MŰEGYETEM 1782

Budapest University of

Technology and Economics

Department of Networked Systems and Services

Optical Solutions for Quantum Key Distribution Transmitters

Ph. D. Thesis

Author:
Ágoston Schranz

Supervisor:
Dr. Eszter Udvary Ph. D.
Associate Professor

Budapest, 2021

Acknowledgement

> For in much wisdom is much grief: and he
> that increaseth knowledge increaseth sorrow.
>
> *Ecclesiastes 1:18*

First and foremost, I would like to express my gratitude to my supervisor, Eszter Udvary. She provided me immense help and opened up research options and possibilities for me through her involvement in new cooperations with other groups. Her continued support within the university also helped me not become overwhelmed by tasks unrelated to my research. She also gave me a free hand in my investigations, allowing me to find what I kind of research I am most suited for. I am also thankful to many of my colleagues at BME for providing a friendly and helpful atmosphere, especially Balázs Matolcsy, who helped me both with discussions and practical issues. I would also like to thank Ericsson Hungary Ltd. for their continued scholarship during my years as a Ph. D. student, and especially Dr. Gábor Kovács for organizing student conferences in each semester, where I could regularly present my work. Finally, I would like to thank my family and friends, many of whom showed interest in my research, even though they are working in completely different fields, and supporting my decision of pursuing a Ph. D.

Abstract

The field of quantum technologies has been going through significant development during the past few decades. Quantum computation and quantum communications, specifically, are capable of providing solutions for problems that are either impossible or computationally expensive with purely classical methods. Significant theoretical achievements have been already presented in the form of quantum algorithms, which can solve given problems in a fraction of the time made possible by current technology. However, these new opportunities will also be available to those who wish to exploit them for despicable purposes, such as breaking widely used public-key cryptosystems, rendering them unsecure.

The good news is that the practical realization of these algorithms is still in its infancy, giving us time to be prepared for a post-quantum era. Quantum key distribution (QKD) protocols have been invented to facilitate eavesdropper detection in provably secure private-key cryptography systems, utilizing the fact that any measurement conducted on a quantum bit alters its state. The first chapter of this thesis introduces some important concepts of quantum cryptography: symmetric key systems, discrete variable quantum key distribution, and quantum random number generation (QRNG). The physical realization of QKD protocols is always an imperfect approximation of the ideal devices and the communication medium; namely, transmitters, receivers and the channel. We need to turn our attention to the problems introduced by such imperfections, which potentially offer loopholes for eavesdropping. Additionally, alternative implementations of protocols can be of particular interest, if they provide improvements over previous designs. Chapter 2 proposes a new transmitter setup for the forefather of all QKD, BB84. It exploits the fact that vertical cavity surface-emitting lasers (VCSELs) can be modulated in polarization between two orthogonal states, effectively reducing the number of necessary light sources in the transmitter. Different aspects of realization are given in detail, also including how to avert an attack specifically targeting this design.

The generation of truly random, indeterministic number sequences is also necessary for QKD systems. The fundamentally probabilistic nature of quantum measurements offer a wide range of options for random number generation. Amongst these, different optical phenomena are popular choices for a variety of reasons. These include the relative simplicity of light generation, the potential use of highly developed equipment originally designed for optical communication networks, etc. The second half of this thesis describes two optical quantum random number generators, which are based on the difference of time intervals between successive photon detections. Chapter 3 introduces the mathematical model of an already existing, robust scheme, which produces bits from an inherently uniform distribution. Two important figures of merit are the focal point of the discussion, the bit generation efficiency and the bit generation rate. Chapter 4 details an improvement of the previous method in terms of its efficiency, complemented with an error analysis assuming non-ideal circumstances. The claims of improvement are verified experimentally, also proving that the quality of randomness can be upheld under the changes made to the original method.

Contents

List of Publications

This is a full list of my publications as of writing this manuscript; not all of them are related to my theses. Thesis related publications are directly mentioned in Chapter 5, and referenced when necessary in previous chapters.

Journal Papers

J1 Ágoston Schranz and Eszter Udvary. "Mathematical analysis of a quantum random number generator based on the time difference between photon detections". In: *Optical Engineering* 59.4 (2020), p. 044104. DOI: 10.1117/1.OE.59.4.044104

J2 Ádám Marosits, Ágoston Schranz, and Eszter Udvary. "Amplified spontaneous emission based quantum random number generator". In: *Infocommunications Journal* 12.2 (2020), pp. 12–17. DOI: 10.36244/icj.2020.2.2

J3 Ágoston Schranz and Eszter Udvary. "Error probability in polarization sensitive communication systems in terms of moments of the channel's rotation angle". In: *Optical and Quantum Electronics* 53.1 (Jan. 2021), p. 62. ISSN: 0306-8919. DOI: 10.1007/s11082-020-02690-1

J4 Balázs Matolcsy, Eszter Udvary, and Ágoston Schranz. "Common-mode noise filtering with space-divided differential 2x2 VLC for V2V applications". In: *Optical and Quantum Electronics* 53.4 (2021), p. 182. DOI: 10.1007/s11082-021-02808-z

J5 Ágoston Schranz, Eszter Udvary, and Balázs Matolcsy. "Efficiency Improvement of a Time-of-Arrival Quantum Random Number Generator". In: *Optical Engineering* 60.3 (2021), p. 034112. DOI: 10.1117/1.OE.60.3.034112

Conference Papers

C1 Ágoston Schranz, Eszter Udvary, and Zsolt Kis. "Photon statistics determination for single photon based quantum key distribution". In: *18th International Conference on Transparent Optical Networks (ICTON)*. IEEE. 2016, pp. 1–4. DOI: 10.1109/ICTON.2016.7550483

C2 Eszter Udvary, Ágoston Schranz, and Balázs Matolcsy. "Dispersion and off-set filtering in RSOA based networks". In: *18th International Conference on Transparent Optical Networks (ICTON)*. 2016, pp. 1–4. DOI: 10.1109/ICTON.2016.7550690

C3 Ágoston Schranz. "Experimental Investigation of VCSEL for Quantum Communications". In: *Mesterpróba 2016*. 2016, pp. 8–11

C4 Ágoston Schranz. "Investigation of VCSEL Polarization for Quantum Key Distribution". In: *International Interdisciplinary PhD Workshop 2016*. 2016, pp. 117–120

C5 Gábor Szabó, Ágoston Schranz, and Eszter Udvary. "Nonlinear Modulation Characteristics of LEDs in Radio on Visible Light Systems". In: *International Interdisciplinary PhD Workshop 2016*. 2016, pp. 6–10

C6 Ágoston Schranz and Eszter Udvary. "Transmitter Design Proposal for the BB84 Quantum Key Distribution Protocol using Polarization Modulated Vertical Cavity Surface-emitting Lasers". In: *Proceedings of the 6th International Conference on Photonics, Optics and Laser Technology*. INSTICC. SciTePress, 2018, pp. 252–258. ISBN: 978-9-897-58286-8. DOI: 10.5220/0006638002520258

