Experimental evidence of quantum randomness incomputability

Cristian S. Calude[*] and Michael J. Dinneen[†]
Department of Computer Science, University of Auckland, Private Bag 92019, Auckland, New Zealand

Monica Dumitrescu[‡]
Faculty of Mathematics and Computer Science, University of Bucharest, Str. Academiei 14, RO-010014 Bucharest, Romania

Karl Svozil[§]
Institute for Theoretical Physics, Vienna University of Technology, Wiedner Hauptstrasse 8-10/136, A-1040 Vienna, Austria
(Received 9 April 2010; revised manuscript received 4 June 2010; published 6 August 2010)

In contrast with software-generated randomness (called pseudo-randomness), quantum randomness can be proven incomputable; that is, it is not exactly reproducible by any algorithm. We provide experimental evidence of incomputability—an asymptotic property—of quantum randomness by performing finite tests of randomness inspired by algorithmic information theory.

DOI: 10.1103/PhysRevA.82.022102 PACS number(s): 03.65.Ta, 02.50.Fz, 03.67.Lx, 89.70.Cf

I. QUANTUM INDETERMINACY

The irreducible indeterminacy of individual quantum processes postulated by Born [1–3] implies that there exist physical "oracles," which are capable of effectively producing outputs which are incomputable. Indeed, quantum indeterminism has been proved [4] under some "reasonable" side assumptions implied by Bell-, Kochen-Specker-, and Greenberger-Horne-Zeilinger-type theorems. Yet, as quantum indeterminism is nowhere formally specified, it is important to investigate which (classes of) measurements lead to randomness, what are the reasons for possible distinctions, whether or not the kinds of randomness "emerging" in different classes of quantum measurements are "the same" or "different," and what are the phenomenologies or signatures of these randomness classes. Questions about "degrees of (algorithmic) randomness" are studied in algorithmic information theory. Here are just four types, among an infinity of others: (i) standard pseudo-randomness produced by software such as MATHEMATICA or MAPLE which are not only Turing computable but cyclic; (ii) pseudo-randomness produced by software which is Turing computable but not cyclic (e.g., digits of π, the ratio between the circumference and the diameter of an ideal circle, or Champernowne's constant); (iii) Turing incomputable, but not algorithmically random; and (iv) algorithmically random [5–7]. In which of these four classes do we find quantum randomness? Operationally, in the extreme form, Born's postulate could be interpreted to allow for the production of "random" finite strings; hence quantum randomness could be of type (iv). (Here the quotation marks refer to the fact that randomness for finite strings is too "subjective" to be meaningful for our analysis. The legitimacy of the experimental approach comes from characterizations of random sequences in terms of the degrees of incompressibility of their finite prefixes [5–7].) A sequence which is not algorithmically random but Turing incomputable can, for instance, be obtained from an algorithmically random sequence $x_1x_2\cdots x_n\cdots$ by inserting a 0 in between any adjacent original bits, i.e., obtaining the sequence $x_10x_20\cdots0x_n0\cdots$. This transformation destroys algorithmic randomness because obvious correlations have appeared; Turing incomputability is invariant under this transformation because a copy of the original sequence is embedded in the new one. Yet much more subtler correlations among subsequences of Turing incomputable sequences may exist, thus making them compressible and algorithmically nonrandom. There is no *a priori* reason to interpret Born's indeterminism by its strongest formal expression (i.e., in terms of algorithmic randomness).

Quantum randomness produced by quantum systems which have no classical interpretation can be proven [4] Turing incomputable. More precisely, if the experiment would run under ideal conditions "to infinity," the resulting infinite sequence of bits would be Turing incomputable; that is, no Turing machine (or algorithm) could reproduce exactly this infinite sequence of digits. This result has many consequences. Here is one example: The experiment could produce a billion 0's, but not all bits produced will be 0. A stronger form of incomputability holds true: Every Turing machine (or algorithm) can reproduce exactly only finitely many scattered digits of that infinite sequence. Yet this proof stops short of showing that the sequence produced by such a quantum experiment is algorithmically random; that is, it is unknown whether or not such a sequence is or is not algorithmically random. One of the strategies toward answering this question is to empirically perform tests "against" the algorithmic randomness hypothesis.

Our (more modest) aim is to present tests capable of distinguishing computable from incomputable sources of "randomness" by examining (long, but) finite prefixes of infinite sequences. Such differences are guaranteed to exist by [4], but, because computability is an asymptotic property, there was no guarantee that finite tests can "pick" differences in the prefixes that we have analyzed.

[*]cristian@cs.auckland.ac.nz; http://www.cs.auckland.ac.nz/~cristian
[†]mjd@cs.auckland.ac.nz; http://www.cs.auckland.ac.nz/~mjd
[‡]mdumi@fmi.unibuc.ro; http://fmi.unibuc.ro/ro/dumitrescu_monica
[§]svozil@tuwien.ac.at; http://tph.tuwien.ac.at/~svozil

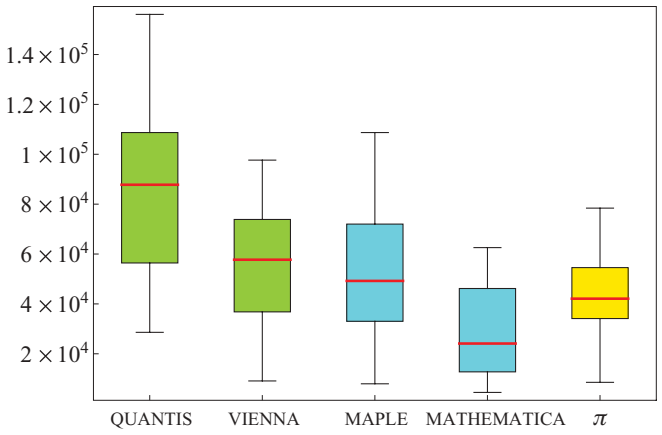

FIG. 1. (Color online) Box-and-whisker plot for the results of the "book stack" randomness test.

