How Random Can Quantum Random Number Generators Be?

Matthew Yeh
John A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, MA 01238

(Dated: December 9, 2020)

True random numbers are an important resource in many cryptographic, scientific, and commercial applications. Random number generators based on quantum processes are particularly appealing because they cannot be predicted, even in principle. However, this unpredictability assumes that the device can be trusted and has not been corrupted by an adversary. Thus, it is important to develop random number sources that can be certified, or proven, random. We assess practical trusted quantum random number generators, and compare them to two possible routes toward certifiable randomness: 1) Bell tests, and 2) quantum supremacy. We conclude that recent claims of quantum supremacy may soon lead to certifiable randomness with practically useful generation rates.

I. INTRODUCTION

The probabilistic nature of quantum mechanics has fascinated scientists from the beginning of the theory, with Einstein famously musing that “God does not play dice with the universe.” It defies our everyday intuition, where we might expect that the world is in principle deterministic. Yet, it is precisely these idiosyncracies that give rise to quantum technologies that could reshape communications and information science.

In this review, we will specifically consider how the innate randomness of quantum physics may be exploited in random number generators (RNGs). Randomness plays a key role in numerous areas of science and engineering. For example, both classical and quantum cryptography utilize random numbers to generate keys for encrypting data [1,2]. Randomness is also employed in scientific studies to perform Monte Carlo simulations or fundamental tests of physics. Notably, as first proposed by John Bell, experiments testing his eponymous inequality should choose the measurement basis time-of-flight and randomly in order to close the locality loophole [3-5]. Randomness even finds application in commercial venues such as lotteries. As such, for many years considerable research effort has been devoted to creating high quality sources of random numbers.

To characterize the quality of various RNGs, it is important to distinguish between true randomness and pseudorandomness. While true randomness is difficult to characterize, it is generally associated with unpredictability. In the context of outputting a random bitstring, imagine flipping a perfectly unbiased coin over and over. Each outcome (string of heads and tails) is equally likely, following the statistics of a uniform distribution. In contrast, pseudorandomness by nature is predictable. Typically, a pseudo-RNG (PRNG) begins with an input bitstring known as the “seed,” and then an algorithm based off number theory operates on the seed to output bits that are uniformly distributed and uncorrelated. Although the randomness is only apparent, for many applications this is sufficient. As it is, modern PRNGs are generally well-designed and will successfully spoof statistical test suites designed to check for hidden correlations in RNG outputs.

Ultimately though, the seed completely determines the output sequence of a PRNG. Even with a period of 2^{19937} as in the widely-used Mersenne Twister implementation [6], PRNGs are not cryptographically secure as an adversary could in principle reverse-engineer a PRNG given a sufficently long output sequence. To this end, hardware RNGs are often considered the “gold standard” for producing true random numbers. These devices operate by measuring an unpredictable physical process, either classical or quantum. For example, a classical hardware RNG might use external events such as the time between user keystrokes as the source of randomness [7]. However, such a source is not necessarily well-characterized by physics and is potentially vulnerable to attacks by a well-informed adversary who can manipulate the events to generate a particular outcome. Quantum RNGs (QRNGs) are then particularly appealing because the unpredictability stems from fundamental physical theories – it is native to the process.

We structure our discussion as follows. We first provide a brief overview of different implementations of QRNGs. Given the success of quantum mechanics as a theory, QRNGs would appear to already fulfill the goal of generating true randomness. However, in any realistic scenario, there is always the possibility of classical noise that corrupts the quantum nature of the setup, as well as the possibility of an adversary or malicious manufacturer biasing the device in such a way that renders the QRNG predictable. Thus, we will also consider the growing field of certifiable randomness, which is the idea that one can prove a RNG is random given a few reasonable assumptions. We conclude with a brief discussion on the future of QRNGs, and suggest that recent demonstrations of quantum supremacy could soon lead to high-throughput sources of certifiably random numbers.

II. PRELIMINARIES

Some concepts that are key to any discussion of random number generation are entropy and randomness extractors, which we describe below.