C7 Ágoston Schranz and Eszter Udvary. "Quantum Bit Error Rate Analysis of the Polarization based BB84 Protocol in the Presence of Channel Errors". In: *Proceedings of the 7th International Conference on Photonics, Optics and Laser Technology - Volume 1: PHOTOPTICS*. INSTICC. SciTePress, 2019, pp. 181–189. ISBN: 978-9-897-58364-3. DOI: 10.5220/0007384101810189

C8 Ágoston Schranz, Ádám Marosits, and Eszter Udvary. "Effects of Sampling Rate on Amplified Spontaneous Emission Based Single-Bit Quantum Random Number Generation". In: *21st International Conference on Transparent Optical Networks (ICTON)*. IEEE. 2019, pp. 1–4. DOI: 10.1109/ICTON.2019.8840188

Book Chapters

B1 Ágoston Schranz and Eszter Udvary. “Polarization Modulated Vertical-Cavity Surface-Emitting Lasers in Quantum Key Distribution”. In: *Optics, Photonics and Laser Technology 2018*. Vol. 223. Springer Series in Optical Sciences. Springer, Cham., 2019, pp. 75–92. DOI: 10.1007/978-3-030-30113-2_4

Chapter 1

Basics of Quantum Cryptography

1.1 Introduction

To provide motivation for my research discussed in the following chapters, I shall begin with introducing the theoretical background of three important aspects of quantum cryptography. The first of these is the field of symmetric key cryptography, with special attention towards the one-time pad scheme. The need of introducing the two other topics follows naturally: both of them are invaluable for the practical realization of provably secure cryptosystems.

The second, quantum key distribution (QKD), is the study of protocols utilizing quantum phenomena to distribute encryption keys between parties, which allow for the detection of any eavesdropping. An unauthorized person obtaining another perfect copy of the exact encryption method is the greatest menace of all for those who want to communicate in secrecy. After a brief general overview, discrete variable QKD is introduced through a selection of well-known protocols, such as BB84, B92 or E91.

The last of the three is the topic of random number generation—a field with many different applications. However, cryptography is arguably the most important among those possibilities; good quality generators of uniformly distributed and independent bits are an absolute necessity for key generation in symmetric key cryptography, but also for the proper operation of QKD transmitters and receivers. State-of-the-art solutions of optical quantum random number generators (QRNGs) are then presented, which, as opposed to algorithmic generators, are suitable for the most demanding situations as well.

1.2 Symmetric Key Cryptography

In cryptography, one party (commonly referred to as "Alice") wants to send a message $M \in \mathcal{M}$ to another ("Bob"). She encrypts the original message with a secret key $K \in \mathcal{K}$ to obtain the ciphertext $E \in \mathcal{E}$. $\mathcal{M}$, $\mathcal{E}$ and $\mathcal{K}$ are the message, ciphertext and key spaces, respectively. Bob then uses another key to decipher E and obtain—hopefully—the original message. The ultimate goal is to deny eavesdroppers or cryptanalysts ("Eve") from accessing the message.

There are two main approaches regarding the nature of keys. The first is the group of symmetric key methods, where Alice and Bob have a pre-shared secret: two identical keys—or two, which are easy to transform into each other—for encryption and decryption. This relies on a prior key distribution phase, which is problematic in real-life scenarios, and the adversaries could be hoping for capturing the key in secret.

The second group overcomes this concern: in public-key or asymmetric cryptography each party has their own secret key, but also its pair, a public key, which may be known by anyone who is interested. The secret and public keys are connected through a one-way function: a function that is difficult to invert. In such schemes, Alice encrypts the message with the public key of the intended recipient, Bob, who then uses his private key for decryption. Although no key distribution is necessary, some of these algorithms rely on computational complexity, rather than provable security. Nowadays, public-key cryptography and hybrid solutions wherein symmetric keys are distributed while encrypted with a public-private key pair, are widely adopted. One such example is the RSA algorithm [15], which is based on the complexity of integer factorization of large prime numbers. It has been shown that Shor's algorithm [16]—a quantum algorithm—is theoretically capable of solving this problem significantly faster than any classical known method. At this moment, however, the largest integer which has been factorized using the algorithm is 21 [17] back in 2012. Similar quantum computing solutions may exist for other computational problems, which could eventually render wide-spread schemes unsecure. Even if present-day technology is still a long way off from breaking public-key systems, it is wise to conduct research into both public-key algorithms resistant against quantum attacks (post-quantum cryptography) and the possibility of utilizing suitable symmetric key protocols shared with secure key distribution.

An information-theoretically secure option is the symmetric key protocol called the *one-time pad* (OTP) or *Vernam cipher*, originally invented in 1882 by Frank Miller [18] and later independently patented by Gilbert Vernam [19]. In OTP, every "letter" of the message is paired with a randomly chosen key letter from the same alphabet. The key is a shared secret between Alice and Bob; its length should exceed that of the message, and it should not be reused. The ciphertext is obtained by the modulo n addition of M and K, where n is the number of letters in the alphabet. On Bob's side, modular subtraction of the same key from the ciphertext yields the original message. If the alphabet is the set of binary digits $\{0, 1\}$ with cardinality two, the modular addition is the same as the bitwise XOR operation $\oplus$ defined as

$$u \oplus v = \begin{cases} 0, & \text{if } u = v, \\ 1, & \text{if } u \neq v, \end{cases} \tag{1.1}$$

where $u, v \in \{0, 1\}$. Also, since modulo 2 addition and subtraction are essentially the same, the bitwise XOR of E and K can be used for decrypting the message. Figure 1.1 shows an example of the process.

$$\begin{array}{cr@{\qquad}cr}
M & 0\,0\,1\,0\,0\,1\,0\,0\,1\,1\,1\,0 & E & 1\,1\,1\,1\,0\,1\,1\,0\,0\,0\,1\,1 \\
K & \oplus 1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\,1 & K & \oplus 1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\,1 \\
\cline{2-2} \cline{4-4}
E & 1\,1\,1\,1\,0\,1\,1\,0\,0\,0\,1\,1 & M & 0\,0\,1\,0\,0\,1\,0\,0\,1\,1\,1\,0
\end{array}$$

Figure 1.1: Encryption (left) and decryption (right) of message M and ciphertext E using key K in a binary one-time pad scheme.

Claude E. Shannon showed in his work titled *Communication Theory of Secrecy Systems* [20] that the one-time pad is indeed a cryptosystem providing what one may call *perfect secrecy.* Perfect secrecy defines the notion that Eve should not be able to gain information about the message M based on the ciphertext E. Therefore, the *a priori* probabilities $P\left(M\right)$ of messages—a property of the underlying language—should be equal to their *a posteriori* probabilities conditioned on the ciphertext, $P_E\left(M\right) = P\big(M \mid E\big)$. Using Bayes' theorem,

$$P_E\left(M\right) = \frac{P\left(M\right) P_M\left(E\right)}{P\left(E\right)}, \tag{1.2}$$

where $P_M\left(E\right) = P\big(E \mid M\big)$, it becomes obvious that perfect secrecy is only achieved if $P\left(E\right) = P_M\left(E\right)$, meaning that the probability of getting ciphertext E should be independent of the message to be encrypted. This comes with several consequences and requirements regarding perfect cryptosystems. First, an *endomorphic* system is needed—the number of possible messages should be equal to the number of ciphertexts:

$$\left|\mathcal{M}\right| = \left|\mathcal{E}\right|, \tag{1.3}$$

and only one key should transform every M into every possible E. Second, the number of possible keys should be at least as many as that of the messages:

$$\left|\mathcal{K}\right| \geq \left|\mathcal{M}\right|. \tag{1.4}$$

$\left|\mathcal{K}\right| = \left|\mathcal{M}\right|$ also means that each key should be equally likely—the keys should be uniformly distributed! This is the reason why quantum random number generators, discussed later, are designed to generate bits as close to an uniform distribution as possible. An OTP scheme that utilizes such a generator fulfills the criteria of perfect secrecy.