II. TESTS OF EXPERIMENTAL QUANTUM INDETERMINACY

Based on Born's postulate, several quantum random number generators employing beam splitters have recently been proposed and realized [8–15]. In what follows a detailed analysis of bit strings of length 2^{32} obtained by two such quantum random number generators will be presented. (The size correlates well with the square root of the cycle length used by cyclic pseudo-random generators; randomness properties of longer strings generated in this way are impaired.) We will compare the performance of quantum random number generators with software-generated number generators on randomness inspired by algorithmic information theory (which complement some commonly used statistical tests implemented in "batteries" of test suites such as, for instance, DIEHARD [16], NIST [17], or TESTU01 [18]). The standard test suites are often based on tests which are not designed for physical random number generators, but rather to quantify the quality of the cyclic pseudo-random numbers generated by algorithms. As we would like to separate "truly" random sequences from software-generated random sequences, the emphasis is on the former type of tests.

The tests based on algorithmic information theory directly analyze randomness and thus the strongest possible form of incomputability. They differ from tests employed in the standard randomness batteries as they depend on irreducible algorithmic information content, which is constant for algorithmic pseudo-random sequences. Some tests are related to each other, as for instance sequences which are not Borel normal (cf. the following) could be algorithmically compressed; the analysis of results helps understand subtle differences at the edge of incomputability or algorithmic randomness. All tests depend on the size of the analyzed strings; the legitimacy of our approach is given by the fact that algorithmic randomness of an infinite sequence can be "uniformly read" in its prefixes (cf. [7]).

III. DATA SOURCES

The analyzed quantum data consist of 10 quantum random strings generated with the commercially available QUANTIS device [19], based on research of a group in Geneva [11], as well as 10 quantum random strings generated by the Vienna Institute for Quantum Optics and Quantum Information (IQOQI) group [20]. The pseudo-random data consist of 10 pseudo-random strings produced by MATHEMATICA 6 [21], and 10 pseudo-random strings produced by MAPLE 11 [22], as well as 10 strings of 2^{32} bits from the binary expansion of π obtained from the University of Tokyo's supercomputing center [23].

The signals of the QUANTIS device are generated by a light-emitting diode (LED) producing photons which are then transmitted toward a beam splitter (a semitransparent mirror) and two single-photon detectors (detectors with single-photon resolution) to record the outcomes associated with the symbols "0" and "1," respectively [19]. Due to hardware imbalances which are difficult to overcome at this level, QUANTIS processes these raw data by unbiasing the sequence by a von Neumann–type normalization: The biased raw sequence of zeros and ones is partitioned into fixed subsequences of length two; then the even-parity sequences "00" and "11" are discarded, and only the odd parity ones "01" and "10" are kept. In a second step, the remaining sequences are mapped into the single symbols $01 \mapsto 0$ and $10 \mapsto 1$, thereby extracting a new unbiased sequence at the cost of a loss of original bits ([24], p. 768).

This normalization method requires that the events are (temporally) uncorrelated and thus independent. (For the sake of a simple counterexample, the von Neumann normalization of the sequences $010101\cdots$ or $1100110011\cdots$ are the constant-0 sequence $000\cdots$ and the empty sequence.) Under the independence hypothesis, the normalized sequences are Borel normal with probability one [25]; e.g., all finite subsequences of length n occur with their expected asymptotic frequencies 2^{-n}. (Alas, see [26] for some pitfalls when transforming such sequences.)

The signals of the Vienna IQOQI group were generated with photons from a weak blue LED light source, which impinged on a beam splitter without any polarization sensitivity with two output ports associated with the codes "0" and "1," respectively [10]. There was *no* pre- or post-processing of the raw data stream, in particular no von Neumann normalization as discussed for the QUANTIS device; however, the output was

TABLE I. Statistics for the results of the "book stack" randomness test.

	Minimum	$Q1$	Median	$Q3$	Maximum	Mean	Standard deviation
MAPLE	796 4	344 90	492 20	696 30	108 700	534 10	330 68.58
MATHEMATICA	450 8	130 20	241 10	434 50	625 70	279 40	194 06.03
QUANTIS	286 00	604 80	877 80	106 700	156 100	899 90	415 45.76
VIENNA	911 0	384 20	577 20	732 20	976 60	538 60	279 38.92
π	855 1	354 80	421 00	528 70	784 10	412 80	207 58.46

TABLE II. Statistics for the results based on the Solovay-Strassen probabilistic primality test.

