University *of Ljubljana*
Faculty *of Mathematics and Physics*

Seminar 1 - 1. letnik, 2. stopnja

Quantum random number generators

Author: Jure Pirman
Mentor: dr. Erik Zupanič
Ljubljana, march 2022

Technological advances have led to the need for methods and systems that they may seem unnecessary at first glance. Such is the requirement for a robust random number generator (RNG). Its use cases span a wide range of applications. For example, they play a key role in banking, telecommunications, datacenters and cryptography where security is crucial. Use cases have also been found in various automotive technologies, different simulations, etc. Most frequently pseudo-random number generators (PRNG) are used which rely on different mathematical operations to generate a sequence of numbers that are typically sufficiently random. Due to its formulaic nature, there exists a way to predict the results no matter the particular algorithm. This is why true-random number generators (TRNG) are important. These typically use specific hardware as a source of randomness. High-end random number generators commonly use quantum phenomena as a fundamental source of randomness, thus they are called Quantum Random Number Generators (QRNG). These are extremely important in quantum communications such as quantum key distribution.

Contents

1 Fundamental ideas

True random number generators (TRNG) typically consist of two parts, a source of entropy and various post-processing methods. Typical entropy sources are e.g. voltage noise, metastable ring oscillators, internal workings of computer hardware, etc. These commonly use processes that are hard to predict as a source of randomness but can be susceptible to external tampering. For example, a change in the temperature of a TRNG can cause a bias in the outgoing stream of numbers.

A more robust source of randomness are different quantum phenomena. Ideally, such a device outputs only readings of the observed quantum system. In reality, this is almost impossible, as we need to use different classical devices to set up and read from our system, all of which contribute to thermal noise in the result.

Almost every TRNG has a way of post-processing the output of the entropy source. Output is typically a number stream that adheres to certain rules, defined by algorithms that use said random numbers. Such a system can e.g. remove any biases, compensate for unwanted number distributions or even out the output bitrate. In the case of a QRNG, it should primarily evaluate and remove the thermal contribution from the quantum entropy source. Methods used in post-processing are selected based on the entropy source and the requirements of the number generator. Hence we won't be looking into post-processing methods or QRNGs that rely heavily on them.

There are typically two main characteristics used to describe any sort of RNG. The first is bitrate which simply tells us how many bits of data a device generates every second. The second property is randomness, which is harder to quantify. For example, we can take a look at a sequence of random numbers, that is finite in length. It's impossible to guarantee that this number is entirely unpredictable, but it may pass any randomness test. Thus there are many different tests to quantify different aspects of quantum number generators like their distributions and unpredictability. Furthermore, most of the tests can be passed by many of TRNGs or even PRNGs.

2 Random number analysis

Random number analysis is an important aspect when creating and designing a generator. There are several tests that attempt to analyse different properties of a generated string of numbers [1, 2]. Using these tests we can both analyse randomness and check whether random numbers conform to rules posed by latter algorithms. It is very important to note that testing random numbers can prove that data isn't random, but it cannot show or prove that data is random. This also implies that it is crucial to understand the methods and principles used in an RNG. For example, a finite length of random numbers from a PRNG can appear as random, but a larger string of numbers from the same generator may not be random.

2.1 Statistical prerequisite testing

These are statistical tests that assess certain properties, such as bias, serial correlation coefficients and more. Using these we're testing whether a series of random numbers conform to their requirement rather than their randomness.

2.1.1 χ^2 Test

To check whether a string of numbers x_i is biased, we have to perform a χ^2 test. The reasoning behind this is that the mean value $\bar{x}$ can also contain some amount of random walk. To calculate the value of χ^2 for a string of random numbers x_i between 0 and $k-1$, we have to count how many times a number has appeared in the sequence. We'll denote this as S_i for number i. χ^2 can now be calculated as:

$$\chi^2 = \sum_{i=0}^{k-1} \left(S_i - \frac{k}{n} \right)^2 . \tag{1}$$

Calculated χ^2 follows the $\chi^2(n-1)$ statistic, which allows us to determine the certainty, by which the string of numbers is unbiased.

