Testing quantum randomness in single-photon polarization measurements with the NIST test suite

David Branning[1,*] and Matthew Bermudez[1,2]

[1]*Department of Physics, Trinity College, Hartford, Connecticut 06106, USA*
[2]*Current address: Department of Physics, University of Connecticut, 2152 Hillside Road, Storrs, Connecticut 06269-3046, USA*
**Corresponding author: david.branning@trincoll.edu*

Received March 11, 2010; revised June 14, 2010; accepted June 26, 2010; posted June 29, 2010 (Doc. ID 125384); published July 21, 2010

A binary sequence was constructed from 1.7×10^7 polarization measurements of single photons from a spontaneous parametric downconversion source, under pumping conditions similar to those used in optical quantum cryptography. To search for correlations in the polarization measurement outcomes, we subjected the sequence to a suite of tests developed at the National Institute of Standards and Technology (NIST) for the assessment of algorithmic random-number generators. The bias of the sequence was low enough to allow all fifteen tests to be applied directly to the polarization outcomes without using any numerical unbiasing procedures. No statistically significant deviations from randomness were observed, other than those related to this small uncorrected bias. © 2010 Optical Society of America

OCIS codes: 270.0270, 270.5568, 270.5290.

1. INTRODUCTION

Over the past few decades, single photons have been used to investigate the most nonclassical features of quantum theory [1]. Quantum interference and entanglement [2,3], in particular, have been vigorously explored in optical systems, not only because of their relevance to the foundations of quantum mechanics, but also because of their practical value as resources for quantum information processing schemes such as quantum key distribution [4,5] and quantum computation [6,7].

However, another central feature of quantum mechanics has received less attention: the randomness of measurement outcomes from superposition states. The nondeterministic character of these outcomes—the inability to predict what the next one will be, given knowledge of all previous outcomes from the same state—is usually assumed in order to give meaning to the probabilistic interpretation of the quantum state [8]. The lack of correlations between successive measurement outcomes is also important for quantum key distribution schemes, in which the results of a sequence of quantum measurements on superposition states are used to generate a secret key, or "one-time pad," that enables secure communication between two parties [4,5].

But as pointed out by Erber [8] some years ago, the randomness of quantum phenomena might well be considered as a *physical assumption* that is testable in its own right, independent of the other foundational aspects of quantum theory. In recent years, experimental tests of randomness have been performed on time-binned sequences of radioactive decays [9–11], and on implementations of "quantum optical random number generators." The latter use photon detection times [12–15] or polarizations [16] or both [17] to generate random number sequences; both degrees of freedom are commonly used as the basis for quantum key distribution schemes [4,5].

Here we report on a comprehensive set of tests of the randomness of single-photon polarization measurement outcomes, using pairs of photons generated by spontaneous parametric downconversion, under conditions similar to those of many quantum cryptographic schemes. One member of each pair was used as a detection trigger, while the other was put into a superposition state of horizontal (H) and vertical (V) polarization, and then measured in the H–V basis. The time sequence of H and V outcomes was subjected to a suite of tests developed at the National Institute of Standards and Technology (NIST) to assess the quality of computer-based random-number generators [18]. Several of the tests require many distinct low-bias sequences of at least 10^6 bits in order to be meaningful; to our knowledge, this is the first direct application of these tests to sequences of two-level quantum events.

2. SINGLE-PHOTON POLARIZATION MEASUREMENTS

The single-photon polarization measurements were conducted as shown in Fig. 1. Continuous-wave light from a diode "pump" laser at 405 nm was incident on a 3.0 mm BBO crystal serving as a parametric downconverter (PDC) [19]. The crystal was cut and oriented for Type-I downconversion, with an output angle of 3 degrees for the frequency-degenerate downconverted light at 810 nm. Each downconverted light beam (signal and idler) was launched into an optical fiber via a lens and sent to a

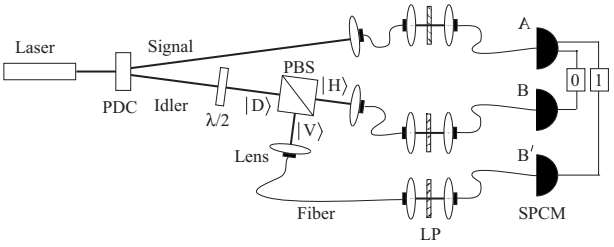

Fig. 1. Experimental arrangement for measuring single-photon polarizations. Signal and idler photon pairs are created in the PDC and counted in coincidence either at detectors AB or AB′, depending on the measurement outcome for the diagonally polarized idler photon in the H–V basis. A binary sequence is created by assigning "0" to the coincidence events AB and "1" to the events AB′.

single-photon counting module (SPCM) for detection. Background counts were reduced by long-pass filters (LP) inserted into each optical channel, which absorbed light with wavelengths shorter than 780 nm. These filters were inserted further down the fiber path to allow a 5 mW, 785 nm diode laser (not shown) to be coupled into the detection system at various points for alignment.

Using a single-mode (monochromatic) approximation for the signal and idler beams, the PDC source produces a quantum state whose leading terms are [20]

$$|\psi\rangle = M|vac\rangle + \eta|H\rangle_s|H\rangle_i, \tag{1}$$

where M is a normalization constant and η is a small number characterizing the size of the PDC's perturbation on the initial vacuum state of the signal and idler modes s and i. The second term represents a product state of one signal and one idler photon, each horizontally polarized as they emerge from the PDC. Before entering the collection fiber, each idler photon passed through a zero-order half-wave plate ($\lambda/2$), oriented to rotate its polarization state from $|H\rangle$ to $|D\rangle = 1/\sqrt{2}(|H\rangle_i + |V\rangle_i)$. The idler then impinged onto a polarizing beamsplitter (PBS), which transmitted H-polarized photons and reflected V-polarized photons to separate lens-fiber collection channels.

The state (1) reflects the fact that the signal and idler photons must be created together to conserve energy and momentum in the PDC process. Therefore the detection of a signal photon can be used as a trigger, allowing another detector to look for an idler photon only during a brief interval afterward, creating a close approximation to a localized one-photon state for the idler [21]. To implement this "heralded" single-photon source [1], the 5-volt TTL pulse from detector A was sent to one input of an electronic AND gate, while the pulse from an idler detector (B or B′) was sent to the other input. The AND gate produced a TTL output only when both of its inputs were at 5 volts simultaneously, and this "coincidence count" was used to register the polarization outcome for each idler photon. Thus, a coincidence count AB between detectors A and B represented a horizontal polarization outcome for the idler photon, while a coincidence event AB′ represented a vertical outcome. The durations of the signal and idler TTL pulses were adjusted to be 10 ns before coming to the AND gate, so that the gate would only produce a coincidence count when the signal and idler inputs overlapped within a 10 ns "coincidence window" [22].