The average uncertainty (information content) or *(Shannon) entropy* $H\left(X\right)$ of a discrete random variable X is defined as [21]

$$H\left(X\right) = -\sum_{x\in\mathcal{X}} P\left(X = x\right) \log P\left(X = x\right), \tag{1.5}$$

where $\mathcal{X}$ and P are the sample space and the probability measure of the underlying probability space $(\mathcal{X}, \mathcal{S}, P)$. Since the entropy is bounded from above by the max-entropy $\log|\mathcal{X}|$, it follows that the uncertainty of messages is at most $\log|\mathcal{M}|$. This amount of information needs to be concealed by a key with at least as much uncertainty, further proving the requirement of Eq. 1.4.

However, this produces a significant practical drawback: infinitely long messages require infinitely long keys for perfect secrecy. Altogether, a key generating device should be able to create the key at a rate matching the information generation rate of the source. Moreover, keys should not be used more than once. Thus, key distribution (sharing new keys between the two participants) becomes an important task, with special care taken to avoid giving unauthorized parties the possibility of obtaining a copy.

1.3 Discrete Variable Quantum Key Distribution

The task of secure key distribution between the authorized parties has been made possible thanks to the research of the past 30 years. Quantum key distribution protocols have been designed in order to tackle the seemingly insurmountable problem of eavesdropper detection [22]. This field is arguably the best-developed aspect of quantum communications and computing. Quantum computing utilizes quantum bits (qubits) instead of regular bits, which can not only take on two discrete values, but an arbitrary superposition of those two. QKD exploits the fundamental traits of quantum physics, which allow one to statistically monitor whether someone tried to gain information about the qubits—which inevitably changes the quantum system in question.

Let me start with a short summary of notation and concepts relevant to quantum information theory [23]. A quantum state can be described by a column vector in a relevant Hilbert space over the field of complex numbers $\mathbb{C}$. Such a vector is denoted by a *ket* $|\psi\rangle$, where ψ is a label, not a direct representation of the vector elements. Its conjugate transpose or Hermitian conjugate, a row vector, is a *bra* $\langle\psi| = |\psi\rangle^\dagger$. The inner product (dot product) of two states $|\psi_i\rangle$, $|\psi_j\rangle$ is then written as $\langle\psi_i|\psi_j\rangle$. States are normalized so that $\langle\psi|\psi\rangle = 1$. As a consequence, any two basis states from an orthonormal basis spanning the vector space obey the relationship $\langle\psi_i|\psi_j\rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta

$$\delta_{ij} = \begin{cases} 0, & i \neq j; \\ 1, & i = j. \end{cases} \tag{1.6}$$

As usual, a zero inner product implies orthogonality of the states. Measurement of quantum states is, to the best of our knowledge, a totally probabilistic event. During a projective measurement, measuring a state in a certain basis could yield either of the basis vectors as a result, with probabilities depending on the overlap between the measured and the basis states. The inner products are intrinsically linked to measurement probabilities; the latter can be expressed as the squared absolute value of some inner product. The result is only certainly "correct" if the state to be measured is a basis vector of the measurement basis, as the overlap (inner product) of different orthogonal basis vectors is zero.

Protocols where the underlying qubits dwell in a finite-dimensional (mostly 2D) Hilbert space are usually called discrete-variable (DV-QKD) protocols, while those in infinite-dimensional spaces are continuous-variable (CV-QKD) solutions [22]. In this thesis, only the topic of DV-QKD is covered.

Qubits used by DV-QKD protocols are quantum states which can be described in two-dimensional vector spaces, e.g. the polarization of a single photon. A general qubit state is then given by

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{1.7}$$

$$\langle\psi| = \alpha^*\,\langle 0| + \beta^*\,\langle 1| \tag{1.8}$$

in its ket and bra forms, with $\alpha, \beta \in \mathbb{C}$ being the *probability amplitudes* and * denoting complex conjugation. $|0\rangle\,, |1\rangle$ are two orthogonal basis vectors. Normalization constraints require $|\alpha|^2 + |\beta|^2 = 1$. For the given example of polarization states, $|0\rangle$ and $|1\rangle$ usually denote the *rectilinear* basis of horizontal and vertical linear polarizations, respectively. Another basis, the *diagonal*, is given by the vectors representing linear polarizations angled at ±45°,

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \text{ and} \tag{1.9}$$

$$|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \tag{1.10}$$

The rectilinear and diagonal bases are mutually unbiased, meaning that measuring a state lying along one of the basis states in the other basis yields a completely random outcome with equal probabilities. As an example, measuring $|1\rangle$ in the diagonal basis yields results $|+\rangle$ and $|-\rangle$ with probabilities $\left|\langle +|1\rangle\right|^2 = \left|\langle -|1\rangle\right|^2 = 0.5$ each. Measurements also change the quantum state by collapsing the wave function into the measured state.

1.3.1 The Basic Idea Behind QKD

The underlying notion of QKD is *eavesdropper detection.* Rather than trying to conceal and distribute keys so that no one is able to gain any information about them—which would be a noble but futile attempt—, one should design protocols that allow the detection of this information leakage, thus preventing the use of compromised keys.

Quantum key distribution is made possible by the *no cloning theorem* [24], which states that an arbitrary unknown quantum state cannot be copied. States known beforehand, and those being pairwise orthogonal to each other can be reproduced faithfully, since measurement in the corresponding eigenbasis always yields the input state as a result. Therefore, every QKD protocol uses at least a pair of states which are not orthogonal to each other; indeed, any two non-orthogonal states suffice for a simple method, the B92 [25].

As a consequence of the theorem, Eve is not able to copy incoming qubits and store them for further manipulation; any eavesdropping strategy works as a quantum measurement, thus affecting and potentially altering the states in question. Even if for the purposes of security analysis, Eve is supposed to be capable of everything not explicitly forbidden by physics, this alteration will result in a mismatch between the raw keys of Alice and Bob, quantified by the quantum bit error rate (QBER). The two parties can compare a sufficiently long random substring of their raw keys on a public channel. If the QBER exceeds a pre-defined limit—depending on the exact strategy of Eve—, they would conclude that someone is listening and has gained a significant amount of information, and the keys would be discarded, the process aborted, and potentially started anew. Obviously, since the eavesdropping method is unknown to anyone but Eve, one should expect the worst-case scenario of an optimal strategy, and use the corresponding lowest acceptable QBER limit. Even if it is impossible to eavesdrop on such a protocol unnoticed, key distribution between Alice and Bob can be permanently blocked.