2.1.2 Serial correlation coefficient

Another important metric for QRNGs and any RNG in general is the serial correlation coefficient (SCC). As we'll see, consecutive measurements can exhibit a correlation, which ruins the randomness of the generated string. SCC is a special case of correlation as it measures just the Lag-1 autocorrelation, and can be calculated as:

$$\mathrm{SCC} = \frac{n\left(\sum_{i=0}^{n-1} x_i x_{i+1}\right) - \left(\sum_{i=0}^{n-1} x_i\right)^2}{n\left(\sum_{i=0}^{n-1} x_i^2\right) - \left(\sum_{i=0}^{n-1} x_i\right)^2}. \tag{2}$$

The result is bounded as $-1 < \mathrm{SCC} < 1$. A negative value of SCC means that a bit is more likely to be the opposite value of the previous bit. When the value of SCC is positive, it means that a bit is more likely to be the same value as the previous bit. Therefore a sequence of good random numbers should have a value of SCC as close as possible to 0.

2.2 Randomness tests

Previously we have established that it is impossible to certify that a string of numbers is truly random. Because of this there are many tests used to evaluate the randomness or unpredictability of a sequence of numbers.

2.2.1 Shannon Entropy estimation

Another way of looking at random numbers is through the amount of information that is attached to them. Shannon entropy [3] estimates the average entropy of an output since it takes into account every possible outcome. We can define it for a source X which produces n different outcomes $x_1, x_2, ..., x_n$, each with probability $P(x_i)$, the Shannon entropy $\mathrm{H}(X)$ is

$$\mathrm{H}(X) = -\sum_{i=1}^{n} P(x_i)\log(P(x_i)). \tag{3}$$

If we calculate $\log_2$ instead of $\log$, the output will be expressed in terms of the number of bits of information. In simpler terms, we can use this to evaluate a number of random bits extracted for a single output from an entropy source. This can be a useful tool for the optimisation of a system, where parameters of the output distribution can be controlled. For example, by evaluating a one-bit system, we can write the probability of 0 as P_0 and the probability of 1 as $1 - P_0$. We can now estimate the number of output bits if a bias is present as:

$$\mathrm{H}(X) = -P_0\log_2(P_0) - (1-P_0)\log_2(1-P_0). \tag{4}$$

Maximum $\mathrm{H}(x)$ is as expected, at $P_0 = 0.5$, which corresponds to unbiased output. We can also see that a presence of bias can quickly lower the amount of generated bits.

2.2.2 Model Equivalence Testing

Typically a TRNG device is constructed on a system whose model is known. This permits a simulation of a TRNG, which consequently allows us to perform a theoretical analysis. A real-world device can then be compared and tested for equivalence. If the behaviour is equal or at least similar enough it is safe to assume that the data generated is random. This is especially true in the case of QRNGs since randomness from a QRNG can not be distinguished from a typical TRNG or even a PRNG [4].

3 Quantum entropy source

Every QRNG relies on quantum phenomena like a superposition of states or uncertainties of variables. Ideally, a QRNGs output string should be independent of any auxiliary components or sub-systems that are used inside a QRNG and it should only depend on the properties of the governing quantum system.

We'll take a look at two different approaches based on relatively simple principles which are usually used as a source of quantum randomness in generators.

3.1 Qubit approach

Qubit is essentially a two-level quantum-mechanical system, defined as a superposition of two states. It can be written as a combination of $|0\rangle$ and $|1\rangle$ states along with complex probability amplitudes α and β:

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{5}$$

where $|\alpha|^2 + |\beta|^2 = 1$. Only when measured, our system will assume one of two possible states. Furthermore, the probability for the system to assume state $|0\rangle$ is $|\alpha|^2$. Similar can be said for $|1\rangle$. E.g., the electron spin in the famous Stern-Gerlach experiment can be regarded as a qubit.