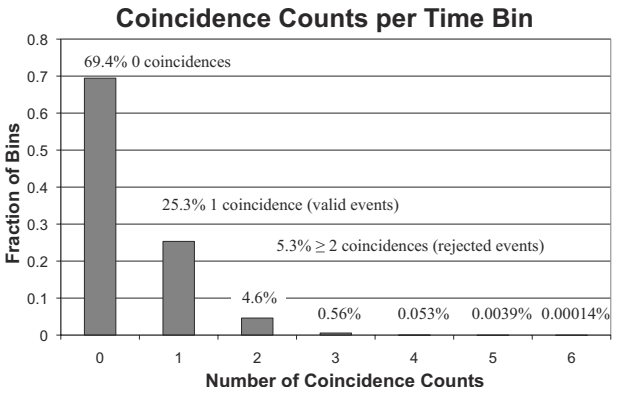

Fig. 2. Numbers of coincidence counts per 0.1 ms time bin. Roughly 25% of the bins contained exactly one coincidence count. These events were used to construct the binary sequence for testing.

The data collection rate was 10 kHz. In each consecutive time interval of duration 0.1 ms, the number of coincidence counts AB and AB′ was recorded. Data were collected over a period of 1.87 h, or 6.72×10^7 bins. A histogram of the number of coincidence counts collected per bin is shown in Fig. 2. Because the spontaneous downconversion process from the cw laser did not occur at regular time intervals, it was possible for two or more coincidence events to occur within a given time bin. The time-ordering of these multiple-count events could not be determined, and so they were discarded (see Appendix A). To reduce the probability of these unwanted multiple-events, the pump laser power was adjusted to make the average number of photon pairs per time bin, μ, much less than 1. A fit of the occupation numbers in Fig. 2 to a Poisson distribution (as expected from integrating the output of a multi-mode parametric downconversion source over many coherence times [20]) yielded a mean number of coincidence events per time bin of $\mu = 0.364 \pm 0.002$, with a reduced χ^2 of 1.16 for 6 degrees of freedom. The probability of a larger χ^2 value arising by chance is 32.5%.

A binary sequence was then constructed from just those time bins in which exactly one coincidence event occurred—about one fourth of the bins, as shown in Fig. 2, giving a net random bit rate of 2.5 kHz. The coincidence event AB (see Fig. 1), indicating a horizontal polarization outcome for the idler photon, was designated as "0" in the sequence, while the event AB′, indicating a vertical idler polarization outcome, was designated as "1." The resulting sequence contained 1.7×10^7 bits, with a bias of 0.04% toward the value "0." To apply the NIST suite of tests, the sequence was divided into sub-sequences of either 10^5 or 10^6 bits as required.

3. RANDOMNESS TESTS

Each of the fifteen tests comprising the NIST Statistical Test Suite [18] analyzed the polarization sequence for a particular statistical facet of randomness, according to the same general procedure: first, some numerical characteristic of the sequence, such as the sum of all its digits, was obtained. Next, the probability p that a truly random source could produce this value of the characteristic, or

one that is even farther away from the ideal, was computed. This probability, the "p-value," represents the degree of randomness of the string with regard to the characteristic being tested. A high p-value means that the string appears random, while a low p-value suggests that it is unlikely that a truly random source could produce this value of the chosen characteristic through chance alone.

However, a single low p-value is not conclusive evidence that a string is non-random. For an ideal random number generator, the p-values for each test are expected to be distributed uniformly on the interval $0<p\leqslant 1$, so that one in every 100 sequences will have $p\leqslant 0.01$, and one in every ten will have $p\leqslant 0.1$. Thus, on average one in 100 sequences from an ideal random source will fail each test, by chance, at the "confidence threshold" of 0.01, while one in ten will fail at the confidence threshold of 0.1. The appropriate confidence threshold for a given test is therefore determined by the number of sequences available to be tested. For example, if more than 100 sequences are available, a confidence threshold of 0.01 can be meaningfully tested; if more than 1% of the sequences fail at this threshold, then the randomness is suspect.

No test can prove definitively that a sequence is random, but if many sequences are tested, a poor random-number generator may be revealed by greater incidence of failure of a particular test than is expected for a particular confidence threshold. For example, if five sequences out of 100 have $p\leqslant 0.01$, then the source almost certainly fails the test at the 1% threshold. But if only nine of these sequences have $p\leqslant 0.1$, then the source would still pass at the 10% level.

Ten of the NIST tests require sequences of up to 10^5 bits as their inputs, while the other five require at least 10^6 bits per sequence. Therefore, the first ten tests (our order differs from that of [18]) were run using 170 sub-sequences of 10^5 bits, enabling a meaningful confidence threshold of 0.01, while the five more stringent tests were run on 17 sequences of 10^6 bits, allowing a meaningful confidence threshold of only 0.1.

A. Bias (Frequency) Test

The first test checks the bias of the sequence toward the outcome "0" or "1." This is the simplest test, and passing it is a prerequisite for many of the other tests in the NIST suite. Given many sample sequences from a truly random source, the number of bits in excess of a completely unbiased sequence of either outcome follows a known distribution, explained in more detail in Appendix A. For a given confidence threshold, the acceptable number of excess 1's or 0's grows as the square of the length of the sequence. Figure 3 shows a histogram of the results of the Bias Test for the 170 sub-sequences of 10^5 polarization measurements. The higher-than-expected number of failures (shown in white) revealed the bias in the source. Collectively, the p-values from this test fit the expected uniform distribution with a reduced χ^2 of 1.397, giving a probability of 18% of obtaining a worse fit by chance alone.

B. Bias (Frequency) Within a Block Test

The second test checks the bias of "blocks," or subsequences of the main sequence, of lengths up to 10^4 bits.