Entropy: Entropy quantifies disorder or randomness. Although there are numerous mathematical definitions of entropy, in general higher entropies imply a greater degree of randomness than lower entropies. One commonly-used entropy is the min-entropy [7].

Definition (min-entropy). For a random variable X with support S, where the support is $S = \{\, x : P(X = x) > 0 \,\}$, the min-entropy is defined as:

$$H_\infty(X) = -\log_2 \max_{x \in S} P(X = x)$$

Min-entropy determines the max number of uniform bits that can be extracted from a distribution. That is, given min-entropy h, the probability of each outcome is bounded by $P(X = x) \leq 1/2^h$ and thus at most h random bits can be extracted. Note that this is a necessary but not sufficient condition to extract h bits.

Randomness Extractors: The output of a QRNG will not necessarily follow a uniform distribution [10]. The role of a randomness extractor is thus to take the outputs of an imperfectly random source and extract a sequence of near uniformly distributed random bits – in other words, to increase the entropy. The earliest example of a randomness extractor is often attributed to von Neumann [11], who showed that a Bernoulli sequence with success probability p can be converted into a uniform sequence by exploiting the fact that $P(10) = P(01) = p(1-p)$. Specifically, his algorithm split the sequence into consecutive, non-overlapping pairs of bits, discarded '00' and '11', and mapped '01' to 0 and '10' to 1.

III. TRUSTED QRNG IMPLEMENTATIONS

Quantum random number generators have a long history. The first implementations can be traced back to the mid-20th century and operated with radioactive decays as the randomness source [8-9]. In recent years, radioactive QRNGs have been phased out in favor of photonic implementations, which benefit theoretically from recent advances in the field of quantum optics and experimentally from high-quality optical instrumentation. In this section, we discuss the historical progression of QRNG implementations, up to its status as one of the most mature quantum technologies.

A. Radioactive Decays

The original QRNGs based on radioactive decays incorporated a radioactive source (typically β decays, i.e. emitted electrons) with a Geiger-Mueller tube for detection. Assuming that the nuclei decay independently and in a one-step process, the number of detector clicks in a time period T can be described by a Poisson distribution

$$P_m(T) = \frac{(\lambda T)^m}{m!} e^{-\lambda T}$$

where m is the number of detector clicks and λ is the Poisson parameter which characterizes the average decay rate. The main distinguishing feature between different implementations is then the method by which the arrival times are converted into random numbers, which can roughly be classified into two methods. In the first method, the internal state of an electronic counter switching between M states is recorded with each detection, producing random numbers modulo M [8]. In the second method, the number of random detection events is counted every T seconds to produce the digits [9]. In both cases, some post-processing has to be done to ensure the distribution of values is uniform.

While effective QRNGs, radioactive decays can only be used in limited scenarios. In particular, for these devices to achieve high throughputs, the randomness source has to be highly radioactive, precluding widespread usage. Moreover, Geiger-Mueller tubes are bulky and require large operating voltages. However, as we will see, in many ways photonic implementations of QRNGs are spiritual successors to these historical devices.

B. Photon Qubits

Conceptually, beamsplitter measurements of photon qubits provide a straightforward path toward QRNGs. If the photon qubit is dual rail-encoded, then a 50:50 beamsplitter may be used; if the qubit is polarization-encoded and linearly polarized at ±45°, then a polarizing beam splitter may be used instead [12, 13]. Detectors placed at the end of each path will then click with equal probability, producing one random bit per event (Fig. 1A). Some of the first optical QRNGs were based on this principle, and in fact such a device forms the basis of ID Quantique's commercial product [14].