1.3.2 Problems of Practical Realization

Some DV-QKD protocols are theoretically using qubits implemented on single photons. The quantum state of radiation describing one qubit is thus a pure Fock state $|n = 1\rangle$. Fock states $\{\ |\, n\rangle\ \mid\ n = 0, 1, 2, \ldots\}$ have a well-defined number (n) of photons, and are thus eigenstates of the *number operator* $\hat{N} = \hat{a}^{\dagger}\hat{a}$ with an eigenvalue n. Here, $\hat{a}^{\dagger}$ and $\hat{a}$ are the creation and annihilation operators of the quantum harmonic oscillator, respectively.

$$\hat{N}\left|n\right\rangle = n\left|n\right\rangle \tag{1.11}$$

Mathematically, Fock states are orthogonal to each other $\big(\langle m|n\rangle = \delta_{mn}\big)$, and provide a simple basis expansion for any state of radiation, but their realization is difficult. True single-photon sources—albeit they exist—have not yet found their way into commercial applications; therefore, practical implementations of QKD transmitters require accessible substitutes, which are usually weak, highly attenuated laser pulses [22].

Low-power laser light can be well approximated with a *coherent state* $\left|\alpha\right\rangle$, where the label α is an arbitrary complex number, conveniently written in its polar form $\alpha = |\alpha| \cdot \mathrm{e}^{\mathrm{i}\phi}$. A coherent state can be expressed in terms of Fock states as

$$\left|\alpha\right\rangle = \mathrm{e}^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} \left|n\right\rangle . \tag{1.12}$$

Coherent states are eigenstates of the annihilation operator $\hat{a}$ with eigenvalue α, leading to the identities

$$\hat{a}\left|\alpha\right\rangle = \alpha\left|\alpha\right\rangle \tag{1.13}$$

and, since $\big(\hat{a}\left|\alpha\right\rangle\big)^{\dagger} = \left\langle\alpha\right|\hat{a}^{\dagger}$,

$$\left\langle\alpha\right|\hat{a}^{\dagger} = \left\langle\alpha\right|\alpha^{*}. \tag{1.14}$$

The mean number of photons $\langle n\rangle$ in a coherent state is

$$\langle n\rangle = \left\langle \alpha \,\middle|\, \hat{N} \,\middle|\, \alpha \right\rangle = \left\langle \alpha \,\middle|\, \hat{a}^{\dagger}\hat{a} \,\middle|\, \alpha \right\rangle = \left\langle \alpha \,|\, \alpha^{*}\alpha \,|\, \alpha \right\rangle = |\alpha|^2 \left\langle \alpha \,|\, \alpha \right\rangle = |\alpha|^2 , \tag{1.15}$$

whereas the probability of measuring exactly n photons in the coherent state is given by

$$P\left(N = n\right) = \left|\left\langle n \,|\, \alpha \right\rangle\right|^2 = \mathrm{e}^{-|\alpha|^2} \frac{|\alpha|^2}{n!}. \tag{1.16}$$

Therefore, the average power of a coherent state is proportional to $|\alpha|^2$, but does not depend on the phase ϕ. The above probability mass function is exactly that of a Poisson distribution [26] with parameter $\lambda = |\alpha|^2$, which is equal to both its mean and variance. A peculiar property of any Poisson distribution is that its support is the set of non-negative integers, and no probability $p_n = P\left(N = n\right)$ is exactly zero, regardless of the parameter.

Thus, if one would like to use a weak coherent source as a substitute for a true single-photon source, there are two types of problems. The first is significantly more alarming: if the mean is large—only around $\lambda = 1$ is sufficient—, a significant portion of pulses will contain multiple photons (Table 1.1). This gives Eve the opportunity to perform a photon number splitting (PNS) attack, originally discussed in [27], or an extended PNS [28] attack. In a BB84 setting, this roughly corresponds to blocking single-photon pulses; storing one (or more) photons of multiphoton pulses, letting the others pass through a lossless channel between Eve and Bob, and measuring the polarization of stored photons only after the basis choices had been disclosed on the public channel, gaining significant information about the key. Assuming a lossless channel may seem strange, but there is nothing in theory prohibiting Eve to use such a transmission medium. In case the original channel losses, without Eve and her trickery, are higher than a certain limit, the PNS attack allows Eve to gain full information about the key [27]. On top of that, an extended PNS attack would also incorporate manipulation of the photon statistics to resemble a Poisson distribution, as expected, which is otherwise modified by the eavesdropping.

Table 1.1: Probabilities of measuring N photons for different Poisson distribution parameters.

λ	$P(N=0)$	$P(N=1)$	$P(N=2)$	$P(N>2)$
0.1	0.90484	0.09048	0.00452	0.00016
1.0	0.36788	0.36788	0.18394	0.08030

The second problem does not have security issues, but can be an annoyance. For smaller parameters, the probability of measuring a blank state with zero photons, p_0, increases. This greatly reduces the achievable key distribution rate. A general rule of thumb is that $\lambda \approx 0.1$ is a good trade-off between blank and multiphoton pulses (see Table 1.1 for exact probabilities); even so, the key rate is decreased to less than 10% of its original value, but only about 0.45% of pulses contain more than one photon.

1.3.3 Examples of DV-QKD Protocols

To provide a basis for further discussion, let us go through some of the most basic and/or most widely used discrete variable QKD protocols.

BB84. Developed by the pioneers of quantum cryptography, Charles Bennett and Gilles Brassard in 1984, what later became known as BB84 is the earliest example of a quantum key distribution protocol [29]. It often serves as a starting point when introducing QKD, since the main principle is easy to understand, and provides a nice illustration of how the no-cloning theorem is exploited. BB84 uses four different quantum states. Each state is orthogonal to exactly one other state from the set, forming a basis of a two-dimensional Hilbert space. The two bases are conjugate; therefore, not all four states are pairwise orthogonal. One of the possible implementations operates with the linear polarization of single photons; I will use this formulation in all further discussion. However, the protocol could be implemented with phase and frequency encoding as well [30].

The four possible linear polarization states are defined by their angles. Horizontal (0°, →, $|0\rangle$) and vertical (90°, ↑, $|1\rangle$) polarizations make up the rectilinear basis (+), while polarization angles of ±45°(↗/$|+\rangle$ and ↘/$|-\rangle$) constitute the diagonal basis (×). Alice generates two independent, uniformly distributed random bit streams $\{s_\mathrm{A}\}$ and $\{m_\mathrm{A}\}$. s_A is the key bit Alice wants to share with Bob. m_A, on the other hand, decides the basis in which the key bit is to be encoded. Each ordered pair $(s_\mathrm{A},\ m_\mathrm{A})$ corresponds to one of the four possible quantum states (see Table 1.2 for the exact assignment) that is then sent to the receiver.

Table 1.2: Key bits, basis bits, and the corresponding quantum state sent by Alice in the BB84 protocol.