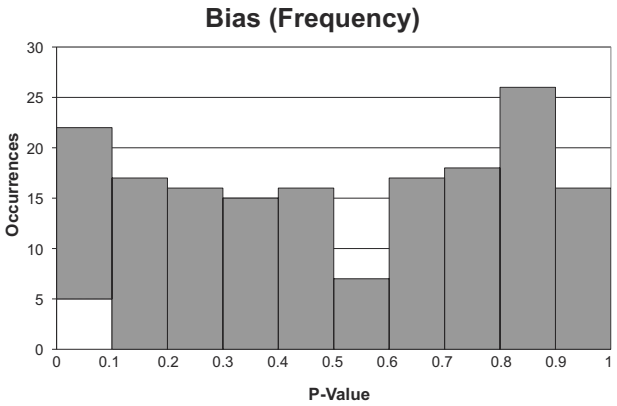

Fig. 3. P-values for the Bias Test. Each of the 170 sub-sequences produced one p-value. The white space in the first bin represents the five sequences, or 2.94% of them, which failed at the 0.01 level ($p\leqslant 0.01$). This was significantly more than the one or two failures expected at this level, indicating that the source was not perfectly unbiased.

The results are shown in Fig. 4. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 1.327 and a probability of 22% of obtaining a worse fit by chance alone. Because only shorter blocks are used, this test is not as sensitive to the overall source bias as the first test, and no significant bias was detected among the blocks.

C. Number of Runs Test

This test determines whether there is an appropriate number of uninterrupted repetitions of 1 or 0 of various lengths within the sequence. The results of this test are shown in Fig. 5. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 0.444, for a probability of 91% of obtaining a worse fit by chance alone.

D. Longest Run of Ones in a Block Test

For subsequences, or "blocks," of various lengths M, this test finds the lengths of the longest uninterrupted runs of "1" outcomes, and compares their actual occurrences to an expected distribution [23]. The results are shown in Fig. 6. The p-values from this test fit the expected uniform dis-

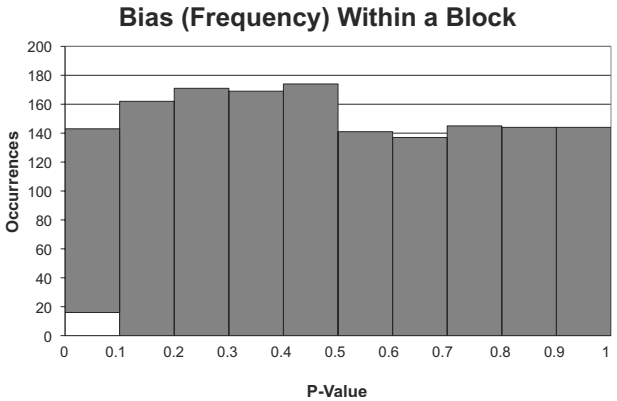

Fig. 4. P-values for the Block Bias Test, which produced nine p-values for each of the 170 sub-sequences. Sixteen of the 1530 p-values, or 1.05% of them, were below 0.01.

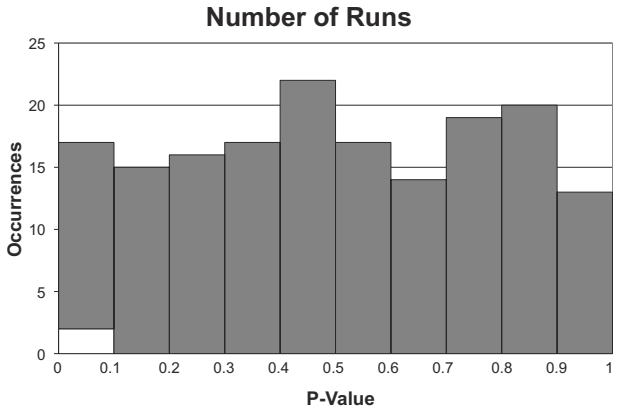

Fig. 5. P-values for the Number of Runs Test. P-values for two of the sub-sequences, or 1.18% of them, were below 0.01.

tribution with a reduced χ^2 of 0.588 and a probability of 81% of obtaining a worse fit by chance alone.

E. Binary Matrix Rank Test

This test examines the number of linearly independent rows and columns (the rank) of square matrices made from successive bits of the original sequence. Deviation from the expected distribution of ranks would indicate a level of periodicity in the sequence that is lower or higher than expected for a random sequence. The results are shown in Fig. 7. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 1.12 and a probability of 34% of obtaining a worse fit by chance alone.

F. Discrete Fourier Transform Test

This test analyzes the distribution of peak heights in the discrete Fourier transform of the sequence. Too many large peaks would indicate a cyclic process in the generation of the sequence. The results are shown in Fig. 8. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 0.797 and a probability of 61% of obtaining a worse fit by chance alone.

Although there was an error in this test as formulated in the original NIST publication, the Mathematica program used to generate the results in Fig. 8 included a correction to the reference distribution [24,25]. This correction has also been incorporated into the most recent version of the NIST Statistical Test Suite [18].

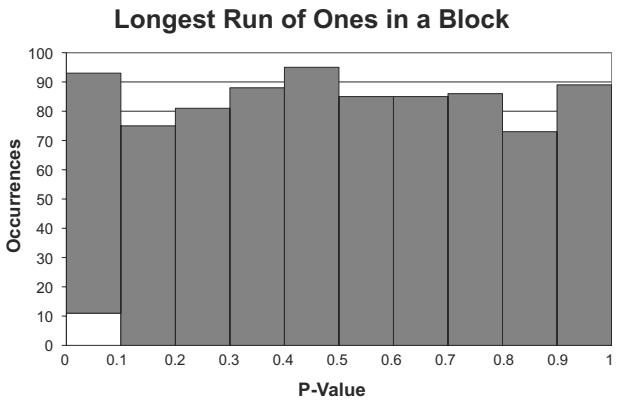

Fig. 6. P-values for the Longest Run of Ones in a Block Test. The test generated five p-values for each sub-sequence. Eleven of these 850 p-values, or 1.29% of them, were below 0.01.

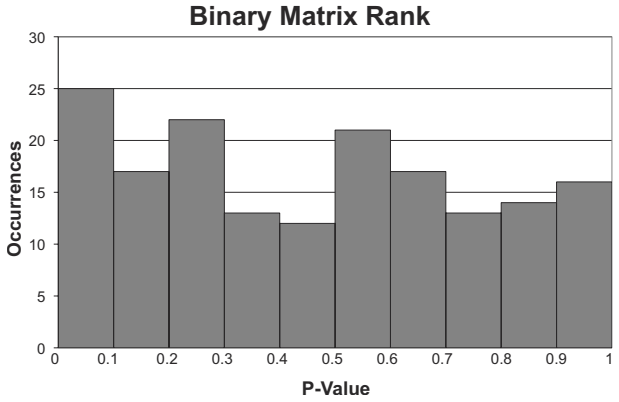

Fig. 7. P-values for the Binary Matrix Rank Test. There were no p-values below 0.01.

