is much shorter than the photon duration. Careful adjustment of experimental parameters ensures a spectral, temporal and spatial mode overlap of photons close to unity. This allows us to achieve a high two-photon interference quality limited mostly by two-photon emission effects of a single atom. The joint measurement performed...
on these photons distinguishes two out of the four Bell states and ideally projects the atoms into either of the two states $|\psi^\pm\rangle = (|\uparrow\rangle_x |\downarrow\rangle_y \pm |\downarrow\rangle_x |\uparrow\rangle_y)/\sqrt{2}$ according to the outcome. Depending on the loading rate of the traps, 1 to 2 successful Bell state measurements are obtained per minute. At each success, a signal is sent to Alice and Bob and triggers setting choices. The analysis basis is selected by the output of a fast quantum random number generator. The measurement outcome is obtained by a spin-state dependent ionization with a fidelity of 97% on a timescale $\leq 1.1\mu s$. Given that Alice's and Bob's locations are separated by 398 m, this warrants space-like separation of the measurements. Although (strict) space-like separation is not a necessary condition for self-testing, it is a strong guarantee that Alice and Bob's measurement devices are indeed separated from each other and that information about the setting of one party is not available to the other one upon measurement, i.e. for assumptions 2 and 4 above.

**CHSH Bell inequality**– Let us label the measurement settings $x = 0, 1$ and $y = 0, 1$ for Alice and Bob respectively, with outcomes $a = 0, 1$ and $b = 0, 1$ for each spin measurement. For each pair of settings, we define the correlator $E_{xy} = \sum_{ab} (-1)^{a+b} P(a,b|x,y)$ where $P(a,b|x,y)$ is the conditional probability of observing outcome $a$ and $b$ when choosing the settings $x$ and $y$. This allows us to define the Bell-CHSH value, given by

$$S = E_{00} + E_{01} + E_{10} - E_{11}. \quad (1)$$

The latter is upper bounded by 2 for any local causal theory [24]. A significant violation of this bound can thus rule out this possibility, as conclusively demonstrated earlier [23], see also [19][21][22].

**Self-testing a Bell state**– Given assumption 1 above, we can associate to each measurement of Alice and Bob quantum observables $A_x$ and $B_y$ acting on two Hilbert spaces $H_A$ and $H_B$ of unknown dimension. Also, we can define the quantum state shared by the two parties as $\rho_{AB} \in L(H_A \otimes H_B)$. We emphasize that the internal functioning of the source and measurement boxes do not need to be known. We simply attribute a quantum state and measurement operators to the actual implementation. Our aim is to identify the actual state $\rho_{AB}$ from the observed statistics only. More precisely, we wish to estimate its fidelity with respect to a maximally entangled state of two qubits, that is

$$F(\rho_{AB}) = \max_{\Lambda_A, \Lambda_B} \text{Tr} (\langle \Lambda_A \otimes \Lambda_B |\rho_{AB} \rangle |\psi^-\rangle \langle \psi^-\rangle), \quad (2)$$

where the maximization is over all local trace-preserving maps $\Lambda_A, \Lambda_B : H_A \otimes H_B \rightarrow \mathbb{C}^2$. The role of these maps $\Lambda_{AB}$ is to identify the subsystems inside the unknown Hilbert spaces $H_A \otimes H_B$ in which $\rho_{AB}$ can be compared to the desired state. Given an observed Bell-CHSH value $S$, the Bell state self-testing fidelity is defined as the minimum fidelity of the unknown quantum state $\rho_{AB}$ which is compatible with the violation, i.e.

$$F = \min_{\rho_{AB} : \rho_{AB} \in S} F(\rho_{AB}) \quad \text{s.t.} \quad E_{00} + E_{01} + E_{10} - E_{11} = S, \quad (3)$$

where the correlators are now given by $E_{xy} = \text{Tr} (\rho_{AB} A_x B_y)$. This quantity captures the relation between $\rho_{AB}$ and the singlet state $|\psi^-\rangle$, one representative Bell state, that can be inferred from observed statistics: if the quantum state has no special relation to a Bell state, then $F \leq 1/2$; on the other hand, if $F = 1$, then we have the guarantee that local maps exist which identify perfectly a Bell state within the state $\rho_{AB}$, because this is the case for all admissible quantum realizations.