3.1.1 Photonic implementation

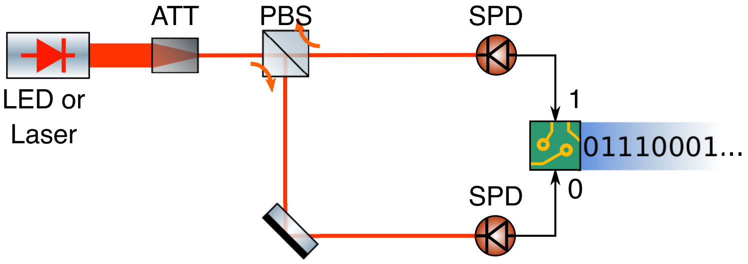

Figure 1: Schematic of a QRNG using a photonic qubit approach. The photon is initially in a superposition of horizontal (H) and vertical (V) polarization. Polarising beam splitter (PBS) separates photons based on their polarisation. Photons are detected using two SPDs, which is used for random bit generation. Laser or an LED can be used as a source of photons with appropriate attenuation (ATT). Adapted from [5].

The polarization of a photon can be interpreted as a quantum state. This means, that we can use simple optics to separate photons based on their polarization. Both polarizations can then be assigned a logical value of the output bit. As can be seen in Figure 1 we can use a Polarizing Beam Splitter (PBS) to separate photons based on their polarization. Two Single-Photon Detectors (SPD) are used to detect photons along both possible paths. Ideally, the number of detected events per time interval should be equal on both SPDs. This is achieved by using polarizers before the PBS.

Until this point, we haven't yet looked at photon sources. We used single photons to describe the inner workings of our systems. Using single-photon sources is unnecessary as these devices are expensive state-of-art equipment. A simple low power LED or laser suffices as a photon source. The output light must be attenuated significantly. For example, a green 1mW laser will output approximately $3 \cdot 10^{16}$ photons every second, but only $1 - 10 \cdot 10^{6}/s$ are required. Usually driving a LED or laser in pulsed mode and using strong attenuators suffices to create a series of single photons. which are usually used as a source of quantum randomness in generators. A key aspect when choosing a light source is its coherence time. It is crucial, that the coherence time of light is significantly shorter than the mean time between counts to avoid any correlation in properties from consecutive photons.

Lastly, we need to look at SPDs. Multiple devices can be used to detect single photons such as single-photon avalanche diodes (SPAD), photomultiplier tubes (PMT) or even superconducting nanowire single-photon detectors (SNSPD or SSPD). There are multiple factors when choosing between SPDs, which are most commonly cost, size and capabilities. Thus PMTs and SPADs are commonly used when designing and making a QRNG. These are typically small and relatively cheap. Another important characteristic is the maximum count rate, which is also the main limiting factor of the maximum output bitrate of these types of QRNGs. SPDs also contribute some amount of noise, typically referred to as dark count. These are simply events where an SPD outputs a signal without detecting a photon. They can be ignored to some degree as they are of quantum origin and can be in some cases also used as a source of quantum randomness. Another important characteristic of an SPD is their quantum efficiency, which is the measure of the effectiveness of a device to convert incident photons into counts. This parameter differs for every device and is also subject to ageing. Too large differences between sensors can cause skews or biases in the outputs. Attention must also be paid to phenomena like afterpulsing where an SPD can report a count shortly after a photon detection. Such outputs are heavily correlated and need to be taken into account since they strongly impact the performance of a QRNG. Thus many photodetectors employ dead times, which disable the detector for a certain amount of time after a count.