Key bit s_A	**Basis bit** m_A	**Sent state**
0	0	→
1	0	↑
0	1	↗
1	1	↘

Bob, independently from Alice, also generates a random bit sequence $\{m_\mathrm{B}\}$, which chooses the basis used for measurement to obtain bits s_B. Assume now that the channel used for transmission is noiseless and there is no eavesdropping. If $m_\mathrm{A} = m_\mathrm{B}$, Bob measures the state in the basis it was encoded in, and he gets a correct result in 100% of such cases ($s_\mathrm{A} = s_\mathrm{B}$). However, measurement in the conjugate basis results in a completely random, uniformly distributed result. Afterwards, both parties disclose their basis choices on a classical channel, and keep only those bits, for which their choices agreed. This process is called key sifting, its output being the raw key of Alice and Bob, respectively.

Now, in a practical scenario, both channel imperfections and eavesdropping attempts can and will alter the quantum states before it reaches the receiver, causing Bob's measurements in the correct basis to fail sometimes. This is, however, beneficial in terms of secrecy: during the information reconciliation phase, the two parties can compare a randomly selected subset of their raw keys, and calculate an estimation for the quantum bit error rate. Whenever the QBER is above a certain limit, Alice and Bob can suspect that there is an eavesdropper trying to gain information, and abort the key distribution process. Due to this, even if the adversary is unable to obtain a copy of the secret keys, it is possible to permanently block encrypted communication between the parties. The upper limits of QBER for which any DV-QKD protocol with an underlying two-level system can operate safely, is $1/2 - \sqrt{1/8} \approx 0.1464$ for optimal attacks on individual qubits [31–33], and 0.11 against stronger coherent attacks [33].

B92. Arguably the most simple DV-QKD protocol is B92, described by Charles Bennett in 1992, eight years after the original BB84 article. As mentioned earlier, its simplicity lies in the fact that it only uses two, necessarily non-orthogonal states [25]. Although the original paper does not specify anything about the two states apart from their non-orthogonality—and proves that any pair of states fulfilling this condition suffices—, I am going to introduce it based on a wide-spread method, which utilizes linearly polarized single photon states $|0\rangle$ and $|+\rangle$ from mutually unbiased bases.

Alice prepares $|0\rangle$ whenever she wants to send a key bit $s_\mathrm{A} = 0$, and $|+\rangle$ for a bit $s_\mathrm{A} = 1$. Bob measures the received state randomly in either the rectilinear or diagonal basis, depending on a bit m_B, and interprets the results the same way as he would in BB84: $|0\rangle$ and $|+\rangle$ as zeroes, $|1\rangle$ and $|-\rangle$ as ones. His measurement can yield a zero, whatever the sent state and the measurement basis; however, if he got a one as result, he knows definitely that he used the conjugate basis—the rectilinear for $|+\rangle$ or the diagonal for $|0\rangle$—, and he knows the corresponding state/bit sent by Alice. Therefore, his measurement basis bits m_B (0 for rectilinear, 1 for diagonal) are perfectly anticorrelated with Alice's key bits s_A whenever he measured $|1\rangle$ or $|-\rangle$. Bob's key bits s_B are then obtained as $s_\mathrm{B} = \overline{m}_\mathrm{B}$. This process is commonly referred to as *unambiguous state discrimination* (USD) [22, 34].

Bob discloses the corresponding time slots to Alice on a public channel, and they keep only the suitable bits for their key. Obviously, assuming perfectly random and independent bits and measurement choices, this only yields a key bit 25% of the time, the remaining three quarters are discarded. After they agreed on their raw keys, the two parties once again sacrifice a random subset of bits to check for errors, deciding to continue or abort the protocol based on the estimated QBER.

Decoy state. Decoy state protocols were invented with a very specific target in mind: to make already existing protocols, especially BB84, less susceptible towards photon number splitting attacks, when a true single-photon source is unavailable. The main idea, originating from a 2003 paper written by Won-Young Hwang [35], is to replace a subset of signal states with different intensities, mostly multiphoton states. These are decoys, and their main role is to check whether the losses for decoy states are significantly smaller than for signal states. Polarizations of decoy pulses are also randomized, so that they cannot be distinguished from multiphoton signal pulses.

Losses for signal and decoy states can be calculated once Bob has received all states, and Alice announced publicly, which of the qubits were decoys. The difference of losses is a clear indicator of PNS attacks, when single photon pulses are blocked by Eve. Although in theory, one would need infinitely many decoy states, realizations with only finite values have been shown to be sufficiently secure [22].

SARG04. Another protocol designed to improve security against PNS attacks in weak coherent state QKD is SARG04 (Scarani–Acín–Ribordy–Grisin 2004). In their paper, the authors provided a general set of protocols, as well as a specific example, which is a modification of BB84 [34]. On the quantum level, the two protocols agree: Alice sends either $|0\rangle$, $|1\rangle$, $|+\rangle$ or $|-\rangle$, while Bob measures the qubit randomly in the rectilinear or diagonal basis. However, the sifting process is radically different, as it denies the comfort of basis disclosure from the PNS attacker. Instead of telling Bob the basis of encoding, Alice announces a pair of non-orthogonal states among the four possibilities, one of which is the state she sent. Now Bob can perform an unambiguous state discrimination measurement—it is not a surprise, that even the authors commented on the similarity between SARG04 and B92—, discarding inconclusive results. Altogether, the raw key rate will be 0.25, as opposed to the 0.5 of BB84. If this is compensated by increasing the mean photon number from 0.1 to 0.2, the attenuation allowed for perfect secrecy is increased so that the distance covered goes up from 50 to around 100 km.

For the attenuation region allowed for SARG04, the PNS attack does not yield full information to Eve [22]; the best she can do is an IRUD (intercept-resend with unambiguous discrimination) attack, where she blocks pulses with less than three photons, conducts an USD measurement, and resends the resulting qubit to Bob—however, this strategy only provides full information for when the attenuation is greater than 25.6 dB. BB84 can be shown to lose security at around 13 dB. These limits can be pushed further out if not only four, but six or more non-orthogonal states are used [34]. Furthermore, it needs to be noted that SARG04 is perfectly suitable for single-photon implementations as well, even if its main purpose becomes superfluous.

E91. Artur K. Ekert's E91 [36] is the odd one out among the listed protocols, as it is based on quantum entanglement. Alice and Bob share a pair of entangled photons (or in the original: spin-1/2 particles), in a singlet state, so that if they use the same basis for measuring their polarizations, the results would be perfectly anticorrelated, assuming a perfect channel and no eavesdropper. Both parties choose randomly between three measurement bases; they use qubits for which their measurements agreed as key bits, and all others (after disclosing the basis choices) to check the Clauser–Horne–Shimony–Holt (CHSH) inequality. The violation of the inequality would mean that the entanglement was not broken—e.g. by a measurement performed by Eve—, and the anticorrelated key qubits can be safely used for encryption and decryption of data. The opposite, however, would tell Alice and Bob to refrain from applying the key, as there is a possibility that someone with malicious intentions tried to access their results.