It has been shown that the self-testing fidelity $F$ can be directly related to the sole knowledge of the Bell-CHSH value $S$ [32]. The tightest known relation is given by [34]

$$F \geq f(S) = \max \left(40, 12 + (4 + 5\sqrt{2}) (5S - 8)\right)/80. \quad (4)$$

**Statistical analysis**– The previous formula holds in the limit where the CHSH value $S$ is known exactly. In order to analyze a real experiment with finite statistics, we consider that each run $i$ is chosen independently by both parties and with a maximum bias $\tau$ with respect to a uniform distribution, i.e. $1/2 - \tau \leq P(x), P(y) \leq 1/2 + \tau$, we construct in the Appendix [48] a confidence interval for the average CHSH violation. This allows us to show that $[\hat{F}, 1]$ with

$$\hat{F} = f \left(8 \left( I_{\alpha^{-1}}(n\bar{t}, n(1 - \bar{t}) + 1) - \tau - \tau^2 \right) - 4 \right) \quad (6)$$

is a one-sided confidence interval for $F$ with confidence level $1 - \alpha$. Here, $\bar{t} = (4 + 3\hat{S})/8$ with $\hat{S}$ the average

---

1 Note that here the values of 0 and 1 for the settings and outcomes were assigned arbitrarily, therefore, any of the 8 relabellings of Eq. (1) equally qualifies as a valid definition of the quantity $S$ [27]. With fixed measurement settings, such equivalent rewritings of the CHSH expression may be necessary to obtain a violation of the local bound with different Bell states.
CHSH value observed over the n rounds assuming a uniform sampling of the settings \( \mathbb{P} \), and \( I^1 \) is the inverse regularized incomplete Beta function, i.e., \( I_\gamma(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \) for \( y = (a, b) \). We emphasize that this bound on the average Bell state fidelity does not rely on the I.I.D. assumption.

**Preselection**— In contrast to results presented in Ref. [23], where all registered events were taken account, here we use a pre-selected set of events to compute the Bell-CHSH violation and the subsequent self-testing fidelities of heralded atomic states. This selection is based on a physical model which takes into account detrimental two-photon emission effects of a single atom and allows to define pre-selection criteria, here a time-window for acceptance of photons in the BSM, to improve the fidelity of the entangled atom-atom state. Details can be found in the Appendix [48]. Importantly, these considerations are not based on the results observed during the experiment. They are based on an ab-initio model of the underlying excitation and emission processes. This selection can then be seen as a pre-selection of the data, or equivalently, as a state preparation. In particular, it does not open the detection loophole or introduce expectation bias.

\[ \hat{F} \]

**Results**— For the evaluation we use the data of events heralding the \( |\psi^-\rangle \) state from the loophole-free Bell test [23] taken between 05.02.2016 and 24.06.2016 (25189 events). Fig. 2 shows the resulting lower bound of the average fidelity \( \hat{F} \) for the ab-initio model and the data set using the same pre-selection as a function of the acceptance window starting time \( t_s \) for different confidence levels. The model allows to determine the acceptance time window start time \( t_s \) for an optimal expected lower bound for the fidelity \( \hat{F} \) for each confidence level shown in Fig. 2. The results for the pre-selected data are shown in Tab. I. For calculation of the lower bound of the fidelity \( \hat{F} \) we consider bias of the RNGs bounded by \( \tau = 6.3 \times 10^{-4} \) (arising from the “paranoid” model for the predictability [23]).

The lower bound of fidelity exceeding the value of 0.5 (at a confidence level of up to \( 1 - 9.3 \times 10^{-8} \) for \( t_s = 746 \) ns) represents the first fully device-independent certification of a distributed entangled state. Moreover, an evaluation of the full data set without any pre-selection yields a Bell state fidelity of 0.5119 at 99% confidence and a Bell state fidelity larger than 0.5 can be certified even at a significance level as high as 99.8%.

Additionally, we applied our method to data from [19] obtained at a distance of 1.3 km. Due to the limited number of events (545), the method can only guarantee a fidelity larger than 50% with a confidence of \( \sim 96.8\% \). Still, this demonstrates that the method can be used in different systems without the need of knowing their specific details.

**Discussion and outlook**— We have derived a bound on the average fidelity of a measured state with respect to a Bell state from the sole knowledge of the observed Bell-CHSH value which is free of the I.I.D. assumption. We used it to quantify device-independently the quality of a bipartite state distributed over 398 m using an elementary quantum network link. These results guarantee that this link is suitable for an integration in a quantum network. This can be seen as a first step towards an implementation of device-independent quantum key distribution. Even though it is possible to get correlated bits from measurement results in this scenario, the fidelity still needs to be improved to retrieve a key for

\[ \begin{array}{|c|c|c|c|}
\hline
\text{CL} & \hat{F} & t_s & n \\
\hline
99\% & 0.5554 & 748 \text{ ns} & 13141 & 2.2589 \pm 0.0287 \\
99.9\% & 0.5411 & 746 \text{ ns} & 14807 & 2.2554 \pm 0.02713 \\
99.99\% & 0.5275 & 745 \text{ ns} & 15671 & 2.2505 \pm 0.0264 \\
\hline
\end{array} \]

**TABLE I.** Fidelity \( \hat{F} \) at different confidence levels. For each the data is pre-selected with an optimal start time for the acceptance time window \([t_s, t_e = 895 \text{ ns}] \) resulting in the corresponding in \( n \) events and an average CHSH value \( \bar{S}_{\psi} \).