As previously mentioned, the output bitrate is limited by the capability of used detectors. Present fastest devices can detect a photon approximately every few 10-100ns [6], which corresponds to about 10Mbps. This has been achieved by multiple teams [7, 8]. Compared to other QRNGs, devices using the qubit approach are comparatively slow. For example, QRNGs based on temporal mode, which we will look at next, achieved rates on the order of 100Mbps. The fastest achieved rates were on the order of a few 10Gbps while using phase noise of light as a source of randomness [5]. Due to constraints, that arise from SPDs, an approach, that does not rely on detection rate alone becomes crucial.

Another issue arises from using different optical paths for each photon polarization. These can differ in length, alignment and properties of individual optical components, etc. This will effectively cause the paths to have slightly different transmittances. We also have to be mindful of the differences between the two SPDs. All of this results in a slight bias in the output. It is possible to compensate for this bias by adjusting the polarizer that sits before the PBS or changing the working parameters of SPD such as bias voltage in a SPAD. This is a time-consuming process, as the whole system has to be regularly monitored and adjusted.

3.2 Temporal approach

Using quantum states as a source of randomness is a powerful concept but can be rather difficult to execute in the real world. Realistically it is impossible to remove any kind of bias from such a device since the adjustments have to be extremely accurate. Furthermore, we have to account for any biases that arise from the ageing of SPDs and other components. By simplifying the physical setup as much as possible and introducing extraction methods that compare successive measurements, we can vastly improve the system. It also implies that a new source of randomness is required. So far we haven't looked at the uncertainty of position or time.

By doing this we'll move attention from the quantum properties of a single photon to the statistical properties of a stream of photons. One approach is to use the directional randomness in the emitted photon. Such devices have been made using arrays of SPDs [9]. These systems require a lot of post-processing to compensate for differences between individual SPDs in the array and non-uniform light intensity across the array. Almost all of these are specific to the setup that is used.

Another approach is to use randomness in creation and detection times. This approach requires just a single SPD and a light source. By doing this, we have also guaranteed that every photon has the same chance of being detected. A drawback is a requirement for more complex methods used to extract randomness.

3.2.1 Implementation

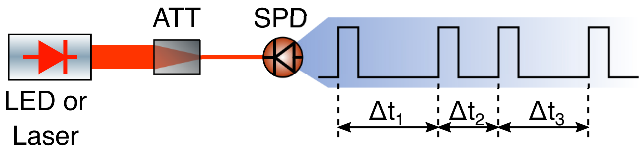

Figure 2: Schematic of a QRNG using a temporal approach. Random bits are generated from photon arrival times. Laser or an LED with significant attenuation (ATT), can be used as a source of photons, which are detected by the single-photon detector (SPD). Adapted from [5].

The hardware of QRNG based on the temporal approach can be relatively simple. Most of the work is done by electronics, which extract random numbers from photon detection times. This is why such a device can be constructed only from a light source and an SPD. The light source needs to be weak enough so that the SPD doesn't get saturated. Thus the maximum detection rate is usually similar to the qubit approach.

The origin of the randomness of time intervals between photon counts is the quantum nature of the photon source and the SPD. The choice of a source depends mostly on the required properties of the emitted light. Additionally, it is beneficial for a source to allow simple output power regulation. Thus a laser or an LED is the preferred source. Choice of a light source also affects the distribution of time intervals between counts, which is crucial both in terms of output and operation. Coherent laser light will have a Poissonian distribution of time intervals, while an LED will have a sub-Poissonian distribution. This implies that photons emitted from an LED will be more probable of being bunched together [10]. This is one of the reasons, why some prefer the usage of laser instead of an LED [11].