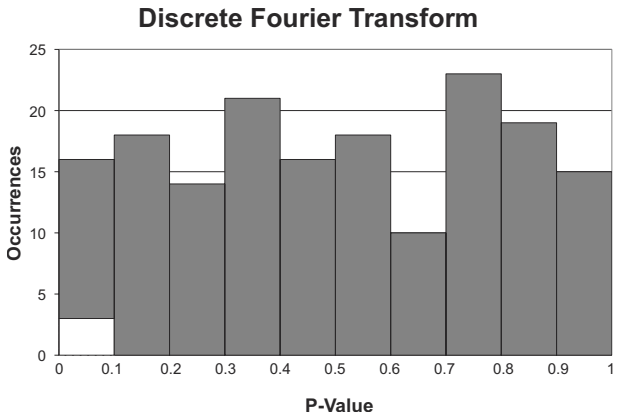

Fig. 8. P-values for the Discrete Fourier Transform Test. P-values for three of the sub-sequences, or 1.76% of them, were below 0.01.

G. Approximate Entropy Test

This test determines how often patterns of different lengths overlap, revealing whether there is too much or too little regularity in the sequence associated with groups of patterns. The results are shown in Fig. 9. The p-values from this test fit the expected uniform distribu-

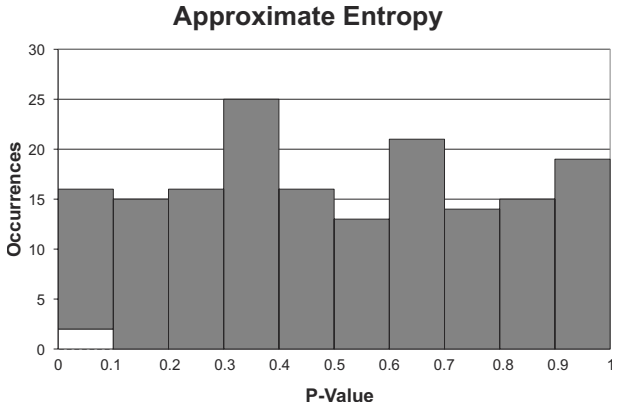

Fig. 9. P-values for the Approximate Entropy Test. P-values for two of the sub-sequences, or 1.18% of them, were below 0.01.

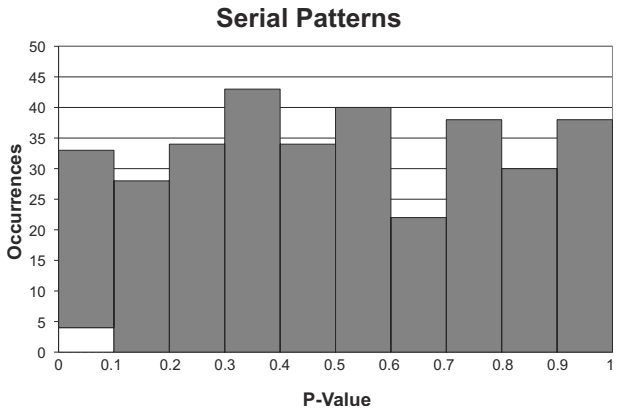

Fig. 10. P-values for the Serial Patterns Test. The test produced two p-values for each sub-sequence. Four of the 340 p-values, or 1.18% of them, were below 0.01.

tion with a reduced χ^2 of 0.784 and a probability of 63% of obtaining a worse fit by chance alone.

H. Serial Patterns Test

This test examines the number of occurrences of all possible patterns of up to three bits in length. Deviations from the expected rates of occurrence of these patterns would indicate non-randomness. The results are shown in Fig. 10. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 1.130 and a probability of 34% of obtaining a worse fit by chance alone.

I. Cumulative Sums Test

This test examines the cumulative sum of the sequence, with every "0" replaced by "–1," from the first bit to each subsequent bit in the sequence. This partial cumulative sum, S_k, represents the total "distance" from the starting point, at each position in the sequence. The largest value that S_k should take is directly related to the length n of the complete sequence [26]. Failures of this test indicate uneven weights of zeros or ones in some part of the sequence. The sums were evaluated going forward from the first bit, and also in reverse order from the last bit, generating two independent p-values for each sequence. The results are shown in Fig. 11. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 0.796 and a probability of 62% of obtaining a worse fit by chance alone.

There were 12 failures of this test at the 0.01 level, significantly more than the 3 or 4 expected. The failure of this test is related to the bias in the source, as observed after data were collected (Section 2), and as detected by the Bias Test (Subsection 3.A).

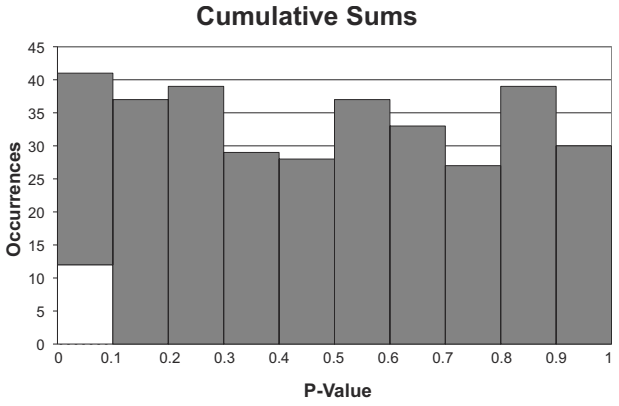

Fig. 11. P-values for the Cumulative Sums Test. The test produced two p-values for each sub-sequence. The average is $p=0.4799$. Twelve of the 340 p-values, or 3.53% of them, were below 0.01.

J. Non-Overlapping Template Match Test

This test checks for the occurrence of a specified "template" pattern, with a focus on aperiodic patterns. The template pattern {0,1,0,1,1,0,0,0,1} was used here, but any would do. The information in the distribution of template occurrences throughout the sub-sequences depends on the template chosen; for the pattern chosen here, too many or too few occurrences would not indicate any problem with bias, for example, but would indicate some unknown process taking place during collection that favors this pattern. The results are shown in Fig. 12. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 0.980 and a probability of 45% of obtaining a worse fit by chance alone.

The remaining tests (Subsections 3.K–3.O) were performed on 17 sequences of 10^6 bits.