	Minimum	$Q1$	Median	$Q3$	Maximum	Mean	Standard deviation
MAPLE	93.0	96.0	101.0	113.5	120.0	104.9	10.577 23
MATHEMATICA	93.0	97.0	109.0	132.3	142.0	113.5	19.608 67
QUANTIS	99.0	103.3	113.0	121.3	130.0	112.6	10.668 75
VIENNA	82.0	100.3	104.5	109.0	119.0	103.5	11.037 81
π	84.0	91.8	106.0	110.8	128.0	104.7	10.668 75

constantly monitored (the exact method being subject to a pending patent). In very general terms, the setup needs to be running for at least one day to reach a stable operation. There is a regulation mechanism which keeps track of the bias between "0" and "1" and tunes the random generator for perfect symmetry. Each data file was created in one continuous run of the device lasting over hours.

We have employed the extended cellular automaton generator default of MATHEMATICA 6's pseudo-random function. It is based on a particular five-neighbor rule, so each new cell depends on five nonadjacent cells from the previous step [21]. MAPLE 11 uses a Mersenne Twister algorithm to generate a random pseudo-random output [22].

IV. TESTING INCOMPUTABILITY AND RANDOMNESS

The tests we performed can be grouped into (i) two tests based on algorithmic information theory, (ii) statistical tests involving frequency counts (Borel normality test), (iii) a test based on Shannon's information theory, and (iv) a test based on random walks.

In Figures 1–5 the graphical representation of the results is rendered in terms of box-and-whisker plots, which characterize groups of numerical data through five characteristic summaries: test minimum value, first quantile (representing one fourth of the test data), median or second quantile (representing half of the test data), third quantile (representing three fourths of the test data), and test maximum value. Mean and standard deviation of the data representing the results of the tests are calculated. Tables containing the experimental data and the programs used to generate the data can be downloaded from our extended paper [27].

A. Book stack randomness test

The *book stack* (also known as "move to front") test [28,29] is based on the fact that compressibility is a symptom of less randomness.

The results, presented in Fig. 1 and Table I, are derived from the original count, the count after the application of the transformation, and the difference. The key metric for this test is the count of ones after the transformation. The book stack encoder does not compress data but instead rewrites each byte with its index from the top (front) with respect to its input characters being stacked (moved to front). Thus, if a lot of repetitions occur (i.e., a symptom of nonrandomness), then the output contains more zeros than ones due to the sequence of indices generally being smaller numerically.

B. Solovay-Strassen probabilistic primality test

The second algorithmic test, based on the Solovay-Strassen probabilistic primality test, uses Carmichael (composite) numbers, which are "difficult" to factor, to determine the quality of randomness by computing how fast the probabilistic primality test reaches the verdict "composite" [30,31].

To test whether a positive integer n is prime, we take k natural numbers uniformly distributed between 1 and $n-1$, inclusive, and, for each chosen i, check whether the predicate $W(i,n)$ holds. If this is the case we say that "i is a witness of n's compositeness." If $W(i,n)$ holds for at least one i then

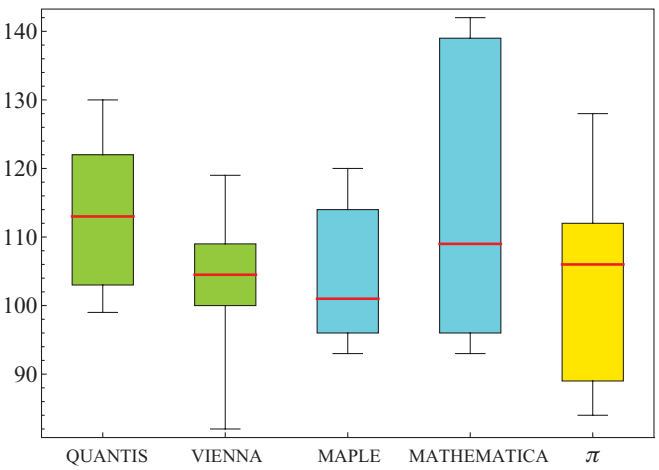

FIG. 2. (Color online) Box-and-whisker plot for the results based on the Solovay-Strassen probabilistic primality test.

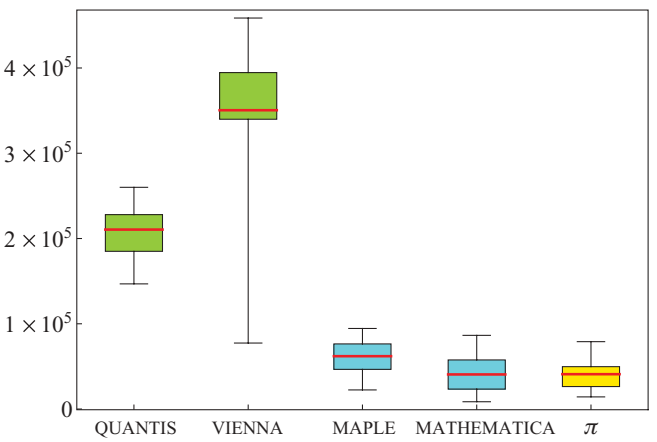

FIG. 3. (Color online) Box-and-whisker plot for the results for tests of the Borel normality property.

TABLE III. Statistics for the results for tests of the Borel normality property.

	Minimum	$Q1$	Median	$Q3$	Maximum	Mean	Standard deviation
MAPLE	224 30	471 70	619 90	761 30	945 10	602 10	219 33.52
MATHEMATICA	8572	255 00	405 90	556 50	864 30	418 70	232 29.77
QUANTIS	146 800	185 100	210 500	226 600	260 000	207 200	335 15.65
VIENNA	774 10	340 200	350 500	392 500	260 000	337 100	103 354.3
π	142 60	288 60	408 80	478 60	790 30	402 20	179 06.21

n is composite; otherwise, the test is inconclusive, but in this case if one declares n to be prime then the probability of being wrong is smaller than 2^{-k}.