---

\[ ^2 \] Here we assume that the average CHSH value is evaluated according to Eq. (23) from the Appendix [48].
secure communication. This leaves us with some work for the future on the experimental as well as theoretical aspects.

J.-D.B., P.S. and N.S. acknowledge funding by the Swiss National Science Foundation (SNSF), through the Grant PP00P2-179109 and 200021-175527, by the Army Research Laboratory Center for Distributed Quantum Information via the project SciNet and from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 820445 and project name Quantum Internet Alliance. K.R. and W.R. acknowledge funding by the German Federal Ministry of Education and Research via the project Q.com-Q.

[1] N. Gisin and R. Thew, Nature Photonics 1, 165 (2007)
[2] H.-J. Briegel, W. Dur, J.I. Cirac and P. Zoller, Phys. Rev. Lett. 81, 5932 (1998)
[3] N. Sangouard, C. Simon, H. De Riedmatten, and N. Gisin, Rev. Mod. Phys. 83, 33 (2011)
[4] For protocols that do not rely on entanglement creation in independent links, but are based exclusively on error correction codes, see E. Knill and R. Laflamme, arXiv:quant-ph/9908012.
[5] S. Pirandola, arXiv:1601.00966
[6] E. Schoute, L. Mancinska, T. Islam, I. Kerenidis, and S. Wehner, arXiv:1610.05238
[7] M. Pant et al., arXiv:1708.07142
[8] J. Wallnofer, M. Zwerger, C. Muschik, N. Sangouard, and W. Dur, Phys. Rev. A 94, 052307 (2016)
[9] M. Epping, H. Kampermann, and D. Bruss, New J. Phys. 18, 053036 (2016)
[10] S. Das, S. Khatria and J.P. Dowling, arXiv:1709.07400
[11] H.J. Kimble, Nature 453, 405 (2008)
[12] C.W. Chou, A. Yao, QIC, Vol.4 No. 4 July 5, pp273-286 (2004)
[13] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt, Phys. Rev. Lett. 23, 880 (1969)
[14] D. Mayers, A. Yao, QIC, Vol.4 No. 4 July 5, pp273-286 (2004)
[15] Z.S. Yuan et al.
[16] K. Lee et al., Science 334, 1253 (2011)
[17] J. Hofmann et al., Science 337, 72 (2012)
[18] S. Ritter et al. Nature 484, 195 (2012)
[19] B. Hensen et al. Nature 526, 682 (2015)
[20] A. Acín, N. Gisin, and L. Masanes, Phys. Rev. Lett. 97, 120405 (2006)
[21] L.K. Shalm et al. Phys. Rev. Lett. 115, 250402 (2015)
[22] M. Giustina et al. Phys. Rev. Lett. 115, 250401 (2015)
[23] W. Rosenfeld, et al. Phys. Rev. Lett. 119, 010402 (2017)
[24] J.S. Bell, Physics 1, 195 (1964)
[25] V. Scarani, Acta Physica Slovaca 62, 347 (2012)
[26] S. Popescu and D. Rohrlich, Phys. Lett. A 169, 411 (1992)
[27] S.L. Braunstein, A. Mann and M. Revzen, Phys. Rev. Lett. 68, 3259 (1992)
[28] J. F. Clauser, M. A. Horne, A. Shimony and R. A. Holt, Phys. Rev. Lett. 23, 880 (1969)
[29] D. Mayers, A. Yao, QIC, Vol.4 No. 4 July 5, pp273-286 (2004)
[30] M. Tomamichel, E. Hänggi, J. Phys. A: Math. Theor. 46, 055301 (2013)
[31] C. Miller, Y. Shi, 8th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2013, 22, 254 (2013)
[32] C-E. Bardyn, T. C. H. Liew, S. Massar, M. McKague, V. Scarani, Phys. Rev. A 80, 062327 (2009)
[33] T. H. Yang, T. Vértesi, J-D. Bancal, V. Scarani, M. Navascués, Phys. Rev. Lett. 113, 040401 (2014)
[34] J. Kaniewski, Phys. Rev. Lett. 117, 070402 (2016)
[35] M. McKague, Theory of Computing 12, 1 (2016)
[36] C. Bamps, S. Pironio, Phys. Rev. A 91, 052111 (2015)
[37] A. Natarajan, T. Vidick, Proc. of STOC ’17, pp. 1003-1015 (2017)
[38] A. Coladangelo, K. T. Goh, V. Scarani, Nature Communications 8, 15485 (2017)
[39] I. Supić, A. Coladangelo, R. Augusiak, A. Acín, arXiv:1707.06531
[40] F. Magniez, D. Mayers, M. Mosca, and H. Ollivier, in Proceedings of the 33rd International Colloquium on Automata, Lang Lecture Notes in Computer Science No. 4052, edited by Bugliesi (2006) pp. 72. arXiv:quant-ph/0512111
[41] B.W. Reichard, F. Unger, and U. Vazirani, Nature 496, 456 (2013)
[42] J-D. Bancal, M. Navascues, V. Scarani, T. Vértesi, and T. H. Yang, Phys. Rev. A 91, 022115 (2015)
[43] M. Dall’Arno, S. Brandsen, F. Buscemi, Proc. R. Soc. A, 473, 20160721 (2017)
[44] J. Kaniewski, Phys. Rev. A 95, 062323 (2017)
[45] P. Sekatski, J.-D. Bancal, S. Wagner, and N. Sangouard, arXiv:1802.02170
[46] T.R. Tan et al., Phys. Rev. Lett. 118, 130403 (2017)
[47] D. Rosset, J.-D. Bancal, N. Gisin, J. Phys. A: Math Theor. 47, 424022 (2014).
[48] See Supplemental Material
[49] B. Hensen et al. Scientific Reports volume 6, 30289 (2016)
[50] J. Hofmann, M. Krug, N. Ortegel, C. Kurz, M. Weber, W. Rosenfeld, H. Weinfurter, in preparation
[51] K. Redeker, R. Garthoff, W. Rosenfeld, H. Weinfurter, in preparation
[52] W. Hoeffding, Annals of Math. Stat. 27, 5 (1956)