1.4 Quantum Random Number Generation

Regardless of what the exact key distribution protocol is, the one-time pad technique always requires truly random numbers for information-theoretical security. Here, and from now on, "truly random" refers to numbers—mostly bits—drawn from a discrete uniform distribution. A random variable B_i describing a uniformly distributed bit is defined on the finite sample space $\Omega = \{ 0, 1 \}$ and follows the probability mass function (PMF)

$$P\left(B_i = b\right) = \begin{cases} 0.5, & b = 0; \\ 0.5, & b = 1. \end{cases} \tag{1.17}$$

Consequently, since all bits need to be pairwise independent to assure indeterminacy, n-long bit sequences for any positive integer n are uniformly distributed on $\Omega^n = \{ 0, 1 \}^n$. All protocols, including the transportation of sealed envelopes, need the encryption key to be random; in addition, several DV-QKD protocols also need additional, independent random sequences to decide the encryption and decryption bases on the transmitter and receiver sides, respectively. Aside from uniformity, the random number generators (RNGs) responsible for generating the key bits are required to have a bit generation rate higher than the key transmission rate.

Most RNGs available on computers are able to provide excellent generation rates, and their output sequences show excellent statistical properties, indicating a uniform distribution. However, since they are generated by an algorithm from a starting state (a seed), their outcome is deterministic. For this reason, they are called pseudorandom number generators (PRNGs). Anyone knowing the seed and the algorithm itself can predict future values with absolute certainty. This, obviously, makes them unsuitable for cryptographic applications. Another drawback is that all PRNGs have a predefined maximum sequence length, after which their state returns back to the seed; thus, they provide a periodic sequence of bits. This might be the least of one's concerns, since well-designed PRNGs may have very long periods, such as Mersenne Twisters with an astronomical length of $\left(2^{19937}-1\right) \sim 10^{6001}$ [37]. The extinction of PRNGs is unlikely even in the post-quantum era, as there exist a myriad of applications which do not necessarily require indeterminacy, only an adequately high level of pseudo-randomness. Such areas include, but are not restricted to, scrambling and descrambling of bits in classical communications to disperse the energy evenly within the allowed bandwidth and to avoid burst errors—a weakness of forward error correction algorithms—, as well as Monte Carlo method simulations [38].

Generators that use physical processes to produce random sequences of numbers instead of algorithms are generally called true (or hardware) RNGs. Within this group, one would find devices exploiting the inherent randomness of measurements of a quantum system, *quantum random number generators* (QRNGs). The earliest implementations detected radioactive decay, specifically β radiation, using Geiger–Müller tubes. In a slow-clock implementation [39], the number of detections were counted between rising edges of a clock signal with frequency $f \ll \lambda$, where λ is the average detection rate. On the other hand, fast clock methods such as in Ref. [40], count the rising edges of a clock with frequency $f \gg \lambda$ between successive detections. In either case, counts, which are non-uniform random variables, are used to generate random numbers after the necessary transformations to achieve a distribution close to uniform, e.g. keeping only the parity bit. However, recent research almost exclusively focuses on alternative methods (almost, but that is not to say there have been no new radioactive QRNG architectures introduced since the '70s [41]). This is due to several factors: safety and health issues regarding the storing and management of large amounts of radioactive materials, and practical considerations, since the bit generaton rates are rather limited.

The most popular field of physics in terms of quantum random number generation is indisputably quantum optics. Photonic solutions provide a wide variety of options to harness the entropy of quantum phenomena. Photon sources do not get depleted such as stacks of radioactive materials do, and there are no hazards or safety concerns—unless one decides to look directly into a fiber transmitting high-intensity light, which is fairly easy to avoid.

1.4.1 Optical Quantum Random Number Generators

The extensive 2017 monography *Quantum Random Number Generators* [38] by Miguel Herrero-Collantes and Juan Carlos García-Escartín discusses the different types of optical QRNGs in great detail and also provide a classification as a starting point. *Time-of-arrival* (ToA) generators, using the randomness of time elapsed between photon detections, are discussed in detail in Chapter 3. In general, the research described in both Chapter 3 and 4 deals with specific optical ToA QRNG methods. Different generator principles are shortly outlined here to obtain a better understanding of the multitude of possibilities provided by optics for good quality random number generation. Note that this description is not exhaustive—other principles of operation are e.g. vacuum fluctuations, the phase noise of lasers, Raman scattering and optical parametric oscillators.

Branching path. Branching path generators use the spatial superposition of a qubit to extract entropy. This type of QRNG is perfectly suitable for an introductory case study, due to the simplistic basic idea it is built upon. Of course, this can be made as complicated in a specific realization as one wishes. Imagine a photon hitting a balanced beam splitter, which transmits or reflects it with equal probabilities. In a "macroscopic" experiment with millions of photons, we would measure equal optical power in the two branches; but a beam splitter cannot split a single photon. Now place a single-photon detector (either a single-photon avalanche photodiode or a photoelectron multiplier) in both paths. At most one device will signal the photon's arrival, and one can assign a bit based on which detector "clicked". Branching the path of light can also be done based on polarization, rather than intensity. Assume, for example, a single photon prepared in the state

$$|+\rangle = \frac{|0\rangle + |-\rangle}{\sqrt{2}}. \tag{1.18}$$

Passing it through a polarization beam splitter (PBS) with transmission and reflection axes aligned parallel to the basis states $|0\rangle$ and $|-\rangle$, the two detectors will detect it with probabilities $\left|\langle 0|+\rangle\right|^2 = \left|\langle 1|+\rangle\right|^2 = 0.5$ after losses have been taken into account. Such experiments are described in [42–46]. The light source can be a true single photon source [45, 46] or attenuated pulsed/continuous wave (CW) laser diodes or LEDs [42].

Both of these architectures are highly sensitive towards device non-idealities and detector differences. Real beam splitters always have splitting ratios slightly different from 0.5; the perfect alignment of a PBS is also impossible in practice. Along with the different quantum efficiencies—probabilities of detecting a given photon—of the two detectors, the resulting bit sequences will feature a non-zero bias, favoring either zeroes or ones. The dead time of detectors, a short period during which they become insensitive after a detection, on the other hand, may introduce correlations between successive bits. If the arrival of a photon can be expected during the dead time, it becomes less likely that subsequent bits are going to be equal. Different noise sources (dark count rate, afterpulsing probability) also result in deviations from uniformity. Several of these problems can be eliminated with a simple but clever idea: transform the spatial superposition into temporal superposition, as in Ref. [43]. This can be done by placing a delay line (essentially a short optical fiber section) in one of the branches, then coupling them back together, and using one detector with well-defined time windows. If the photon is detected in the earlier window with respect to a reference trigger signal, assign a zero, if it is detected in the later window, assign a one.

The bit generation rates of branching path generators are limited to several Mbps at best [38, 42], often only reaching tens or hundreds of kbps [43, 45, 46]. The main limitation factors include the repetition rates of pulsed lasers/LEDs, the low efficiency of spontaneous parametric down-conversion used to create heralded single photons, and the dead times or low quantum efficiencies of detectors. Even so, the basic method can be analyzed without much difficulty, the origin of randomness is almost self-explanatory—to no surprise, one of the commercial QRNGs, ID Quantique's Quantis [44] is a branching path generator. One promising alternative to increase the rates is to create superpositions where the number of paths is a higher integer power of 2 [47].