This is because at least half the i values from 1 to $n-1$ satisfy $W(i,n)$ if n is indeed composite, and *none* of them satisfy $W(i,n)$ if n is prime [30]. Selecting k natural numbers between 1 and $n-1$ is the same as choosing a binary string s of length $n-1$ with k 1's such that the ith bit is 1 if and only if i is selected. Reference [31] contains a proof that, if s is a long enough algorithmically random binary string, then n is prime if and only if $Z(s,n)$ is true, where Z is a predicate constructed directly from conjunctions of negations of W.[1]

A Carmichael number is a composite positive integer k satisfying the congruence $b^{k-1} \equiv 1 (\mathrm{mod}\, k)$ for all integers b relative prime to k. Carmichael numbers are composite, but they are difficult to factorize and thus are "very similar" to primes; they are sometimes called pseudo-primes. Fermat's primality test declares significantly more Carmichael numbers as primes than the Solovay-Strassen test. With increasing values, Carmichael numbers become "rare."[2]

We used the Solovay-Strassen test for all Carmichael numbers less than 10^{16}—computed in Refs. [32,33]—with numbers selected according to increasing prefixes of each sample string till the algorithm returns a nonprimality verdict. The metric is given by the length of the sample used to reach the correct verdict of nonprimality for all of the 246 683 Carmichael numbers less than 10^{16}. [We started with $k=1$ tests (per each Carmichael number) and increase k until the metric goal is met; as k increases we always use new bits (never recycling) from the sample source strings.] The results are presented in Fig. 2 and Table II.

C. Borel normality test

Borel normality—requesting that every binary string appears in the sequence with the correct probability 2^{-n} for a string of length n—served as the first mathematical definition of randomness [25]. A sequence is (Borel) normal if every binary string appears in the sequence with the right probability (which is 2^{-n} for a string of length n). A sequence is normal if and only if it is incompressible by any information lossless finite-state compressor [35], so normal sequences are those sequences that appear random to any finite-state machine.

Every algorithmic random infinite sequence is Borel normal [36]. The converse implication is not true: There exist computable normal sequences (e.g., Champernowne's constant).

Normality is invariant under finite variations: Adding, removing, or changing a finite number of bits in any normal sequence leaves it normal. Further, if a sequence satisfies the normality condition for strings of length $n+1$, then it also satisfies normality for strings of length n, but the converse is not true.

[1] In fact, every "decent" Monte Carlo simulation algorithm in which tests are chosen according to an algorithmic random string produces a result which is not only true with high probability but *rigorously correct* [34].

[2] There are 1,401,644 Carmichael numbers in the interval $[1,10^{18}]$.

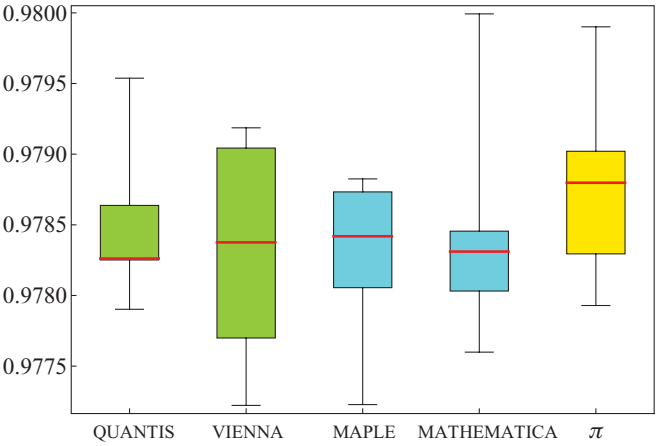

FIG. 4. (Color online) Box-and-whisker plot for average results in "sliding window" estimations of the Shannon entropy.

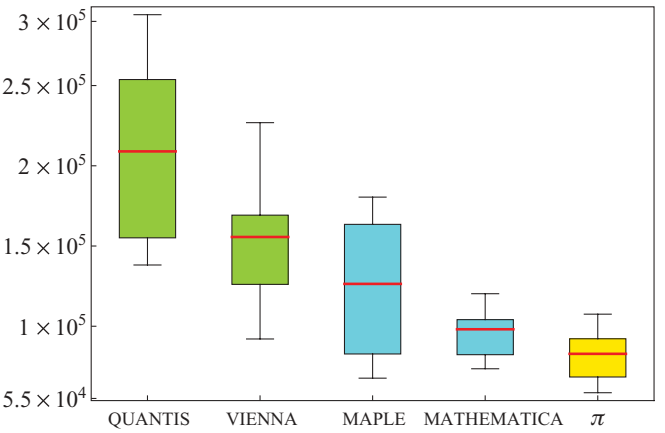

FIG. 5. (Color online) Box-and-whisker plot for the results of the random walk tests.

TABLE IV. Statistics for average results in "sliding window" estimations of the Shannon entropy.