Preselection of heralding events

To allow filtering as a preselection we have developed an ab-initio physical model independently of our measurement results to find an optimum filtering based on the model only. This model describes the photon emission process of a single atom excited by a short laser pulse and takes into account all important processes within its multilevel structure. Thereby we are able to calculate the expected fidelity for the entangled state of two atoms heralded by a two-photon coincidence at a certain time. The full description of the model goes far beyond the focus of the present work and shall be published elsewhere [50][51], here we only present a brief sketch of it.

For generation of a photon whose polarization is entangled with the atomic spin state, the atom is excited...
by a laser pulse resonant to the transition $5^2S_{1/2}, F = 1 \rightarrow 5^2P_{3/2}, F' = 0$. The temporal shape of the pulse is approximately Gaussian with a FWHM of 22 ns, see Fig. 4. After the successful emission of a photon, ideally, the atom should not interact with the excitation laser due to selection rules, see Fig. 3(a). In practice, however, the atomic state remains weakly sensitive to the excitation laser due to two reasons. First, there unavoidably are small polarization misalignments of the excitation laser, i.e. its polarization is not perfectly aligned along the quantization axis (imperfect $\pi$-polarization), allowing for a reexcitation of the $5^2P_{3/2}, F' = 0, m_F = 0$ level, see Fig. 3(b). Second, off-resonant scattering via the $5^2P_{3/2}, F' = 1$ level is possible (Fig. 3(c),(d)). Moreover, before the emission of a photon into the desired mode takes place, there is a finite probability that the atom emitted a first photon in a $\pi$-transition, which is not collected by the optics. These multiple photon emissions are detrimental for the quality of the atomic state announced by detection of the photons in the Bell state measurement in two different ways. On the one hand, the state of the atom can be changed by scattering additional photons. On the other hand, the interference quality of photons is reduced since the temporal shape and coherence of the photonic wavepackets is affected.

Importantly, the unwanted multiphoton processes happen predominantly during the excitation. Thus, to filter them, we only accept the detection events for which the first detector click is obtained after a time $t_s$ when the excitation laser pulse is essentially off, see Fig. 4. Additionally, to reduce the dark counts contribution to the heralding event we define a maximal time $t_e$ for the detection of the second photon. A later start time $t_s$ increases the entanglement swapping fidelity and by this the expected S-value for the CHSH inequality but at the expense of the obtained events number, see Fig. 5. Since the measured S-values will depend not only on the entanglement swapping fidelity but also on other properties, e.g., the atomic state measurement fidelity and the coherence time of the entangled states, we use the experimental parameters as specified in [23] to estimate the expected S-value.

The optimal selection of the time window of $[t_s = 748 \text{ ns}, t_e = 895 \text{ ns}]$, considering Eq. (6) for a 99% confidence interval, reduces the number of heralding events by approximately a factor of 2 but the atoms are expected to be in an entangled state of a higher quality.