Random bit extraction can be done using multiple methods, each with its benefits and flaws. Since most of it is done through digital devices, we can establish a couple of theoretical capabilities of the temporal approach. Let's assume that the mean time between counts is $\bar{t}$. This value is subject to the output power of the source and is limited solely by the capability of SPD. As we have established before, we can expect numbers around 100ns. With current digital devices, it is possible to achieve time resolution δt in the order of 100ps or even less [5]. It is important to keep in mind that time resolution is also subject to the precision of the whole circuit, not just the digitizer. The number of output bits per detection can be estimated using (3). Using Shannon entropy for equal probabilities we get an estimate of $\log_2(\bar{t}/\delta t)$. We have to keep in mind that this number is an upper estimate since the delays between counts aren't equally probable. With previously mentioned numbers, we can estimate the maximum achievable output bitrate to be $\log_2(1000) \cdot 10\mathrm{Mbps} \approx 100\mathrm{Mbps}$.

3.2.2 Clocked method

The easiest approach to randomness extraction is to measure the time intervals between the detections. Issues arise as the output will have Poissonian-like distribution [10]. This requires further postprocessing. Another approach is to compare two consecutive delays Δt_i and Δt_{i+1}. A binary 1 is generated if $\Delta t_i > \Delta t_{i+1}$ and a 0 when $\Delta t_i < \Delta t_{i+1}$. In an ideal case, this produces a perfectly balanced output, at a bit rate equal to or lower than the detection rate. In reality, we have to account for a few different factors. Firstly the time measurement is digital and thus discrete. Therefore exists a non-zero chance that $\Delta t_i = \Delta t_{i+1}$. When this occurs the output bit should be omitted. Secondly, if the clock signal is constant, the counted delays will be correlated between consecutive counts. This is a result of the fact that we are summing up the remainder of the previous clock cycle and the current one. This can be more clearly seen in Figure 3. The issue can be solved by increasing the clock frequency or by using a restartable clock [12]. It is important to mention that this issue is also applicable to systems where we're not comparing measured delays.

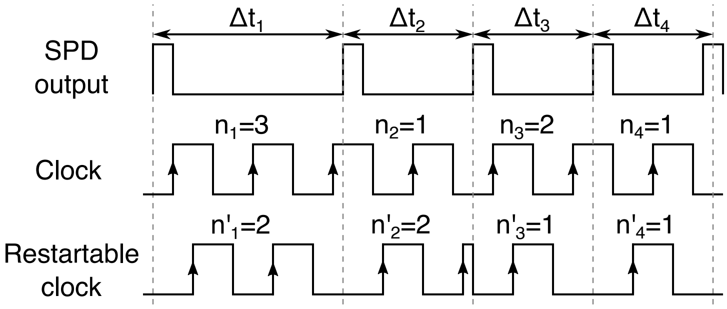

Figure 3: Comparison of time measurement using a regular clock signal and a restartable clock. We can see that with normal clock outputs are correlated since we're summing up the remainder of a cycle from a previous count with the present count. The issue is solved by restarting the clock when a photon is detected, or by increasing the clock frequency. Adapted from [12].

3.2.3 Pulsed method

Another approach is by pulsing the light source for a fixed amount of time so that the probability of detecting a single photon is small. Therefore the source of randomness is the probability that during a pulse a photon is created and detected. The hardware setup is the same as the one described at the beginning of this section with the key difference being that the light source is pulsed. We can create an unbiased output by comparing consecutive repetitions. This is achieved by creating a group of pulses, which contains two subgroups of identical lengths. We assign a bit to each subgroup, which tells us whether a photon was detected or not as can be seen in Figure 4. There are four possible outcomes: 00, 01, 10 and 11. In the case of 00 and 11 nothing is written to the output. A logical 0 is output for a 01 and a 1 for a 10, or vice-versa [13]. Since the probability of photon detection is equivalent between groups, we can safely assume that both 01 and 10 outcomes are equally probable.