	Minimum	$Q1$	Median	$Q3$	Maximum	Mean	Standard deviation
MAPLE	0.977 2	0.978 1	0.978 4	0.978 7	0.978 8	0.978 3	0.000 523 161 7
MATHEMATICA	0.977 6	0.978 1	0.978 3	0.978 5	0.980 0	0.978 3	0.000 665 493 6
QUANTIS	0.977 9	0.978 3	0.978 3	0.978 6	0.979 5	0.978 4	0.000 452 269 9
VIENNA	0.977 2	0.977 7	0.978 4	0.979 0	0.979 2	0.978 3	0.000 695 583 4
π	0.977 9	0.978 4	0.978 8	0.979 0	0.979 9	0.978 8	0.000 606 272 4

Normality was transposed to strings in Ref. [36]. In this process one has to replace limits with inequalities. As a consequence, these two properties, which are valid for sequences, are no longer true for strings.

For any fixed integer $m > 1$, consider the alphabet $B_m = \{0,1\}^m$ consisting of all binary strings of length m, and for every $1 \leqslant i \leqslant 2^m$ denote by $N_i^m(x)$ the number of occurrences of the lexicographical ith binary string of length m in the string x (considered over the alphabet B_m). By $|x|_m$ we denote the length of x over B_m; $|x|_1 = |x|$. A string x is Borel normal if for every natural $1 \leqslant m \leqslant \log_2 \log_2 |x|$,

$$\left| \frac{N_j^m(x)}{|x|_m} - 2^{-m} \right| \leqslant \sqrt{\frac{\log_2 |x|}{|x|}},$$

for every $1 \leqslant j \leqslant 2^m$. In Ref. [36] it is shown that almost all algorithmic random strings are Borel normal.

First we count the maximum, minimum, and difference of nonoverlapping occurrences of m-bit ($m = 1, \ldots, 5$) strings in each sample string. Then we test the Borel normality property for each sample string and found that almost all strings pass the test, with some notable exceptions. We found that several of the Vienna sequences failed the expected count range for $m = 2$ and a few of the Vienna sequences were outside the expected range for $m = 3$ and $m = 4$ (with some less then the expected minimum count and some more than the expected maximum count). The only other bit sequence that was outside the expected range count was one of the MATHEMATICA sequences that had too big of a count for $k = 1$. Figure 3 depicts a box-and-whisker plot of the results. This is followed by statistical (numerical) details in Table III.

D. Test based on Shannon's information theory

The next test computes "sliding window" estimations of the Shannon entropy $L_n^1, \ldots, L_n^t$ according to the method described in [37]: A smaller entropy is a symptom of less randomness. The results are presented in Fig. 4 and Table IV.

E. Test based on random walks

A symptom of nonrandomness of a string is detected when the plot generated by viewing a sample sequence as a 1D random walk meanders "less away" from the starting point (both ways); hence the maximum–minimum range is the metric.

The fifth test is thus based on viewing a random sequence as a one-dimensional *random walk,* whereby the successive bits, associated with an increase of one unit *per* bit of the x coordinate, are interpreted as follows: 1 ="move up" and 0 ="move down" on the y axis. In this way a measure is obtained for how far away one can reach from the starting point (either positive or negative) from the starting y value of 0 that one can reach using successive bits of the sample sequence. Figure 5 and Table V summarize the results.

V. STATISTICAL ANALYSIS OF RANDOMNESS TESTS RESULTS

In what follows the significance of results corresponding to each randomness test applied to all five sources are analyzed by means of some statistical comparison tests. The Kolmogorov-Smirnov test for two samples [38] determines whether two datasets differ significantly. This test has the advantage of making no prior assumption about the distribution of data (i.e., it is nonparametric and distribution free).

The Kolmogorov-Smirnov test returns a p value, and the decision "the difference between the two datasets is statistically significant" is accepted if the p value *is less than* 0.05, or, stated pointedly, if the probability of taking a wrong decision is less than 0.05. Exact p values are only available for the two-sided two-sample tests with no ties.

In some cases we have tried to double-check the decision "no significant differences between the datasets" at the price of a supplementary, plausible distribution assumption. Therefore, we have performed the Shapiro-Wilk test for normality [39] and, if normality is not rejected, we have assumed that the datasets have normal (Gaussian) distributions. In order to be able to compare the expected values (means) of the two

TABLE V. Statistics for the results of the random walk tests.

	Minimum	$Q1$	Median	$Q3$	Maximum	Mean	Standard deviation
MAPLE	676 40	887 30	126 400	162 500	180 500	125 300	429 95.59
MATHEMATICA	735 00	847 60	981 10	103 400	120 300	964 50	146 85.34
QUANTIS	138 200	161 600	209 000	250 200	294 200	211 300	559 60.23
VIENNA	920 70	130 200	155 600	167 600	226 900	152 900	367 17.55
π	585 70	704 20	828 00	919 20	107 500	821 20	148 33.75

TABLE VI. Kolmogorov-Smirnov test p values for the "book stack" tests.

	MATHEMATICA	QUANTIS	VIENNA	π
MAPLE	0.417 5	0.167 8	0.994 5	0.417 5
MATHEMATICA		**0.002 1**	0.167 8	0.417 5
QUANTIS			0.167 8	**0.012 3**
VIENNA				0.417 5

samples, the Welch t-test [40], which is a version of Student's test, has been applied. In order to emphasize the relevance of p values less than 0.05 associated with Kolmogorov-Smirnov, Shapiro-Wilk, and Welch's t-tests, they are printed in boldface and discussed in the text.

A. Book stack randomness test

The results of the Kolmogorov-Smirnov test associated with the "book-stack" tests are enumerated in Table VI. Statistically significant differences are identified for QUANTIS versus MATHEMATICA and π.