K. Random Excursions Test

Like the Cumulative Sums Test (Subsection 3.I), this test concerns the distance of the cumulative sum from the starting point. The test examines the number of times different partial cumulative sums are attained within a single excursion from the starting point, i.e., until the sum reaches zero again. Distances between –4 and 4 were checked. The results are shown in Fig. 13. The p-values from this test fit the expected uniform distribution with a reduced χ^2 of 1.425 and a probability of 17% of obtaining a worse fit by chance alone.

L. Random Excursions Variant Test

Like the previous test, this test also examines the attainment of distances from zero for intermediate cumulative sums. However, it checks the distances –9 through 9, and the occurrence of these distances across all excursions. The results are shown in Fig. 14. The p-values from this

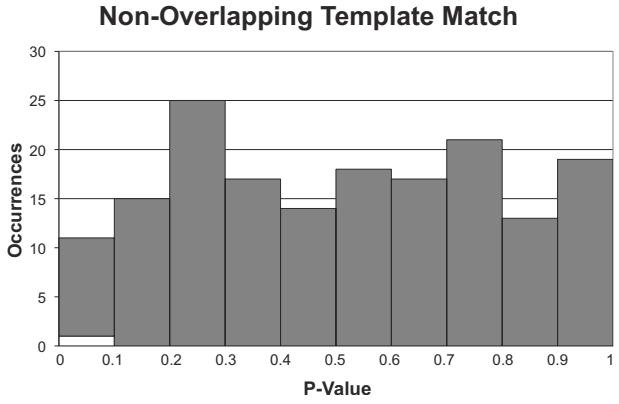

Fig. 12. P-values for the Non-Overlapping Template Match Test. The P-value for one sub-sequence, or 0.588% of them, was below 0.01.

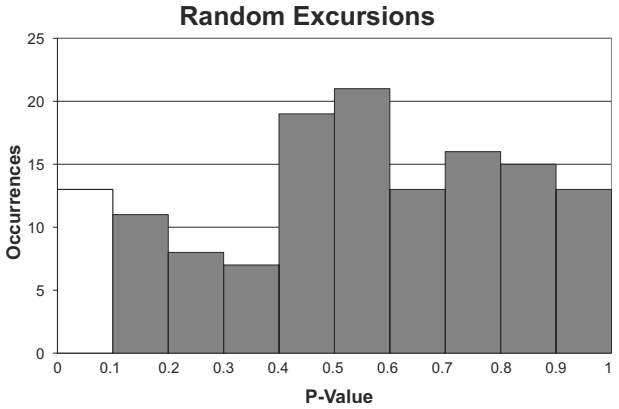

Fig. 13. P-values for the Random Excursions Test. The test generated eight p-values for each of the 17 sub-sequences tested. Thirteen of these 136 p-values, or 9.56% of them, were less than 0.1 (shown in white).

test fit the expected uniform distribution with a reduced χ^2 of 1.258 and a probability of 25% of obtaining a worse fit by chance alone.

M. Overlapping Runs of Ones (Template Match) Test

This test checks for the occurrence of overlapping runs of 1's. In this case, runs of length nine were examined. The results are shown in Fig. 15. Because only 17 p-values were produced (one for each sub-sequence of 10^6 bits), they are shown individually and not as a histogram.

N. Maurer's Universal Statistical Test

This test examines the numbers of bits in between matching bit patterns of various lengths. This property is related to the compressibility of the sequence. It also determines an element of the cryptographic security of the sequence, as it is related to the complexity of some deciphering algorithms. Failure of this test would indicate a sequence that is too easily compressible. The p-values for the 17 sub-sequences of 10^6 bits are shown in Fig. 16; again, they are shown individually and not as a histogram.

O. Linear Complexity Test

This test checks the length of the linear feedback shift register (LFSR) of the sequence, which is also related to compression and deciphering techniques. The sequence's LFSR is a complex algorithm that generates the sequence as output. A LFSR can be used as a pseudo-random number generator, if it is sufficiently complex. This test seeks to invert this process, generating the LFSR from the stream and checking its complexity. The 17 resulting p-values are shown in Fig. 17.

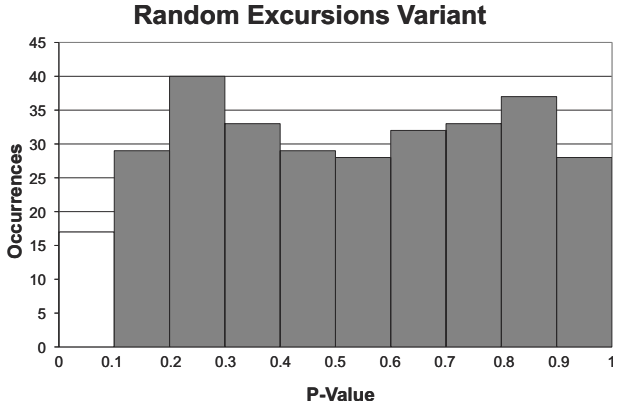

Fig. 14. P-values for the Random Excursions Variant Test. This test generated 18 p-values per sub-sequence. Seventeen of these 306 p-values, or 5.56% of them, were below 0.1.

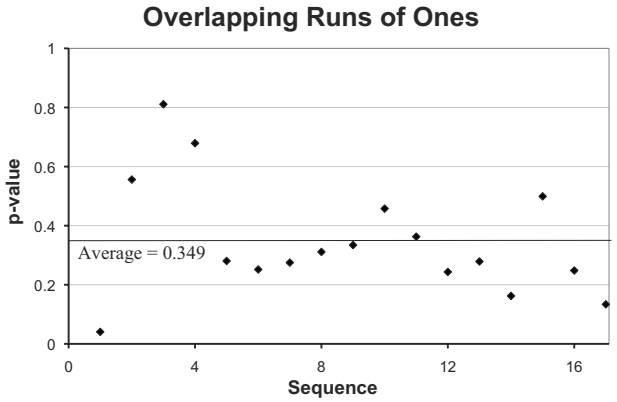

Fig. 15. P-values for the Overlapping Runs of Ones Test. The average is $p=0.349$ (solid line) and the standard deviation is 0.197. Only the first sub-sequence has a p-value below 0.1.