Photon counting. Photon counting (PC) generators, along with time-of-arrival QRNGs, are heirs to earlier radioactivity-based setups, inheriting methods applied to those. PC methods are reminiscent of the slow clock setups mentioned previously. These generators employ single-photon detectors and assign bits based on the number of photons/detections within a fixed time window. The general setup is simplistic: a light source (semiconductor laser or LED) emits pulsed or CW light, which is guided to the detector either in free space or through an optical fiber. The optical intensity should be low enough so that the sensitive receivers do not get damaged. Thus, attenuators are usually placed between the source and detector.

Most PC QRNGs use detectors of limited photon resolving capabilities; they generally provide binary results: no output for no photons vs. a voltage pulse if at least one photon was detected. A popular approach is to count the number of detections within fixed periods of time [48, 49], and assign the parity of the result as a random bit. The dead time of detectors had even proven to be helpful, since it allows for finding average power levels where the resulting bit stream is unbiased. This, however, would be theoretically impossible for the parity-based method if the detectors were ideal [48]. Another option is to use multiple least significant bits (LSBs) from the binary representation of counts, e.g. four in Ref. [50], keeping in mind that this number, as a necessary condition, should not exceed the min-entropy of the source.

More nuanced methods also have been tested to transform the underlying exponential/Poisson distributions into a uniform one. As an example, the authors of Ref. [51] took fixed-length time windows, keeping only those within which only one detection was found. They assigned the binary form of the number of the time bin within the window, where the detection occurred. By cutting off the edges of each window, this resulted in a uniform distribution, yielding high-quality random numbers. (One might argue that this is rather a time-of-arrival generator, but this is a semantic question, which I do not want to get into, and I simply list this generator within the PC group.) Another research group [52] examined the time bins within a large time window, and encoded bits based on how many, and exactly which bins contained a detection. It has also been shown that the dark counts of the detector alone can be used for random number generation, although only at several tens of kbps.

There exist more refined detectors, that are actually able to resolve the number of photons arriving simultaneously (or at least well within the dead time). The generators in Refs. [53] and [54] use a device that is able to discriminate between photon numbers ranging from 0 to 7—the amplitude of its output voltage pulse depends on the number, unlike in case of a “simple” single photon detector. The first of these papers [53] described a system using weak laser pulses as the source of light, comparing successive photon number readouts, and assigning a bit based on their comparison. The second approach [54] is less robust, but more refined: measured numbers were grouped into four bins, where the total probability of each bin is around 25%, as uniformity would demand.

An interesting idea is to illuminate existing image sensors, e.g. the camera of a cell phone, with an LED, then read out the binary representation of each pixel’s voltage. The voltage is, under correct conditions, mainly dominated by the number of detected photons. Using a suitable extractor to increase the entropy of the bit stream, it provides a method for true random number generation with commercial devices [55].

The average bit generation rates exceed those of branching path generators: values in the order of 1–100 Mbps are common, while parallelization of several SPAD pixels can help yield higher rates, e.g. 200 Mbps in Ref. [50]. Another group treated separately in Ref. [38] is that of attenuated pulse generators, but one might argue that these are just a specific subcategory of PC QRNGs. Here, randomness is provided by answering the basic yes–no question “Has there been a detection?”

Amplified spontaneous emission. Amplified spontaneous emission (ASE) is a phenomenon during which spontaneously emitted photons are multiplied through stimulated emission in an optical gain medium. Spontaneous emission is a purely quantum process that cannot be explained by classical electrodynamics [56]. A spontaneously emitted photon has random properties, such as energy/frequency/wavelength—within the range allowed by the widths of relevant energy bands—and polarization. When amplified, the effect produces rapidly fluctuating, easy to measure optical intensity fluctuations, which can be sampled to obtain random bits. ASE-based generators are a promising method of extremely high-speed random number generation.

ASE noise can be obtained from different optical devices. It is the natural self-noise of optical amplifiers, both erbium doped fiber amplifiers (EDFAs) and semiconductor optical amplifiers (SOAs), and it is even more prominent when there is no input signal to boost. Also, superluminescent light emitting diodes (SLEDs) are a type of light source operating based on amplified spontaneous emission. The optical band of ASE noise can cover hundreds of nanometers in the near infrared spectrum, with full width at half maximum (FWHM) values easily reaching tens of nanometers, corresponding to several THz in frequency [57]. Thus, the electrical bandwidth of detected intensity fluctuations is only limited by the narrowest bandwidth element in the light’s path—usually an electronic device, but a well-placed optical filter can help if needed.

Generators in the literature use one or more of the aforementioned devices as ASE sources: SLED [57–61], EDFA or other doped fiber amplifiers [56, 59, 62], and SOA [59]. The bit acquisiton methods encorporate one-bit solutions, mostly employing threshold comparison of samples [62]. In some cases, the inherently asymmetric intensity distribution is symmetrized by differential detection: the signal is split in two using either a polarization beam splitter [56] or a regular 50-50 beam splitter [61], with a small delay introduced to one arm before detection. Since these pairs of signals are assumed to be independent and identically distributed, their difference is a symmetric variable, yielding higher quality random numbers. A particular solution offers a scalable parallelization method: splitting the signal to different disjoint frequency bands, one can individually generate bit streams using each band, then concatenate those for a higher total rate [58]. Once again, this can be done since different spectrum components are independent from each other.

Multi-bit methods are also prevalent, and offer extraordinarily high generation rates. Deliberately oversampling the intensity noise, and digitizing the measured samples at 16 [59] or 32 [60] bits is one way to go. These raw streams are heavily correlated, but this can be eliminated by removing some of the most significant bits (MSBs). The article in Ref. [57] offers both a single-bit and three multi-bit encoding options for the same physical hardware: the noise is split into two, one path having a delay. Two bit streams are generated for both arms, with the correlations and the possible bias being decreased by taking their bitwise XOR, and the potential discarding of MSBs.

Our research group also conducted experiments regarding ASE QRNGs, mostly focusing on the effect of sampling rate on the quality of randomness [2, 13]. The ASE source was an SLED, an the intensity noise was further amplified using an EDFA. To enhance the differences between sequences obtained by different sampling rates, we introduced a small deliberate bias into the bit creation by slightly detuning the comparison threshold.

The bit generation rates of ASE generators are generally higher than those of previously mentioned methods. These values regularly reach 1–20 Gbps [56–58, 61, 62], but can exceed several hundreds of Gbps (multibit methods in Refs. [57] and [59]), or even reach 1.6 Tbps [60]. Lately, a much needed detailed quantification of randomness—neglected by most prior publications—was also conducted for ASE-based QRNGs [63].

1.4.2 Randomness Testing

Given a bit sequence of finite length, it is impossible to certainly prove or disprove whether or not its source was a generator producing uniformly distributed bits [38]. However, if a device frequently outputs "suspicious" sequences, one might be tempted to express doubt about its quality. One way to quantify the confidence in the uniformity of a generator is statistical hypothesis testing.