**Finite statistics analysis**

In this section, we detail the construction of the confidence interval on the average singlet fidelity reported in the main text.

**Model**

In the experimental situation described in the main text, the settings used after the $i$th heralding event can be described by two random variables $X_i$ and $Y_i$. These variables follow a global probability distribution

$$P(\vec{X} = \vec{x}, \vec{Y} = \vec{y}),$$

where $\vec{X} = (X_1, X_2, \ldots)$, $\vec{Y} = (Y_1, Y_2, \ldots)$ and $\vec{x}, \vec{y} \in \{0,1\}^n$ for a binary choice of settings.

Similarly, two random variables $A_i$ and $B_i$ can be used to describe the outcomes observed upon measuring the

---

**FIG. 3.** Structure of the relevant levels in $^{87}$Rb, excitation and decay processes. (a) Generation of atom-photon entanglement in spontaneous decay of the excited $5^2P_{3/2}, F = 0, m_F = 0$ level. The excitation laser is shown in orange. Photons polarized linearly along the quantization axis ($\pi$-decays, gray) are not detected in our system. After the first decay, a second excitation is possible due to polarization misalignment (b), or off-resonant excitation (c). If the $5^2P_{3/2}, F' = 1$ is excited, also decay to $5^2S_{1/2}, F = 2$ level is possible (d).

**FIG. 4.** Time histogram of single photon detection (red) for excitation of a single atom by a short resonant laser pulse (blue). $t_s$ and $t_e$ define the acceptance time window for coincidences.
Before focusing on the fidelity, our figure of merit, let us recall how to estimate the Bell contribution corresponding to a given round \(i\). We introduce the statistic

\[
T_{i|\text{past}_i} = \frac{1}{4} \cdot \frac{\chi(A_i \oplus B_i = X_i Y_i)}{P(X_i, Y_i|\text{past}_i)},
\]

where \(\chi\) is the indicator function, i.e. \(\chi(\text{condition}) = 1\) if the condition is true and \(\chi(\text{condition}) = 0\) for a false condition, and \(\oplus\) is the addition modulo 2. The expectation value of this estimator is directly related to the CHSH violation on round \(i\) given the past:

\[
\mathbb{E}(T_{i|\text{past}_i}) = \sum_{\vec{a}, \vec{b}, \vec{x}, \vec{y}} T_{i|\text{past}_i} \cdot \frac{\chi(\vec{a} \oplus \vec{b} = \vec{x} \cdot \vec{y})}{P(\vec{X} = \vec{x}, \vec{Y} = \vec{y})}
\]

\[
\quad \quad = \frac{1}{4} \sum_{a,b,x,y} \frac{\chi(a \oplus b = xy)}{P(X_i = x, Y_i = y|\text{past}_i)} \cdot P(A_i = a, B_i = b, X_i = x, Y_i = y|\text{past}_i)
\]

\[
\quad \quad = \frac{1}{4} \sum_{a,b,x,y} \chi(a \oplus b = xy) \cdot P_i(a, b|x, y, \text{past}_i)
\]

\[
\quad \quad = 4 + S_{i|\text{past}_i}.
\]

This expression thus provides a good estimation of the Bell violation contribution of round \(i\). Note that the relation (13) is valid for all distribution of the settings which is independent from \(A\) and \(B\)'s behavior according to Eq. (8).

In the case where the settings are chosen uniformly, i.e. \(P(X_i = x_i, Y_i = y_i|\text{past}_i) = \frac{1}{4}\), the random variable \(T_{i|\text{past}_i}\) is a Bernoulli variable whose only possible values are 0 and 1. It can then be interpreted as a binary game which is either won (if \(T_{i|\text{past}_i} = 1\)) or lost (if \(T_{i|\text{past}_i} = 0\). The CHSH contribution of round \(i\) can then be re-interpreted in terms of the winning probability \(q_{i|\text{past}_i} = P(T_{i|\text{past}_i} = 1)\) of this game, such that

\[
8 q_{i|\text{past}_i} = 4 + S_{i|\text{past}_i}.
\]

**Settings choice bias**

In practice, it may be difficult to guarantee that the choice of settings is exactly uniform. One can then resort to a partial characterization of the settings’ distribution. For instance, consider the case where the settings of Alice and Bob are chosen independently as

\[
P(\vec{X} = \vec{x}, \vec{Y} = \vec{y}) = \prod_i P(X_i = x_i, Y_i = y_i|\{x_j, y_j\}_{j<i})
\]
with
\[
P(X_i = x_i, Y_i = y_i|\{x_j, y_j\}_{j<i}) = P(x_i|\tau_i^x)P(y_i|\tau_i^y)
\]
and
\[
P(x_i|\tau_i^x) = \frac{1}{2} + (-1)^{x_i}\tau_i^x
\]
and
\[
P(y_i|\tau_i^y) = \frac{1}{2} + (-1)^{y_i}\tau_i^y,
\]
and we only have the guarantee that the local biases are bounded \(|\tau_i^x|, |\tau_i^y| \leq \tau\) by some maximal value \(\tau \leq \frac{1}{2}\).