As more compression is a symptom of less randomness, the corresponding ranking of samples is as follows: $\langle \text{QUANTIS} \rangle = 899\,88.9 > \langle \text{VIENNA} \rangle = 538\,63.8 > \langle \text{MAPLE} \rangle = 534\,11.6 > \langle \pi \rangle = 412\,77.5 > \langle \text{MATHEMATICA} \rangle = 279\,38.3$. The Shapiro-Wilk tests results are presented in Table VII.

Since normality is not rejected for any string, we apply the Welch's t-test for the comparison of means. The results are enumerated in Table VIII. Significant differences between the means are identified for the following sources: (i) QUANTIS versus all other sources (MAPLE, MATHEMATICA, VIENNA, π) and (ii) VIENNA versus MATHEMATICA and MAPLE (as already mentioned).

B. Solovay-Strassen probabilistic primality test

The Kolmogorov-Smirnov test results for this test are presented in Table IX, where no significant differences are detected.

The Shapiro-Wilk test results are presented in Table X. Since there is no clear pattern of normality for the data, the application of Welch's t-test is not appropriate.

C. Borel test of normality

The results of the Kolmogorov-Smirnov test are presented in Table XI.

Statistically significant differences are identified for

(i) QUANTIS versus MAPLE, MATHEMATICA, and π;

(ii) VIENNA versus MAPLE, MATHEMATICA, and π; and

(iii) QUANTIS versus VIENNA.

Note the following:

(1) Pseudo-random strings pass the Borel normality test for comparable, relatively small (with respect to quantum strings; cf. the following) numbers of counts: If the angle brackets $\langle x \rangle$ stand for the statistical mean of tests on x, then $\langle \text{MAPLE} \rangle = 602\,10$, $\langle \text{MATHEMATICA} \rangle = 418\,70$, $\langle \pi \rangle = 402\,20$.

(2) Quantum strings pass the Borel normality test only for "much larger numbers" of counts ($\langle \text{QUANTIS} \rangle = 207\,200$, $\langle \text{VIENNA} \rangle = 337\,100$).

As a result, the Borel normality test detects and identifies statistically significantly differences between all pairs of computable and incomputable sources of "randomness."

TABLE VII. Shapiro-Wilk test p values for the "book stack" tests.

MAPLE	MATHEMATICA	QUANTIS	VIENNA	π
0.788 0	0.481 9	0.723 9	0.814 6	0.517 2

TABLE VIII. Welch's t-test p values for the "book stack" tests.

	MATHEMATICA	QUANTIS	VIENNA	π
MAPLE	0.053 5	**0.043 6**	0.974	0.341 2
MATHEMATICA		**0.000 9**	**0.028 3**	0.155 1
QUANTIS			**0.036 8**	**0.005 4**
VIENNA				0.269 0

D. Test based on Shannon's information theory

The results of the Kolmogorov-Smirnov test are presented in Table XII. No significant differences are detected. The descriptive statistics data for the results of this test indicate almost identical distributions corresponding to the five sources.

The results of the Shapiro-Wilk test associated with a test based on Shannon's information theory are presented in Table XIII. Since there is no clear pattern of normality for the data, the application of Welch's t-test is not appropriate.

E. Test based on random walks

The Kolmogorov-Smirnov test results associated with test based on random walks are presented in Table XIV. Statistically significant differences are identified for (i) QUANTIS versus all other sources (MAPLE, MATHEMATICA, VIENNA, and π); (ii) VIENNA versus MATHEMATICA, VIENNA (as already mentioned), and π; and (iii) MAPLE versus π.

Quantum strings move farther away from the starting point than the pseudo-random strings (i.e., $\langle \text{QUANTIS} \rangle > \langle \text{VIENNA} \rangle > \langle \text{MAPLE} \rangle > \langle \text{MATHEMATICA} \rangle > \langle \pi \rangle$).

It was quite natural to double-check the conclusion "QUANTIS and VIENNA do not exhibit significant differences." Hence we run the Shapiro-Wilk test, which concludes that normality is not rejected (cf. Table XV).

Next, we apply the Welch's t-test for the comparison of means. The results are given in Table XVI. Significant

TABLE IX. Kolmogorov-Smirnov test p values for the Solovay-Strassen tests.

	MATHEMATICA	QUANTIS	VIENNA	π
MAPLE	0.759 1	0.400 5	0.759 1	0.759 1
MATHEMATICA		0.759 1	0.759 1	0.759 1
QUANTIS			0.400 5	0.759 1
VIENNA				0.988 3

TABLE X. Shapiro-Wilk test p values for the Solovay-Strassen tests.

MAPLE	MATHEMATICA	QUANTIS	VIENNA	π
0.069 6	**0.036 3**	0.437 8	0.696 3	0.431 5

TABLE XI. Kolmogorov-Smirnov test p values for the Borel normality tests.

	MATHEMATICA	QUANTIS	VIENNA	π
MAPLE	0.417 5	$<$ **10^{-4}**	**0.000 2**	0.167 8
MATHEMATICA		$<$ **10^{-4}**	**0.000 2**	0.994 5
QUANTIS			**0.000 2**	$<$ **10^{-4}**
VIENNA				**0.000 2**

TABLE XII. Kolmogorov-Smirnov test p values for Shannon's information theory tests.