However, in practice this first-order model does not account for experimental nonidealities. Specifically, the two detectors required are unlikely to have identical characteristics, resulting in some bias that is difficult to predict, and real detectors have a dead time after a click where the sensitivity is greatly reduced. Although post-processing procedures can asymptotically reduce the bias introduced by instrumentation variation [15], similar to decay-based QRNGs, detector dead time still places a major limitation on the throughput of beamsplitter-based QRNGs especially since only one bit is generated per detection. His-

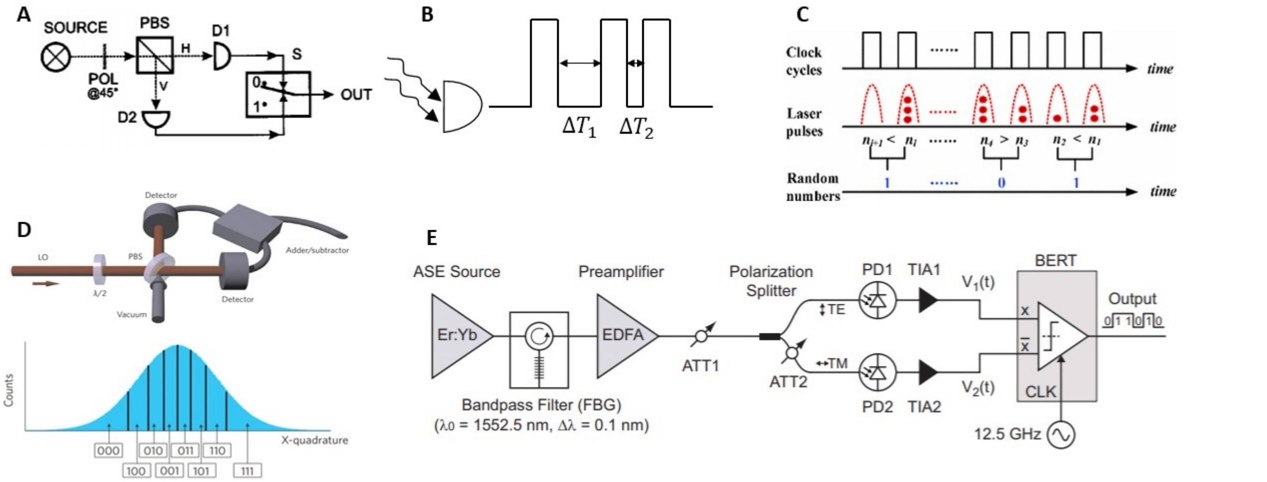

FIG. 1. Operating Mechanism of Selected QRNGs. A) A photon qubit in the $(|H\rangle + |V\rangle)/\sqrt{2}$ state is sent through a polarizing beamsplitter [12]. B) Time difference of arrivals are compared to generate binary bits, or digitized to a certain bit precision to extract multiple bits. Similar principles may be used in decay-based QRNGs. C) Photon numbers between successive clicks are compared [22]. D) Homodyne measurement schematic [25]. The quadrature distribution is binned to generate multiple bits per bin. E) ASE measurement schematic [30]. Noise from a pumped EDFA is filtered, polarization-split, and sent to a comparator.

torically, this led to the development of optical QRNGs based on photon arrival times, which can generate more than one bit per photon and additionally only require a single detector.

C. Photon Arrival Time

Photon-counting statistics indicates that coherent light from an attenuated laser and incoherent light with a sufficiently short coherence time (such that any photon bunching cannot be resolved) will follow a Poisson distribution. Thus, QRNGs based on photon arrival time often take inspiration from decay-based generators.

The first optical QRNG based on temporal information operated by comparing the arrival times successive detection events, e.g. output 0 if $t_1 < t_2$ and 1 if $t_1 > t_2$ (Fig. 1B) [16]. While this demonstration removed the need for two detectors, because the clock only has a finite resolution there was the possibility of the case $t_1 = t_2$ as well as the introduction of unwanted correlations. Thus, post-processing was required which limited bit generation to half a bit per photon. Later work countered this discretization issue by using a highly attenuated laser [17], such that the clock rate was faster than the average photon emission rate, and outputting 1 (0) if the photon was detected on an even (odd) cycle (similar to the method proposed in [8]).