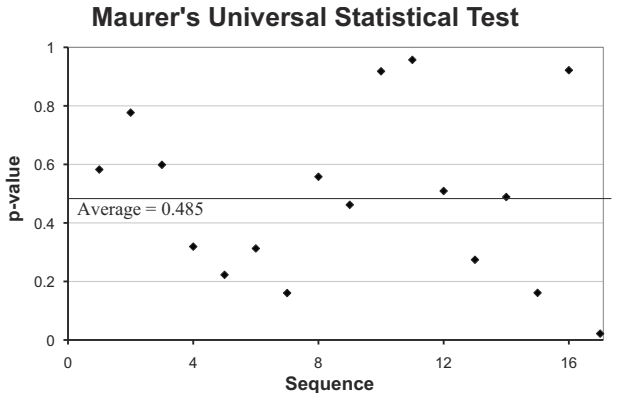

Fig. 16. P-values for Maurer's Universal Statistical Test. The average is $p=0.485$ (solid line) and the standard deviation is 0.286. Only the final sub-sequence (number 17) has a p-value below 0.1.

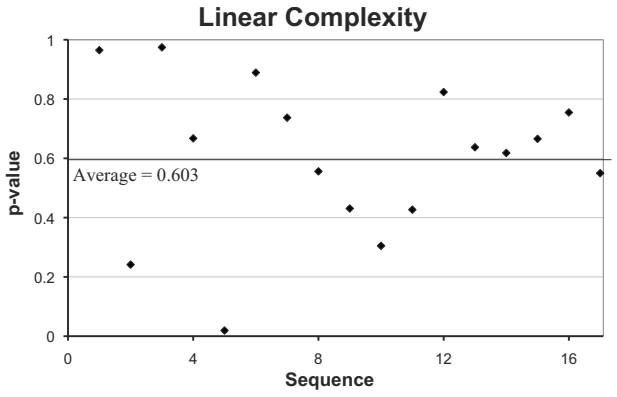

Fig. 17. P-values for the Linear Complexity Test. The average is $p=0.603$ (solid line) and the standard deviation is 0.259. Only sub-sequence 5 produced a p-value below 0.1.

4. ANALYSIS OF RESULTS

A. Cumulative *p*-Value Statistics

The entire test suite produced 4573 p-values in total. For a given confidence threshold α, the proportion of p-values passing each test is expected to be [18]

$$f = (1-\alpha) \pm 3\sqrt{\alpha(1-\alpha)/s}, \tag{2}$$

where s is the number of sequences tested. The two sets of tests, with $s=170$ for the first ten and $s=17$ for the final five, yielded different ranges for the expected passing proportions. Figures 18 and 19 show the passing proportions for each of the first ten tests and the final five tests, respectively.

Another check can be performed by matching the complete set of p-values to their expected distribution. For a completely random process, the p-values from all of the tests should be distributed uniformly on the interval (0,1] with the following characteristics:

$$\langle x \rangle = \int_0^1 xp(x)dx = \frac{1}{2}, \tag{3}$$

$$\langle x^2 \rangle = \int_0^1 x^2 p(x)dx = \frac{1}{3}, \tag{4}$$

$$\sigma_x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = 0.289. \tag{5}$$

A histogram of all 4573 p-values is shown in Fig. 20. The p-values conform quite well to properties of Eqs. 3–5, with an average of 0.501 and a standard deviation of 0.288. These results were also fit to a uniform distribution with an average of 457.3 p-values in each bin (of width $\Delta p = 0.1$). The comprehensive set of p-values fit this "reference" distribution with a reduced χ^2 of 0.967 and a probability of 47% of obtaining a worse fit by chance alone.

B. Bias-Related Failures

Only two of the tests, the Bias Test (Subsection 3.A) and the Cumulative Sums Test (Subsection 3.I) showed a number of failures at the 0.01 confidence level that was significantly higher than expected by chance alone (see Fig. 18). The two tests are related, as a heavily biased sequence will result in unacceptably large (positive or negative) cumulative sums throughout.

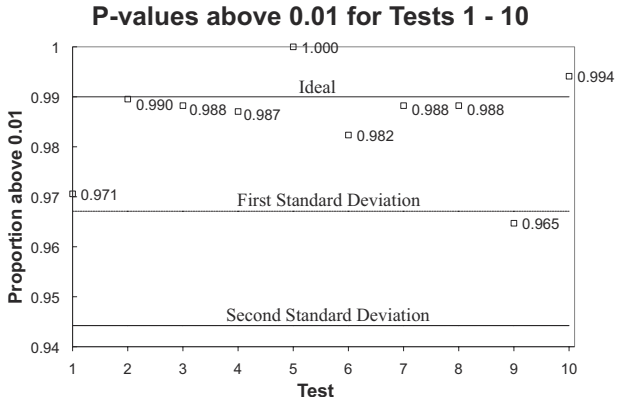

Fig. 18. Proportion of 170 sequences passing the first ten tests at a confidence threshold of 0.01. The outlying points for tests 1 and 9 are due to the bias of the source.

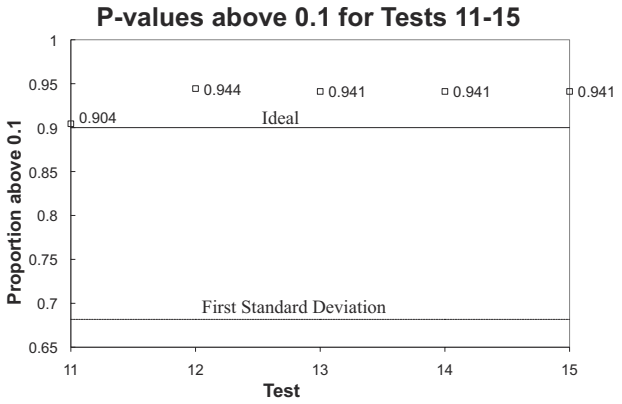

Fig. 19. Proportion of 17 sequences passing the final five tests at a confidence threshold of 0.1.

In the apparatus, the weighting of "0" and "1" events was controlled by the orientation angle θ of the half-wave plate used to set the diagonal polarization of the idler photons under test. The probabilities of reflection P_R and transmission P_T of a photon at the polarizing beamsplitter are given by Malus' Law:

$$P_R = 1 - P_T = \cos^2(2\theta) \tag{6}$$

with the ideal setting being $\theta = \pi/8$ for an unbiased sequence, $P_R = P_T = 1/2$. In the neighborhood around this ideal setting, the rate of change of P_R is

$$\frac{dP_R}{d\theta} = \frac{d}{d\theta}[\cos^2(2\theta)]_{\theta=\pi/8} = -2. \tag{7}$$

Thus, to ensure a bias of less than 1%, or

$$\delta P_R \simeq \delta\theta \frac{dP_R}{d\theta} \leqslant 0.01, \tag{8}$$

the half-wave plate must be oriented to within $\delta\theta \leqslant 0.005$ radians (0.3°) of the ideal setting.