	MATHEMATICA	QUANTIS	VIENNA	π
MAPLE	0.787 0	0.787 0	0.787 0	0.167 8
MATHEMATICA		0.787 0	0.417 5	0.052 5
QUANTIS			0.417 5	0.167 8
VIENNA				0.417 5

TABLE XIII. Shapiro-Wilk test p values for Shannon's information theory tests.

MAPLE	MATHEMATICA	QUANTIS	VIENNA	π
0.196 2	**0.018 9**	**0.034 5**	0.379 0	0.877 4

TABLE XIV. Kolmogorov-Smirnov test p values for the random walk tests.

	MATHEMATICA	QUANTIS	VIENNA	π
MATHEMATICA	0.167 8	**0.012 3**	0.417 5	0.052 5
QUANTIS		$<$ **10^{-4}**	**0.002 1**	0.167 8
VIENNA			0.052 5	$<$ **10^{-4}**
π				**0.000 2**

TABLE XV. Shapiro-Wilk test p values for the random walk tests.

MAPLE	MATHEMATICA	QUANTIS	VIENNA	π
0.200 6	0.926 8	0.546 4	0.888 8	0.957 7

TABLE XVI. Welch's t-test p values for the random walk tests.

	MATHEMATICA	QUANTIS	VIENNA	π
MAPLE	0.069 61	**0.001 3**	0.140 9	**0.011 9**
MATHEMATICA		$<$ **10^{-4}**	**0.000 7**	**0.043 5**
QUANTIS			**0.014 3**	$<$ **10^{-4}**
VIENNA				**0.000 1**

differences between the means are identified for the following sources: (i) QUANTIS versus all other sources (MAPLE, QUANTIS, VIENNA, and π); (ii) VIENNA versus MATHEMATICA, QUANTIS (as already mentioned), and π; (iii) MAPLE versus π.

VI. SUMMARY

Tests based on algorithmic information theory analyze algorithmic randomness, the strongest possible form of incomputability. In this respect they differ from tests employed in the standard test batteries, as the former depend on irreducible algorithmic information content, which is constant for algorithmic pseudo-random generators. Thus the set of randomness tests performed for our analysis could in principle be expected to be "more sensitive" with respect to differentiating between quantum randomness and algorithmic types of "quasi-randomness" than statistical tests alone.

All tests have produced evidence—with different degrees of statistical significance—of differences between quantum and nonquantum sources:

(a) For the test for Borel normality—the strongest discriminator test—statistically significant differences between the distributions of datasets are identified for (i) QUANTIS versus MAPLE, MATHEMATICA, and π; (ii) VIENNA versus MAPLE, MATHEMATICA, and π; and (iii) QUANTIS versus VIENNA. Not only is the average number of counts larger for quantum sources, but the increase is quite significant: QUANTIS is 3.5–5 times larger than the corresponding average number of counts for software-generated sources, and VIENNA is 5–8 times larger than those values.

(b) For the test based on random walks, statistically significant differences between the distributions of datasets are identified for (i) QUANTIS versus all other sources (MAPLE, MATHEMATICA, VIENNA, and π) and (ii) VIENNA versus MATHEMATICA, VIENNA, and π. Quantum strings move farther away from the starting point than the pseudo-random strings (i.e., $\langle \text{QUANTIS} \rangle > \langle \text{VIENNA} \rangle > \langle \text{MAPLE} \rangle > \langle \text{MATHEMATICA} \rangle > \langle \pi \rangle$).

(c) For the "book-stack" test, significant differences between the means are identified for the following sources: (i) QUANTIS versus all other sources (MAPLE, MATHEMATICA, VIENNA, and π) and (ii) VIENNA versus MATHEMATICA and MAPLE.

(d) For the test based on Shannon's information theory, as well as for the Solovay-Strassen test, *no significant differences* among the five chosen sources are detected. In the first case the reason may come from the fact that averages are the same for all samples. In the second case the reason may be because the test is based solely on the behavior of algorithmic random strings and not on a specific property of randomness.

We close with a cautious remark about the impossibility to formally or experimentally "prove absolute randomness." Any claim of randomness can only be secured *relative* to, and *with respect* to, a more or less large class of laws or behaviors, as it is impossible to inspect the hypothesis against an infinity of—and even less so all—conceivable laws. To rephrase a statement about computability ([41], p. 11), "How can we ever exclude the possibility of our presented, some day (perhaps by some extraterrestrial visitors), with a (perhaps extremely

complex) device that 'computes' and 'predicts' a certain type of hitherto 'random' physical behavior?"

ACKNOWLEDGMENTS

We are grateful to Thomas Jennewein and Anton Zeilinger for providing us with the quantum random bits produced at the University of Vienna by the Vienna IQOQI group, for the description of their method, critical comments, and interest in this research. We thank Alastair Abbott, Hector Zenil, and Boris Ryabko for interesting comments; Ulrich Speidel for his tests for which some partial results have been reported in our extended paper [27]; Stefan Wegenkittl for critical comments of various drafts of this paper and his suggestions to exclude some tests; and the anonymous referees for constructive suggestions. CSC gratefully acknowledges the support of the Hood Foundation and the Vienna University of Technology. KS gratefully acknowledges support of the CDMTCS at the University of Auckland, as well as of the Ausseninstitut of the Vienna University of Technology.