Throughput was further increased by using the time of arrival itself up to k bits of precision, with the first demonstration producing 5.5 bits of entropy per detection [18]. Interestingly, some proposals relax post-processing requirements by carefully driving the laser such that the Poisson distribution becomes roughly uniform [19]. Specifically, if the photon emission rate (Poisson parameter) is time-varying, then the distribution becomes

$$\lambda(t)e^{-\int_a^b \lambda(t')dt'}$$

which is uniform if the photon flux goes as $1/(T - t)$. Improvements in photon-counting electronics resulted in outputs of up to 16 bits/photon, paving the way for another commercial product [20].

D. Photon Counting

Although there exist QRNGs based on photon counting that operate along the lines of [9] and output bits based on the number of detection events within a fixed period, recent work has focused on using photon-number resolving detectors to exploit the photon number statistics of coherent states. Coherent states are quantum states of light that can be described as a superposition of photon number states:

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle$$

They can be generated by a laser, and possess many characteristics that mirror the behavior of classical oscillators. Notably, the probability of detecting n photons is Poissonian with average photon number $\langle n\rangle = |\alpha|^2$.

$$P(n) = e^{-|\alpha|^2} \frac{|\alpha|^{2n}}{n!} = e^{-\langle n\rangle} \frac{\langle n\rangle^n}{n!}$$

Initial proposals compared photon fluxes between subsequent measurements (Fig. 1C), outputting 1 if $n_1 > n_2$, 0 if $n_1 < n_2$, and nothing if $n_1 = n_2$ [21]. The throughput can be slightly increased by assigning two bits to each photon number result, and binning the results such that the distribution is approximately uniform (i.e. 0,1 photons correspond to 00 with probability 25%, 2 photons to 01, and so forth) [22]. Alternatively, some approaches design the photon number distribution to have a large entropy, thereby optimizing the efficiency of randomness extraction [23]. Still, the main difficulty associated with photon counting generators is the limitations of number-resolving detectors, which typically involve spatially multiplexed detectors that cannot distinguish if 2+ photons are coincident on the same pixel.

E. Vacuum State Fluctuations

Another interesting quantum state of light is the vacuum state $|0\rangle$, which can be interpreted as the $n = 0$ level in a quantum harmonic oscillator. Analogous to how the harmonic oscillator ground state still has a finite energy, the vacuum state has fluctuations whose statistics can be exploited in QRNGs. In "position" (amplitude quadrature) space,

$$|0\rangle = \int_{-\infty}^{\infty} \psi(x) \, |x\rangle \, dx$$

where $\psi(x)$ is the ground state wavefunction. This weights the measurement outcomes by a Gaussian probability distribution, so random, unbiased numbers can be generated by binning the measurement outcomes such that all the bins have equal probability, or by using the continuous-valued measurements themselves (up to n bits of precision) [25]. Quadrature measurement is performed using balanced homodyne detection (Fig. 1D). Briefly, vacuum (a blocked beamsplitter port) is interfered with a laser beam at a 50:50 beamsplitter and sent to two detectors, whose difference current is proportional to the quadrature amplitude. Although limited common-mode rejection ratios, electronic noise, and detector bandwidth can corrupt the signal, much of the recent progress in this domain has focused on careful signal processing and randomness extraction to achieve Gbps rates [26-29].

F. Amplified Spontaneous Emission

Optical QRNGs need not be limited to explicit manipulation of quantum states of light. Spontaneous emission that accompanies stimulated emission is very much quantum in origin and a good source of entropy. The benefit of such devices is that amplified spontaneous emission (ASE) is a well-studied phenomenon in optical communications, providing the principal source of noise in erbium-doped fiber amplifiers (EDFAs). Implementations accordingly take advantage of mature fiber amplifier technology [30], pumping an EDFA to generate incoherent broadband ASE, splitting the signal into two polarizations, and then performing clocked comparison of the detector voltages to generate random bits (Fig. 1E). The bit generation rate is primarily limited by filter and detector bandwidths, resulting in bit rates of order Gbps that can be further boosted by extracting multiple bits per measurement [31].