In practice, bias in the polarization measurement sequence also arises from inequalities in the efficiencies of the H and V collection channels for the idler photons. Any mismatch in the cumulative transmission and detection efficiencies of the lenses, fibers, filters, and detectors in

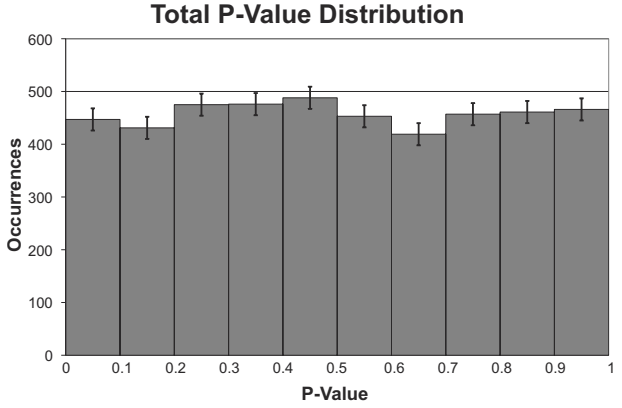

Fig. 20. Histogram of all p-values generated from the NIST Test Suite. Ideally, the p-values follow a uniform distribution.

these channels will show up as a difference in the rates of detected AB and AB′ events.

Because this efficiency mismatch could not be controlled in the apparatus, the setting of θ was used to compensate for it. With the pump laser at maximum power (50 mW), the setting of θ that best equalized the rates of detected AB and AB′ events to within their statistical uncertainties was chosen. The pump laser power was then reduced to give a low time-bin occupation number μ.

In this way, the bias was made low enough for the entire NIST test suite to be applied, but not low enough to pass the Bias and Cumulative Sums tests directly. In principle, the bias could be made lower by measuring a larger number of counts in the AB and AB′ channels (using increased pump power and/or longer integration times) and by scanning θ (possibly under computer control) until they are made equal. The ultimate lower limit of the bias would be determined by the scanning resolution for θ, and by the fluctuations in the larger numbers of AB and AB′ counts.

C. Comparison with Other Quantum-Optical Randomness Tests

With the exception of the Bias Test and the Cumulative Sums Test, the passing p-value proportions in Figs. 18 and 19 were all consistent with the hypothesis of randomness, and with p-values from the NIST test suite performed on another recently implemented optical quantum random-number generator (QRNG) [15]. For that source, the random variable was constructed from the detection times of photons incident on a high-speed gated photodiode, resulting in a net random bit rate of 4 MHz, and was submitted for testing without numerical unbiasing. The larger number of bits obtained for testing in that case (5×10^8) permitted a confidence level of 0.01 to be chosen even for tests 11–15, while in our case a confidence level of only 0.1 could be assigned for those tests. Another recent optical QRNG based on time of arrival had net random bit rates of 1 MHz without unbiasing [14]. Here again a much larger number of bits (10^9) was submitted to the NIST test suite (among others) and reported to have "passed all of the tests" without reporting the confidence level or the passing p-value proportions. Some other recent QRNG's based on photon arrival times [12,13,17] have achieved net random bit rates of 100 kHz, 80 kHz, and 1 MHz, respectively, with numerical unbiasing required. The NIST test suite was not used in these cases.

A QRNG similar to the one reported here, based on polarization measurements of single photons from downconversion, was reported in [16], with a net bit rate of 230 Hz. Although the NIST test suite was not used, several other randomness tests were performed, and passed, albeit with a smaller number of bits (7×10^5). The bits from that source were subjected to numerical unbiasing procedures before being tested.

5. CONCLUSION

For all but two of the tests, the p-values exhibited failure rates and uniform distributions that were consistent with those expected for events from an ideal random source. The exceptions to this were the two tests that were most sensitive to the 0.04% bias of the source: the Bias Test and the Cumulative Sums Test. Source bias of this type is also present in some other quantum optical random-number generators [12,13,16,17], and computational techniques may be used to construct a shorter unbiased sequence from the original [27,28]. In our case, the bias of the raw sequence was low enough to allow direct analysis of the polarization measurements with the NIST tests, with no unbiasing procedure required. To our knowledge, this is the first comprehensive set of such tests performed directly on a sequence of photon polarization measurements.

A weakness of these results stems from the relatively high value of the average time-bin occupation number, $\mu=0.36$, of the source. Optimally, quantum cryptographic schemes operate with a μ of 0.1 or less, to reduce the likelihood of two or more photons occupying the same time bin [4,5]. For the data set reported here, approximately one in five of the occupied time bins had to be rejected for this reason (see Appendix A). This suggests that improvements can be made in at least two ways: by lowering μ significantly, so that these double-events are made more rare, and by applying randomness tests that go beyond the NIST suite to explicitly incorporate the double events that occur.

APPENDIX A: REJECTION OF DOUBLE-COINCIDENCE EVENTS

Approximately one in five of the occupied time bins contained two coincidence events, either of the "mixed" (both an H and a V idler outcome) or "repeat" (2 H outcomes, or 2 V outcomes) varieties. For the mixed cases, since the time ordering of the events cannot be established, it is impossible to assign them as "01" or "10" in the binary sequence. But for the repeat cases, it might be considered possible to include them in their correct positions in the sequence, as "00" or "11," respectively. However, when this was done for the 170 subsequences of length 10^5 bits, it caused over half of them to fail the Bias Test, precluding any further testing within the NIST suite.

A careful examination of the Bias Test reveals why this was so: the test starts by taking the sum s of the entire sequence of n bits, with −1's in place of 0's. The bits of opposite value cancel each other out, so that s is the remaining number of "excess bits" in either direction. For a source with a consistent bias, s should grow with $\sqrt{n}$, so that if s is normalized by $\sqrt{2n}$ (to account for bias in either direction), it can be used to generate a p-value according to [18]

$$p = erfc\left(\frac{|s|}{\sqrt{2n}}\right) = erfc(z). \tag{A1}$$

This form of the p-value takes into account the addition of excess bits as the length of the sequence increases, assuming that they are added in a consistently biased way due to the characteristics of the source. For a sequence of single events from a biased source, e.g. $\{0,0,1,0,1,1,\ldots.n\}$, s will grow larger as n increases, but the normalized "source bias"

$$z = \frac{|s|}{\sqrt{2n}} \tag{A2}$$

will not change, so that the p-value expressing the bias of a given source will remain constant for any length of sequence that is tested.