[1] M. Born, Z. Phys. **37**, 863 (1926).
[2] M. Born, Z. Phys. **38**, 803 (1926).
[3] A. Zeilinger, Nature (London) **438**, 743 (2005).
[4] C. S. Calude and K. Svozil, Adv. Sci. Lett. **1**, 165 (2008).
[5] P. Martin-Löf, Inform. Control **9**, 602 (1966).
[6] G. J. Chaitin, *Exploring Randomness* (Springer Verlag, London, 2001).
[7] C. Calude, *Information and Randomness—An Algorithmic Perspective*, 2nd ed. (Springer, Berlin, 2002).
[8] K. Svozil, Phys. Lett. A **143**, 433 (1990).
[9] J. G. Rarity, M. P. C. Owens, and P. R. Tapster, J. Mod. Opt. **41**, 2435 (1994).
[10] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, Rev. Sci. Instrum. **71**, 1675 (2000).
[11] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, J. Mod. Opt. **47**, 595 (2000).
[12] M. Hai-Qiang, W. Su-Mei, Z. Da, C. Jun-Tao, J. Ling-Ling, H. Yan-Xue, and W. Ling-An, Chin. Phys. Lett. **21**, 1961 (2004).
[13] P. X. Wang, G. L. Long, and Y. S. Li, J. Appl. Phys. **100**, 056107 (2006).
[14] M. Fiorentino, C. Santori, S. M. Spillane, R. G. Beausoleil, and W. J. Munro, Phys. Rev. A **75**, 032334 (2007).
[15] K. Svozil, Phys. Rev. A **79**, 054306 (2009).
[16] G. Marsaglia [http://www.stat.fsu.edu/pub/diehard/].
[17] A. Rukhin *et al.*, NIST Special Publ. 800-22 (National Institute of Standards and Technology, Washington DC, 2001) [http://csrc.nist.gov/groups/ST/toolkit/rng/documents/SP800-22b.pdf].
[18] P. L'Ecuyer and R. Simard, ACM Trans. Math. Software (TOMS) **33**, 22 (2007).
[19] ID Quantique SA, QUANTIS (idQuantique, Geneva, Switzerland, 2001–2010) [http://www.idquantique.com/images/stories/PDF/quantis-random-generator/quantis-whitepaper.pdf].
[20] T. Jennewein, Institut für Quantenoptik und Quanteninformation, Quantum random number generator (private communication).
[21] Wolfram Research, Inc., MATHEMATICA Version 6.0 (Wolfram Research, Waterloo, Ontario, 2007) [http://reference.wolfram.com/mathematica/tutorial/RandomNumberGeneration.html].
[22] Maplesoft, MAPLE Version 11 (Maplesoft, Champaign, IL, 2007) [http://www.maplesoft.com/support/help/Maple/view.aspx?path=rand].
[23] Y. Kanada and D. Takahashi, Calculation of π up to 4 294 960 000 decimal digits, University of Tokyo (1995), [ftp://pi.super-computing.org].
[24] J. von Neumann, National Bureau of Standards Applied Math Series **12**, 36 (1951), reprinted in *John von Neumann, Collected Works*, Vol. V, edited by A. H. Traub (MacMillan, New York, 1963), pp. 768–770.
[25] É. Borel, Rendiconti del Circolo Matematico di Palermo **27**, 247 (1909).
[26] P. Hertling, J. Universal Comput. Sci. **8**, 235 (2002).
[27] C. S. Calude, M. J. Dinneen, M. Dumitrescu, and K. Svozil, Report CDMTCS-372 (Centre for Discrete Mathematics and Theoretical Computer Science, University of Auckland, Auckland, New Zealand, 2009), e-print arXiv:0912.4379.
[28] B. Y. Ryabko and A. I. Pestunov, Prob. Peredachi Inf. **40**, 73 (2004).
[29] B. Y. Ryabko and V. A. Monarev, J. Statist. Plan. Inference **133**, 95 (2005).
[30] R. Solovay and V. Strassen, SIAM J. Comput. **6**, 84 (1977). Corrigendum in [42].
[31] G. J. Chaitin and J. T. Schwartz, Commun. Pure Appl. Math. **31**, 521 (1978).
[32] R. G. Pinch, e-print arXiv:math.NT/9803082.
[33] R. G. Pinch, in *Proceedings of Conference on Algorithmic Number Theory 2007. TUCS General Publication No. 46*, edited by A.-M. Ernvall-Hytönen, M. Jutila, J. Karhumäki, and A. Lepistö (Turku Centre for Computer Science, Turku, Finland, 2007), pp. 129–131.
[34] C. Calude and M. Zimand, Int. J. Comput. Math. **16**, 47 (1984).
[35] J. Ziv and A. Lempel, IEEE Trans. Inf. Theory **24**, 530 (1978).
[36] C. Calude, in *Developments in Language Theory*, edited by G. Rozenberg and A. Salomaa (World Scientific, Singapore, 1994), pp. 113–129.
[37] A. D. Wyner, IEEE Information Theory Society (1994).
[38] W. J. Conover, *Practical Nonparametric Statistics* (Wiley, New York, 1999), p. 584.
[39] S. S. Shapiro and M. B. Wilk, Biometrika **52**, 591 (2005).
[40] B. L. Welch, Biometrika **34**, 28 (1947).
[41] M. Davis, *Computability and Unsolvability* (McGraw-Hill, New York, 1958).
[42] R. Solovay and V. Strassen, SIAM J. Comput. **7**, 118 (1978).