IV. CERTIFIABLE RANDOMNESS

As demonstrated by the examples in the previous section, an overarching challenge is that it is difficult to ascertain how random these QRNGs actually are. In part, this is due to experimental limitations (classical noise, instrumentation variability) that introduce unwanted bias and correlations. Moreover, these devices are "trusted QRNGs", in that it is trusted that the operation is indeed quantum. But in principle, an adversary could bias a beamsplitter, or preset the output sequence such that they retain complete knowledge. Many works claim randomness using statistical test suites, e.g. from NIST [32], but these only provide a necessary rather than sufficient condition. Although there exist proposals that aim to set a compromise between complete trust and complete verification, i.e. only the source or only the detector is trusted [33], at any rate in the worst case it is important to have generators that can be proven random.

In general, certified randomness protocols involve some sort of statistical test that a quantum system/computer can pass while a classical computer cannot. The natural test to use is of course Bell's inequality, which formed the basis of the first (and as of present, only) experimental realizations of certified QRNGs [34-38]. However, spurred by recent claims of quantum supremacy [39, 40], recent proposals have suggested that sampling problems used in these experiments may also provide a path toward certified randomness. Certified random number generation may well be the first practical application of quantum supremacy.

A. Bell Tests

It is helpful to consider a simplified version of how such a QRNG might operate, based on a variant of Bell test called the CHSH game [41, 42]. As illustrated in Figure 2, Alice and Bob are given challenge bits x and y (chosen uniformly at random) and want to produce a and b such that $x + y = xy$. Alice and Bob do not communicate during the game, but can collaborate on a strategy beforehand. The optimal classical strategy is to always choose $a = b = 0$, with win probability 75%. The optimal quantum strategy is more involved. For simplicity, let us consider polarization-encoded qubits. Alice and Bob first share an EPR pair $(|HH\rangle + |VV\rangle)/\sqrt{2}$. Upon receiving challenge bits, Alice measures in the standard basis if $x = 0$ and in the Hadamard basis (polarizer at 45°) if $x = 1$. Similarly, Bob rotates his polarizer to 22.5° if $y = 0$ and to $-22.5°$ if $y = 1$. With this strategy, Alice and Bob win $\cos^2(\pi/8) \approx 85\%$ of the time. If we define

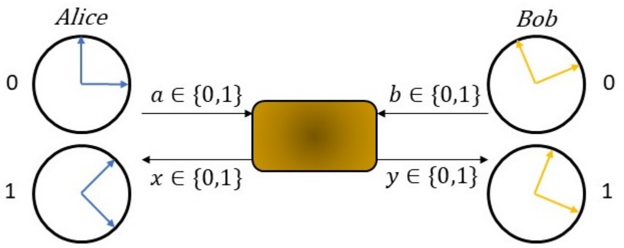

FIG. 2. Diagram of the CHSH Game.

the CHSH score as $J = P(win) - 3/4$, then the corresponding inequality is $J \leq 0$. Violation of this inequality indicates the output bits (a, b) must have some quantum randomness, assuming no-signaling. However, this simple scheme results in a net loss of entropy, since the two output bits are not uniformly distributed whereas the two input bits (x, y) were. The entropy in/entropy out ratio can be increased using a random expansion protocol. In the first experimental demonstration of a Bell-certified RNG, Pironio et al. realized that it suffices to give Alice and Bob one particular (x, y) pair, say $(0, 0)$, with high probability $1 - p$ and the other three possibilities with probability $p/3$ [34]. Intuitively, this is because the process is still random despite the large bias toward one particular challenge bit pair. Since it was assumed that Alice and Bob cannot communicate, whoever received a 0 challenge has no way of discerning between $(0, 1)$, $(1, 0)$ and $(0, 0)$ and thus there is no way to "cheat" using local hidden variables. Employing this protocol, it is possible to obtain approximately quadratic expansion in the entropy.

In spite of this, the throughput is still extremely low due to the nature of the experiment. Entanglement events were heralded by interfering photons emitted by two $^{171}\text{Yb}^{+}$ ions and looking for coincident detections. Occurring on average every 8 minutes, for the 42 certified random bits generated from 3,016 runs this results in a bitrate on the order of μHz [34].