On the other hand, a sequence of n repeat events from the same source will have twice the number of bits, e.g., $\{00,11,11,00,\ldots.2n\}$, and will also generate twice the number of excess bits so that s becomes $2s$. This gives an apparent source bias of

$$z_{repeat} = \frac{|2s|}{\sqrt{2(2n)}} = \frac{2}{\sqrt{2}}\frac{|s|}{\sqrt{2n}} = \sqrt{2}\cdot z. \tag{A3}$$

That is, because each independent repeat event adds two identical outcomes to the sequence, the number of excess bits $2s$ is artificially high compared to that of a sequence of $2n$ independent single events. Taken by themselves, the repeat events do not appear to come from the same source, but rather, from one with a normalized bias that is larger by a factor of $\sqrt{2}$.

Therefore, unless the source is perfectly unbiased ($z=0$), the repeat events cannot be included in the sequence without changing the outcome of the Bias Test. Any initial source bias $z>0$ for the single events will be increased toward $\sqrt{2}\cdot z$ as more of the repeat events are inserted.

Note that if we add n mixed double events of *unknown* time order (01 or 10) to n repeat events, then the normalized source bias for this sequence of $2n$ double events (of total length $4n$ bits) becomes

$$z_{repeat+mixed} = \frac{|2s|}{\sqrt{2(4n)}} = \frac{|s|}{\sqrt{2n}} = z. \tag{A4}$$

The mixed doubles add length, but by definition they cannot add excess bits. Therefore, if all double events—mixed and repeated—are included in the sequence, there is no net effect on the bias of the sequence. But the mixed doubles cannot be included here because their time order is not known; therefore, all double events must be rejected.

ACKNOWLEDGMENTS

We thank Mark Silverman for helpful discussions, and Wayne Strange and Adam Katcher for help with the analysis.

REFERENCES AND NOTES

1. S. Scheel, "Single-photon sources—an introduction," J. Mod. Opt. **56**, 141–160 (2009).
2. L. Mandel, "Quantum effects in one-photon and two-photon interference," Rev. Mod. Phys. **71**, S274–S282 (1999).
3. A. Zeilinger, "Experiment and the foundations of quantum physics," Rev. Mod. Phys. **71**, S288–S297 (1999).
4. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," Rev. Mod. Phys. **74**, 145–195 (2002).
5. V. Scarani, H. Bechmann-Pasquinucci, N. Cerf, M. Dusek, N. Lütkenhaus, and M. Peev, "The security of practical quantum key distribution," Rev. Mod. Phys. **81**, 1301–1350 (2009).
6. A. Steane, "Quantum computing," Rep. Prog. Phys. **61**, 117–173 (1998).
7. A. Galindo and M. A. Martín-Delgado, "Information and computation: classical and quantum aspects," Rev. Mod. Phys. **74**, 347–423 (2002).
8. T. Erber, "Testing the randomness of quantum mechanics: nature's ultimate cryptogram?" Ann. N.Y. Acad. Sci. **755**, 748–756 (1995).
9. M. P. Silverman, W. Strange, C. R. Silverman, and T. C. Lipscombe, "Tests of alpha-, beta-, and electron capture decays for randomness," Phys. Lett. A **262**, 265–273 (1999).
10. M. P. Silverman and W. Strange, "Experimental tests for randomness of quantum decay examined as a Markov process," Phys. Lett. A **272**, 1–9 (2000).
11. M. P. Silverman, W. Strange, C. R. Silverman, and T. C. Lipscombe, "Tests for randomness of spontaneous quantum decay," Phys. Rev. A **61**, 042106 (2000).
12. A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, "Optical quantum random number generator," J. Mod. Opt. **47**, 595–598 (2000).
13. H.-Q. Ma, Y. Xie, and L.-A. Wu, "Random number generation based on the time of arrival of single photons," Appl. Opt. **44**, 7760–7763 (2005).
14. M. Stipcevid and B. Medved Rogina, "Quantum random number generator based on photonic emission in semiconductors," Rev. Sci. Instrum. **78**, 045104 (2007).
15. J. F. Dynes, Z. L. Yuan, A. W. Sharpe, and A. J. Shields, "A high speed, postprocessing free, quantum random number generator," Appl. Phys. Lett. **93**, 031109 (2008).
16. H.-Q. Ma, S.-M. Wang, D. Zhang, J.-T. Chang, L.-L. Ji, Y.-X. Hou, and L.-A. Wu, "A Random number generator based on quantum entangled photon pairs," Chin. Phys. Lett. **21**, 1961–1964 (2004).
17. T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, "A fast and compact random number generator," Rev. Sci. Instrum. **71**, 1675–1680 (2000).
18. 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 (revised)," Natl. Inst. Stand. Technol. (U. S.) Spec. Publ. 800-22rev1 (2008) http://csrc. nist. gov/groups/ST/toolkit/rng/documentation_software. html.
19. D. C. Burnham and D. L. Weinberg, "Observation of simutaneity in parametric production of optical photon pairs," Phys. Rev. Lett. **25**, 84–87 (1970).
20. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, 1995).
21. C. K. Hong and L. Mandel, "Experimental realization of a localized one-photon state," Phys. Rev. Lett. **56**, 58–60 (1986).
22. D. Branning, S. Bhandari, and M. Beck, "Low-cost coincidence-counting electronics for undergraduate quantum optics," Am. J. Phys. **77**, 667–670 (2009).
23. There is an error in the current NIST publication [18] concerning this test: in section 3.D, although all of the formulae appear to be correct, the last three tables of probabilities (for $M=512$, 1000, and 10000) are not. The source code for the Statistical Test Suite provided by NIST, to the extent that it makes use of these incorrect probabilities, is also in error.
24. S.-J. Kim, K. Umeno, and A. Hasegawa, *On the NIST Statistical Test Suite for Randomness* IEICE Tech. Rep. (IEICE, 2003) Vol. 103, 21–27.
25. S.-J. Kim, K. Umeno, and A. Hasegawa, *Corrections of the NIST Statistical Test Suite for Randomness* Report 2004/018 (Cryptology ePrint Archive, 2004).
26. The current NIST publication [18] also contains an error regarding this test: on page 2–32, Eq. (3), the absolute value of S_k should be divided by $\sqrt{n}$.
27. J. Von Neumann, "Various techniques used in connection with random digits," Nat. Bur. Stand. (U. S.) Appl. Math Series No. 12 (GPO, 1951) pp. 36–38.
28. Y. Peres, "Iterating Von Neumann's procedure for extracting random bits," Ann. Stat. **20**, 590–597 (1992).