In this case the statistic \(T_{i|\text{past}}^u\) as well as the CHSH value Eq. (10) cannot be evaluated directly. We can nevertheless bound its behavior.

For this, let us then consider the statistic that would correspond to a uniform choice of settings
\[
T_{i|\text{past}}^u = \chi(A_i \oplus B_i = X_iY_i).
\]

As mentioned before, this statistic is a Bernoulli random variable taking value either 0 or 1. Its winning probability is \(q_{i|\text{past}}^u = \mathbb{E}(T_{i|\text{past}}^u)\), and can be evaluated without having the knowledge of distribution of settings. Its expectation value is given by
\[
\mathbb{E}(T_{i|\text{past}}^u) = \sum_{a,b,x,y} \chi(a \oplus b = xy) P_i(a, b | x, y, \text{past}_i)
\]
\[
\times P(x|x^x)P(y|y^y)
\]
\[
= \sum_{a,b,x,y} \chi(a \oplus b = xy) P_i(a, b | x, y, \text{past}_i)
\]
\[
\times \left( \frac{1}{4} + \frac{(-1)^x}{2} \tau^x + \frac{(-1)^y}{2} \tau^y + (-1)^{x+y} \tau^x \tau^y \right)
\]
\[
= \frac{1}{4} \sum_{a,b,x,y} \chi(a \oplus b = xy) P_i(a, b | x, y, \text{past}_i)
\]
\[
+ \sum_{a,b,x,y} \chi(a \oplus b = xy) P_i(a, b | x, y, \text{past}_i)
\]
\[
\times \left( \frac{(-1)^x}{2} \tau^x + \frac{(-1)^y}{2} \tau^y + (-1)^{x+y} \tau^x \tau^y \right)
\]
\[
\text{Defining } f_{xy} = \sum_{a,b} \chi(a \oplus b = xy) P(a, b | x, y, \text{past}_i) \in [0,1] \text{ we write}
\]
\[
\mathbb{E}(T_{i|\text{past}}^u) - \mathbb{E}(T_{i|\text{past}}) = \sum_{x,y} f_{xy} \left( \frac{(-1)^x}{2} \tau^x + \frac{(-1)^y}{2} \tau^y + (-1)^{x+y} \tau^x \tau^y \right).
\]

Let us now consider this sum. Without loss of generality we set \(0 \leq \tau_0 \leq \tau_x \leq \tau\), all the other cases directly follow by a permutation of the outcomes or the exchange of \(x\) and \(y\). One has
\[
2 \sum_{x,y} f_{xy} \left( \frac{(-1)^x}{2} \tau^x + \frac{(-1)^y}{2} \tau^y + (-1)^{x+y} \tau^x \tau^y \right)
\]
\[
= f_{00} (\tau^x + \tau^y + 2 \tau^x \tau^y) + f_{10} (\tau^x - \tau^y + 2 \tau^x \tau^y) + f_{01} (\tau^x + \tau^y - 2 \tau^x \tau^y) + f_{11} (\tau^x - \tau^y - 2 \tau^x \tau^y)
\]
\[
\leq \tau^x + \tau^y + 2 \tau^x \tau^y + f_{01} (\tau^x - \tau^y - 2 \tau^x \tau^y),
\]
where we used \(f_{xy} \in \{0,1\}\), \(\tau^x + \tau^y + 2 \tau^x \tau^y \geq 0, -\tau^x + \tau^y - 2 \tau^x \tau^y \leq 0\) and \(-\tau^x - \tau^y + 2 \tau^x \tau^y \leq 0\). For the last term one finds
\[
f_{01} (\tau^x - \tau^y - 2 \tau^x \tau^y) \leq \begin{cases} \tau^x - \tau^y - 2 \tau^x \tau^y & \tau^x < \frac{\tau^y}{1+2\tau^y} \\ 0 & \tau^y \geq \frac{\tau^x}{1+2\tau^x} \end{cases}
\]
leading to
\[
\sum_{x,y} f_{xy} \left( \frac{(-1)^x}{2} \tau^x + \frac{(-1)^y}{2} \tau^y + (-1)^{x+y} \tau^x \tau^y \right)
\]
\[
\leq \begin{cases} \tau^x & \tau^y < \frac{\tau^x}{1+2\tau^x} \\ \frac{1}{2}(\tau^x + \tau^y) & \tau^y \geq \frac{\tau^x}{1+2\tau^x} \end{cases}
\]
\[
\leq \tau + \tau^2,
\]
which holds for all allowed values of \(\tau^x\) and \(\tau^y\). Plugging this inequality in Eq. (22) then gives
\[
\mathbb{E}(T_{i|\text{past}}^u) - \mathbb{E}(T_{i|\text{past}}) \leq \tau + \tau^2.
\]