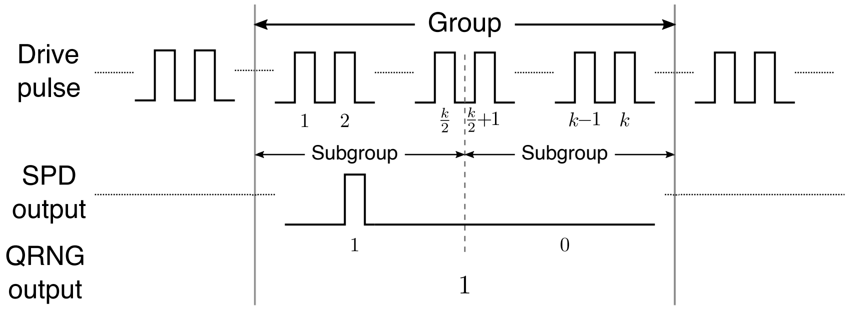

Figure 4: Diagram of a pulsed QRNG. The light source is driven by short drive pulses. A group comprises of k pulses. Every group has two equal length subgroups, each with their corresponding bit. If a photon is detected in a subgroup a 1 is assigned to its bit. When a 10 is detected, a 1 is output and when a 01 is detected a 0 is output. In the case of a 00 or a 11 nothing is output. Adapted from [13].

4 Conclusion

Quantum random number generators are and are becoming even more important. Together with vastly increasing demand in cybersecurity alongside computation capabilities might make them play a key role in everyday life. As we have seen, these devices use relatively simple physics, from which we extract randomness. Even though the theory behind the operation is simple, practical execution can prove to be very difficult. This is amplified by the fact that QRNGs are some of the highest-end random number generators in terms of the randomness of their output, thus the requirements can be very strict. Using various methods presence of biases, output number distributions and classical contributions can be removed. Upon simple concepts, scientists have built increasingly capable QRNGs.

References

[1] L. Bassham, A. Rukhin, J. Soto, J. Nechvatal, M. Smid, S. Leigh, M. Levenson, M. Vangel, N. Heckert, and D. Banks, “A statistical test suite for random and pseudorandom number generators for cryptographic applications,” 2010-09-16 2010.

[2] D. Johnston, *Random number generators - principles and practice : a guide for engineers and programmers*. De G Press, 2018.

[3] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[4] A. Acín and L. Masanes, “Certified randomness in quantum physics,” *Nature*, vol. 540, pp. 213–219, 12 2016.

[5] X. Ma, X. Yuan, Z. Cao, B. Qi, and Z. Zhang, “Quantum random number generation,” *npj Quantum Information*, vol. 2, 06 2016.

[6] M. D. Eisaman, J. Fan, A. Migdall, and S. V. Polyakov, “Invited review article: Single-photon sources and detectors,” *Review of Scientific Instruments*, vol. 82, p. 071101, 07 2011.

[7] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, “A fast and compact quantum random number generator,” *Review of Scientific Instruments*, vol. 71, pp. 1675–1680, 04 2000.

[8] J. Rarity, P. Owens, and P. Tapster, “Quantum random-number generation and key sharing,” *Journal of Modern Optics*, vol. 41, pp. 2435–2444, 12 1994.

[9] Q. Yan, B. Zhao, Q. Liao, and N. Zhou, “Multi-bit quantum random number generation by measuring positions of arrival photons,” *Review of Scientific Instruments*, vol. 85, p. 103116, 10 2014.

[10] M. Fox, *Photon statistics*. Quantum optics: An Introduction, Oxford University Press, 2006.

[11] M. A. Wayne, E. R. Jeffrey, G. M. Akselrod, and P. G. Kwiat, “Photon arrival time quantum random number generation,” *Journal of Modern Optics*, vol. 56, pp. 516–522, 02 2009.

[12] M. Stipcevic and B. M. Rogina, “Quantum random number generator based on photonic emission in semiconductors,” *Review of Scientific Instruments*, vol. 78, p. 045104, 2007.

[13] L. M. Yu, M. J. Yang, P. X. Wang, and S. Kawata, “Note: A sampling method for quantum random bit generation,” *Review of Scientific Instruments*, vol. 81, p. 046107, 04 2010.