Later theoretical work by Vazirani & Vidick indicated that this efficiency could be improved to exponential randomness expansion using a more sophisticated protocol for nonuniformly distributing challenges [42]. In practice, experimental work usually sets the challenges to always be $(0, 0)$ for simplicity, and the main advance has been using spontaneous parametric downconversion (SPDC) instead of trapped ions to produce entangled photons from a nonlinear crystal [35-38]. Using SPDC, trials can be run at a significantly higher frequency, most recently resulting in a rate of 181 bps in a loophole-free Bell test implementation [38].

B. Quantum Supremacy

Quantum supremacy is the experimental realization of some computational task that cannot be solved in a reasonable amount of time by any classical algorithm. Quantum supremacy was recently claimed by Google using a 53-qubit programmable superconducting processor to sample from the output of random quantum circuits (Random Circuit Sampling) [39], and quantum advantage by Zhong et al. using a photon-based sampling task known as BosonSampling [40]. Given this dichotomy of a quantum computer being able to easily perform a task a classical computer could never do, quantum supremacy suggests an alternate route to certified randomness. Note that at the moment only Random Circuit Sampling (RCS) is known to have a relevant statistical test that can verify randomness of output bitstrings based on strong complexity-theoretic arguments [43].

Heavy Output Generation (HOG): HOG is a statistical test proposed by Aaronson & Chen for verifying quantum supremacy [43]. At a high level, the idea is that the output distribution of RCS is nearly uniform, but some outputs strings are more likely ("heavy"). Aaronson & Chen provided strong evidence that it is exponentially hard for a classical computer to do better than simply outputting the uniform distribution. In contrast, a quantum computer can simply perform RCS. Thus, if someone receives k output bitstrings and $>2/3$ are heavy, i.e. probability greater than the median probability, then the HOG test is passed and it can be concluded that there is some quantumness. In other words, no classical computer can generate the same amount of entropy as RCS in a comparable time. We note that Google used a different benchmark, linear cross-entropy benchmarking (XEB) [39, 44]; however the basic idea is similar and in fact HOG and XEB are specific cases of a more generalized measure [45].

Aaronson has recently proposed that Google's quantum supremacy experiment can be used to generate verifiable randomness from moderately sized quantum computers (50-100 qubits) [44]. As depicted in Figure 3, the proposal is as follows: 1) The client (pseudo)randomly generates quantum circuits $C_1 \ldots C_M$ and sends them to a server with access to a quantum computer, 2) the server quickly performs RCS on those circuits and returns a list of k samples $S = (s_1 \ldots s_k)$ from each distribution $C_i \left|0\right\rangle^n$, 3) for a few randomly chosen iterations, the client performs the HOG test, and 4) if the tests pass, feeds the output bitstrings $S = (S_1 \ldots S_M)$ to a randomness extractor.

Such a proposal could potentially produce certified random bits at a more practically useful rate than Bell-based implementations. Google's processor sampled 53 qubits 10^6 times in 200 seconds, setting a very rough upper bound on their bit generation rate as 265 kbps [39]. In practice, the HOG verification takes place on a classical computer, which would only work for moderately sized systems before the computation takes too long. However, HOG only need to be checked for a few circuits to verify quantumness, so in principle this only introduces

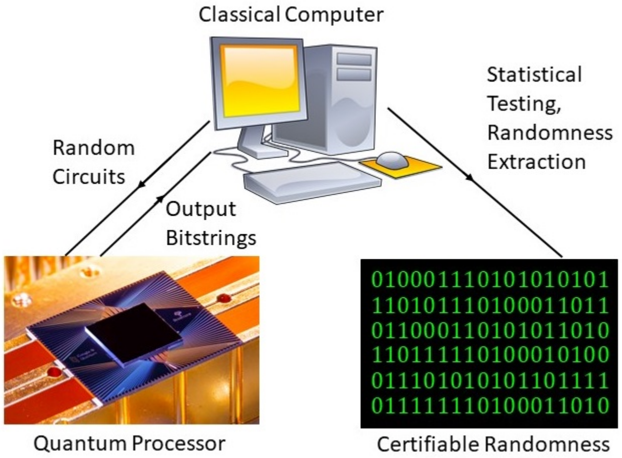

FIG. 3. Proposal for Certified RNG from Quantum Supremacy. A classical computer sends pseudorandom circuits to a quantum computer [39], which sends back RCS bitstrings whose quantum origin can be statistically tested.

a constant overhead. Another limitation that should be studied in greater detail is that Google's processor has a fidelity of 0.2%, so the distribution is 99.8% uniform and may not consistently pass HOG either.