Therefore, we obtain
\[
q_{i|\text{past}}^u \geq q_{i|\text{past}}^w - \tau - \tau^2,
\]
meaning that a lower bound on the winning probability \(q_{i|\text{past}}^w\) of the uniform statistic \(T_{i|\text{past}}^u\) gives rise to a lower bound on \(q_{i|\text{past}}^w\). In order to estimate \(q_{i|\text{past}}\) with a distribution of settings which is not fully known, we can thus safely estimate the CHSH value with the statistic \(T_{i|\text{past}}^u\), effectively assuming that the settings are chosen uniformly, and then correct the winning probability \(q_{i|\text{past}}^w\) according to the value of \(\tau\), as expressed in (28). This provides a lower bound on the actual winning probability \(q_{i|\text{past}}\) of \(T_{i|\text{past}}^w\).

To simplify the notation, we now drop the explicit conditioning on the past, and thus simply write e.g. \(S_i, q_i, T_i\) and \(\mathcal{F}\) for the quantities introduced above.

**Bounding the fidelity**

We construct a statistical parameter for the whole experiment corresponding to the average Bell state fidelity:
\[
\mathcal{F} = \frac{1}{n} \sum_i \mathcal{F}_i.
\]
Thanks to the convex relation between the CHSH violation and the singlet fidelity Eq. (4), this quantity can be bounded from the average CHSH violation \( \mathcal{S} \) as

\[
\mathcal{F} = \frac{1}{n} \sum_i F_i \geq \frac{1}{n} \sum_i f(S_i) \geq f \left( \frac{1}{n} \sum_i S_i \right) = f(\mathcal{S}),
\]

or equivalently, using relation (14), from the average winning probability \( \overline{q} = \frac{1}{n} \sum_i q_i \) as

\[
\mathcal{F} \geq f(8\overline{q} - 4).
\]

In particular, a left-sided confidence interval for \( \overline{q} \) gives rise to a one-sided confidence interval for the singlet fidelity. By relation (28), a left-sided confidence interval for \( \overline{q} = \frac{1}{n} \sum_i q_i \) gives rise to a left-sided confidence interval for \( q \), and thus also for the singlet fidelity:

\[
\mathcal{F} \geq f \left( 8(\overline{q} - \tau - \tau^2) - 4 \right).
\]

Let us thus focus now on the average winning probability \( \overline{q} \).

A confidence interval for the average winning probability

The random variables \( T_i^u \) in Eq. (49) being estimators for the parameters \( q_i^u \), we use their average

\[
\overline{T}^u = \frac{1}{n} \sum_{i=1}^{n} T_i^u
\]

to estimate \( \overline{q}^u \). This gives rise to the following effective CHSH value

\[
\overline{S}^u = 8\overline{T}^u - 4
\]

which can be evaluated in practice directly from the observed data, without assumption on the distribution of measurement settings. Note that each random variable \( T_i^u \) is a Bernoulli variable with parameter \( q_i^u \). Therefore, \( \overline{T}^u \) is a so-called (renormalized) Poisson binomial random variable. The distribution probability of such a random variable in terms of the average parameter \( \overline{q}^u \) has been characterized by Hoeffding [52]. We recall this result here.

**Theorem 1.** (Hoeffding, 1956). Let \( \overline{T} = \frac{1}{n} \sum_{i=1}^{n} T_i \) be the average of \( n \) independent Bernoulli variables \( T_i \) with parameters \( q_i \). If \( c \) and \( d \) are two integers such that

\[
0 \leq c \leq n\overline{q} \leq d \leq n
\]

for \( \overline{q} = \frac{1}{n} \sum_{i=1}^{n} q_i \), then

\[
P(c \leq n\overline{T} \leq d) \geq \sum_{k=c}^{d} \binom{n}{k} (1 - \overline{q})^{n-k}.
\]

This theorem says that within all sets of \( n \) choices of Bernoulli variables \{\( T_i \)\}_{i=1}^{n} with a fixed average parameter \( \overline{q} = \frac{1}{n} \sum_i q_i \), the one producing the largest tail distribution for the average variable \( \overline{T} \) is the set of \( n \) identically-distributed Bernoulli variables with \( q_i = \overline{q} \) \( \forall i \). The tail probability then follows a binomial distribution.