V. DISCUSSION

Table I presents a summary of representative QRNGs discussed in this work. Clearly, many advances have to be made in the field of certified randomness for these generators to be be comparably fast to more common non-certified QRNGs. Still, they are catching up, with Bell-certified QRNGs improving their speed 7 orders of magnitude in the past decade and quantum supremacy proposals showing potential to improve on that another 3 orders.

However, another factor that must be considered is feasibility. In this regard the outlook for Bell certification is quite poor. Beyond improving speed with SPDCs, most advances since the original demonstration by Pironio et al. have been to improve security analysis by closing the detection and locality loopholes using highly efficient detectors and space-like separation [35-38]. These experiments are quite complex and require resources, e.g. 100m of space, that it is unlikely the average user of random bits will have access to. In contrast, while Google's experiment is obviously extremely complex, the certification test in the aforementioned proposal can in principle be performed by the client if they have access to sufficient (classical) computing power, so direct access to a quantum computing setup is not required. Combined with the higher bit generation rate, we expect that if a certified randomness protocol is to take off, it will likely be based on the Aaronson proposal or some variant.

TABLE I. Brief summary of representative QRNG implementations. For certifiable randomness (marked by *), we list all implementations discussed in the manuscript.

Work	Year	Method	Rate
Schmidt [8]	1970	Radioactive Decay	<10 kbps
Jennewein et al. [12]	2000	Photon Qubit	1 Mbps
Stipcevic et al. [16]	2007	Arrival Time	1 Mbps
Wayne et al. [19]	2010	Arrival Time	110 Mbps
Ren et al. [22]	2011	Photon Counting	2.4 Mbps
Applegate et al. [24]	2015	Photon Counting	143 Mbps
Gabriel et al. [25]	2010	Vacuum State	6.5 Mbps
Zheng et al. [29]	2019	Vacuum State	6 Gbps
Williams et al. [30]	2010	ASE	12.5 Gbps
Argyris et al. [31]	2015	ASE	560 Gbps
Pironio et al.* [34]	2010	Bell Test	<29 μbps
Christensen et al.* [35]	2013	Bell Test	0.4 bps
Liu et al.* [36]	2018	Bell Test	114 bps
Bierhorst et al.* [37]	2018	Bell Test	<1.7 bps
Liu et al.* [38]	2018	Bell Test	181 bps
Google, TBD* [39]	TBD	RCS+HOG	<256 kbps

VI. CONCLUSION

In conclusion, we have surveyed recent progress in the field of random number generation. While for most applications PRNGs and trusted QRNGs are sufficient, it is of particular cryptographic interest to produce random numbers that can be proven quantum in origin, and thus cannot be predicted even in principle. The first protocols developed and experimentally demonstrated were naturally based off Bell tests, providing truly random bits albeit at low generation rates. It remains an open problem to develop devices with throughputs more comparable to standard trusted QRNGs. To this end, future experimental work could revolve around sampling tasks originally designed for quantum supremacy demonstrations. As a promising path toward faster certified bit rates and as the first practical application of quantum supremacy, this would be quite remarkable.

[1] C. H. Bennett and G. Brassard, Proceedings of IEEE International Conference on Computers, Systems and Signal Processing **175**, 8 (1984).

[2] R. L. Rivest, A. Shamir, and L. Adleman, Commun. ACM **21(2)**, 120 (1978).

[3] J. S. Bell, Physics Physique Fizika **1(3)**, 195 (1964).

[4] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49(25)**, 1804 (1982).

[5] G. Weihs, et al., Phys. Rev. Lett. **81(23)**, 5039 (98).

[6] M. Matsumoto and T. Nishimura, ACM Transactions on Modeling and Computer Simulation **8(1)**, 3 (1998).

[7] R. Shaltiel, *International Colloquium on Automata, Languages, and Programming* (Springer, 2011).

[8] H. Schmidt, J. Appl. Phys **41(2)**, 462 (1970).

[9] C. H. Vincent, J. Phys. E **3(8)**, 594 (1970).

[10] D. Zuckerman, Algorithmica **16(4-5)**, 367 (1996).

[11] J. von Neumann, Appl. Math Ser. **12.36-38**, 5 (1951).

[12] T. Jennewein, et al., Rev. Sci. Instrum. **71(4)**, 1675 (2000).

[13] A. Stefanov, et al., J. Mod. Opt. **46(4)**, 595 (2000).

[14] M. Troyer and R. Renner, ID Quantique Technical Report (2012).

[15] P. X. Wang, G. L. Long, and Y. S. Li, J. Appl. Phys. **100**, 056107 (2006).

[16] M. Stpicevic and B. M. Rogina, Appl. Phys. Lett. **93(3)**, 031109 (2007).

[17] J. F. Dynes, et al., Appl. Phys. Lett. **93(3)**, 031109 (2008).

[18] M. A. Wayne, et al., J. Mod. Opt. **56(4)**, 516 (2009).

[19] M. A. Wayne and P. G. Kwiat, Opt. Express **18(9)**, 9351 (2010).

[20] M. Wahl, et al., Appl. Phys. Lett. **98**, 171105 (2011).

[21] H. Furst, et al., Opt. Express **18(12)**, 13029 (2010).

[22] M. Ren, et al., Phys. Rev. A **83(2)**, 023820 (2011).

[23] Y. Jian, et al., Rev. Sci. Instrum. **82**, 073109 (2011).

[24] M. Appletgate, et al., Appl. Phys. Lett. **107**, 071106 (2015).

[25] C. Gabriel, et al., Nat. Phot. **4(10)**, 711 (2010).

[26] Y. Shen, L. Tian, and H. Zou, Phys. Rev. A **81(6)**, 063814 (2010).

[27] T. Symul, S. M. Assad, and P. K. Lam, Appl. Phys. Lett. **98(23)**, 231103 (2011).

[28] Y. Shi, B. Chng, and C. Kurstsiefer, Appl. Phys. Lett. **109(4)**, 041101 (2016).

[29] Z. Zheng, et al., Rev. Sci. Instrum. **90(4)**, 043105 (2019).

[30] C. R. S. Williams, et al., Opt. Express **18(23)**, 23584 (2010).

[31] A. Argyris, et al., J. Light. Technol. **30(9)**, 1329 (2012).

[32] e. a. A. Rukhin, NIST Special Publication 800-22 Revision 1a (2010).

[33] M. Avesani, et al., Nat. Comm. **9(1)**, 1 (2018).

[34] S. Pironio, et al., Nature **464(7291)**, 1021 (2010).

[35] B. G. Christensen, et al., Phys. Rev. Lett. **111(13)**, 130406 (2013).

[36] Y. Liu, et al., Phys. Rev. Lett. **120(1)**, 010503 (2018).

[37] P. Bierhorst, et al., Nature **556(7700)**, 223 (2018).

[38] Y. Liu, et al., Nature **562(7728)**, 548 (2018).

[39] F. Arute, et al., Nature **574(7779)**, 505 (2019).

[40] H. Zhong, et al., Science (2020).

[41] J. F. Clauser, et al., Phys. Rev. Lett. **23(15)**, 880 (1969).

[42] U. V. Vazirani and T. Vidick, arXiv preprint arXiv:1111.6054 (2011).

[43] S. Aaronson and L. Chen, arXiv preprint arXiv:1612.05903 (2016).

[44] S. Aaronson and S. Gunn, arXiv preprint arXiv:1910.12085 (2019).

[45] A. Bouland, et al., Nat. Phys. **15(2)** (2019).