We recall that the Binomial cumulative distribution can be expressed in terms of the regularized incomplete Beta function by \( \overline{I}_\alpha^{-1} \), i.e. \( \overline{I}_\alpha^{-1}(a,b) = x \) for \( y = \overline{I}_\alpha^{-1}(a,b) \). We now construct a confidence interval for the average CHSH winning probability \( \overline{q} \).

**Theorem 2.** Given Bernoulli random variables \( T_i \) with parameter \( q_i \) for \( i = 1, \ldots, n \) and \( 0 \leq \alpha \leq 1/2 \), the interval \([\hat{q}, 1]\) with the random variable

\[
\hat{q} = \overline{I}_\alpha^{-1}(n\overline{T}, n(1 - \overline{T}) + 1)
\]

is a confidence interval for \( \overline{q} = \frac{1}{n} \sum_i q_i \) with confidence level \( 1 - \alpha \).

**Proof.** We need to show that

\[
P(\hat{q} \leq \overline{q}) \geq 1 - \alpha
\]

for all possible sets of Bernoulli variables characterized by parameters \( 0 \leq q_i \leq 1 \) with \( \frac{1}{n} \sum_i q_i = \overline{q} \). Condition (42) states that whatever the unknown distribution of the Bernoulli variables \( T_i \) happens to be, the value of \( \hat{q} \) computed from them (the random variable \( \hat{q} \) depends on the observed \( \overline{T} \)) is lower than the actual parameter \( \overline{q} \) only with probability at most \( \alpha \). \( \hat{q} \) then constitutes a reasonable lower bound for the parameter \( \overline{q} \).

But before starting, let us introduce the function

\[
g : \mathbb{R} \to \mathbb{R}
\]

\[
z \mapsto \overline{I}_\alpha^{-1}(z, n - z + 1).
\]

This function describes the trade-off between the parameters \( q \) and \( c \) in the sum (40). Clearly, increasing the value of \( c \) reduces the sum, unless \( q \) increases as well.

Since \( g(c) \) is defined as the value of \( q \) which keeps the sum invariant (and equal to \( \alpha \)), it is a monotonically increasing function.
Let us now write
\[ P(\hat{q} \leq \overline{q}) = \sum_{k=0}^{d} P(nT = k) \times \chi(g(k) \leq \overline{q}) = \sum_{k=0}^{d} P(nT = k) \] (44)

where \( P(nT = k) \) is the probability distribution of the sum \( nT \) of arbitrary Bernoulli random variables and \( d \) is the largest integer such that \( g(d) \leq \overline{q} \). The sum contains all terms between 0 and \( d \) because \( g(x) \) is an increasing function. For \( \alpha \leq 1/2 \), the condition \( d \geq n\overline{q} \) is satisfied, so we can use Thm. (1) to lower bound this probability by the binomial case:

\[ P(\hat{q} \leq \overline{q}) = \sum_{k=0}^{d} P(nT = k) \]

\[ = P(nT \leq d) \]

\[ \geq \sum_{k=0}^{d} \binom{n}{k} \overline{q}^k (1 - \overline{q})^{n-k} \]

Using Eq. (39) we obtain

\[ P(\hat{q} \leq \overline{q}) \geq 1 - I_{\overline{q}}(d + 1, n - d). \] (46)

Here, \( d \) is by definition the maximal integer for which \( g(d) \leq \overline{q} \) holds. It follows that \( g(d + 1) \geq \overline{q} \), or

\[ I_{\overline{q}}^{-1}(d + 1, n - d) \geq \overline{q}. \] (47)

Next, we will apply the regularized incomplete Beta function to both sides of this inequality. But before we do so, remark that the cumulative of a binomial distribution is monotonously decreasing with the parameter \( q \), i.e.

\[ \sum_{k=0}^{d} \binom{n}{k} q^k (1-q)^{n-k} \geq \sum_{k=0}^{d} \binom{n}{k} q'^k (1-q')^{n-k} \]

\[ \forall \ 0 \leq q \leq q' \leq 1. \] (48)

It follows by Eq. (39) that \( I_q(d + 1, n - d) \) is a monotonously increasing function of \( q \):

\[ I_q(d + 1, n - d) \leq I_{q'}(d + 1, n - d) \]

\[ \forall \ 0 \leq q \leq q' \leq 1. \] (49)

Hence, applying this function to both sides of (47) preserves the inequality

\[ I_{\overline{q}}^{-1}(d + 1, n - d) = \alpha \geq I_{\overline{q}}(d + 1, n - d). \] (50)

Combining with Eq. (46) completes the proof

\[ P(\hat{q} \leq \overline{q}) \geq 1 - I_{\overline{q}}(d + 1, n - d) \]

\[ \geq 1 - \alpha. \] (51)