During hypothesis testing, a so-called *null hypothesis* H_0 and its *alternative hypothesis* H_1 are stated, the two being disjoint events. Then a relevant statistic S (a function of the samples) is chosen, and the reference distribution is calculated for S under the assumption that H_0 is true. From the observed value of the statistic, a so called *p-value* is computed, which is the probability of observing a test statistic as extreme as the one calculated, assuming that H_0 is true. After that, a *significance level* α is selected; this is the probability that H_0 is rejected given that it was true (type I error). Type II errors are, on the other hand, when H_0 is wrongly accepted. The two types of error cannot be minimized simultaneously. The null hypothesis is then accepted if and only if the p-value is greater than α; otherwise, H_0 is rejected/H_1 is accepted [64].

For randomness testing of RNGs, the null and alternative hypotheses are generally

$$H_0 = \{\text{the bit sequence was produced by a RNG with uniform distribution}\} \text{ and}$$
$$H_1 = \{\text{the bit sequence was produced by a RNG with non-uniform distribution}\}.$$

In case of cryptographic applications, the significance level is often set to $\alpha = 0.01$, although this is no more than a *de facto* standard, and other values might be more useful if type II errors need to be minimized [64].

However, no single statistic (or no finite number of different statistics) are enough to completely quantify the randomness of a bit sequence. Therefore, testing of RNGs is usually done using test suites, which contain multiple, more or less independent tests, analyzing the randomness from different perspectives. Examples of such aspects are the following: the *bias*, representing the difference of relative bit frequencies from 0.5; periodic features, which can be observed with the help of discrete Fourier transform; the autocorrelation coefficients, capable of implying dependencies between bits at a certain distance; the number of *runs*—successive, uninterrupted subsequences of identical bits—, showing how often the generator switches between zeros and ones on average; etc. Obviously, a truly uniform RNG would have zero bias and zero autocorrelation for lags other than zero. For each perspective, similar expectations and corresponding reference distributions can be found based on the null hypothesis.

There exist several well-known suites designed for the randomness testing of RNGs, such as the Statistical Test Suite of the US National Institute of Standards and Technology (NIST STS) [65], TestU01 [66], the DieHard [67] and DieHarder [68] batteries. Neither of these were designed with true RNGs in mind; therefore, a high-quality PRNG is expected to pass all tests within these suites. For a QRNG, passing is not a sufficient proof of randomness, but a necessary condition towards accepting it as uniform.

During the research leading to this thesis, I exclusively worked with the NIST STS. It consists of 15 different statistical tests, some of them having multiple subtests, from which there are 188 altogether. The exact testing process I adopted is as follows:

1. The significance level is left at its default, $\alpha = 0.01$.

Table 1.3: Test parameters used in randomness testing throughout the thesis.

Test	**Block length**	**Default**
Block frequency	10001	128
Non-overlapping template	9	9
Overlapping template	9	9
Approximate entropy	10	10
Serial	16	16
Linear complexity	500	500

2. All statistical tests in the suite are conducted on $L_\mathrm{s} = 10^6$ long subsequences taken from a file, which contains at least 10^9 random bits. Each test is run on $N_\mathrm{s} = 1000$ distinct subsequences in total. Some tests require parameters, these are selected based on input size recommendations. The parameters are shown in Table 1.3. As it can be seen, only the block length M of the block frequency test is changed from its default value of 128, to comply with the recommendation ($M > 0.01 \cdot L_\mathrm{s}$).

3. After the testing has been finished, the software evaluates the obtained p-values with respect to two different aspects. If either of them fails, the test result is deemed a failure. A subsequence in itself passes a test if its calculated p-value is greater than or equal to α.

    - First aspect: *passing proportion.* The proportion of subsequences passing a certain test should be within the interval

    $$1 - \alpha \pm 3\sqrt{\frac{\alpha\,(1-\alpha)}{N_\mathrm{s}}}. \tag{1.19}$$

    For $\alpha = 0.01$ and $N_\mathrm{s} = 1000$, this means that at least 980 but no more than 999 subsequences should pass. Note that since a perfect generator produces any possible bit sequence of a certain length with the same probability, it is expected that some of them do not “seem” random.

    - Second aspect: *uniformity of p-values.* For a given test, the N_s distinct p-values should be distributed uniformly between 0 and 1. To check this, the test suite breaks the unit interval into ten disjoint bins of width 0.1, counts the p-values found within each bin and conducts a uniformity χ^2-test on these numbers. If the newly calculated uniformity p-value exceeds 0.0001, the hypothesis of uniformity is accepted.

> This process is repeated for the desired subset of the 188 tests—generally all of them.

Raw bit streams output by QRNGs do not always pass all tests, either due to device imperfections, or an inherent non-uniformity of the method. In such a case, it is possible to increase the quality using post-processing methods called randomness extractors [38], which are either deterministic algorithms or require a short, reusable random seed. Randomness extraction usually results in shorter but more uniformly distributed sequences, the amount of losses and the increase of quality depending on the complexity of the algorithm. Although post-processing is an important and well-researched subtopic of random number generation, it is only worth a short mention in context of this dissertation. Personally, I am in favor of designing QRNGs which do pass statistical tests without randomness extraction, as it will become clear in Chapters 3 and 4. This is in opposition to several authors, who prefer generating non-uniform data at higher rates, and then letting the algorithms do their job.

Chapter 2

DV-QKD Protocols with Polarization Modulated Vertical Cavity Surface-Emitting Lasers

2.1 Introduction

Vertical cavity surface-emitting lasers (VCSELs) are semiconductor laser diodes, with a defining characteristic that light propagation is perpendicular to the active region. This is in contrast with more conventional edge-emitting lasers (EELs), like Fabry–Pérot, distributed Bragg reflector and distributed feedback lasers, where emission is parallel to the active region. The main importance of VCSELs is that they can be manufactured and tested on-wafer at large scales, allowing for reduced production costs [69]. Since technological advances have already made it possible to manufacture good quality, reliable VCSELs, they have been mass produced for use in laser printing, optical mice, sensing, and plenty of other applications [70].

In this chapter, I show why VCSELs make good candidates for light sources in weak coherent state DV-QKD implementations, describing their advantages and the potential disadvantage of a unique feature called polarization switching. Based on this, I then propose a new transmitter design for the BB84 protocol that exploits controlled polarization switching of VCSELs, allowing to reduce the number of elements in the device.

2.2 VCSEL Advantages in Low-Power Applications

DV-QKD protocols require very low power levels almost by definition—especially if implemented using quasi-single photon states. Assuming a pulse repetition rate of 100 MHz and a mean photon number of 0.1 per pulse at a wavelength of 850 nm, the average optical power emitted by the transmitter is approximately 2.337 pW. Such a transmitter is inevitably lossy, as practical pulses of off-the-shelf laser diodes are several orders of magnitude stronger than this. Therefore, optical attenuators are required to reduce the power level significantly. Inherently low-power solutions would lead to lesser necessary attenuation, being more energy efficient and environmentally friendly—although this is certainly not the application, where cutting the power waste would resolve the problems of climate change.

VCSELs possess several beneficial properties regarding this matter, specifically when compared to edge-emitting lasers. They typically come with lower threshold currents, lower output powers and high power conversion efficiencies, offering decreased total losses. Also, due to their symmetric cross-section, their emitted beams are circular, unlike the elliptical profile typical of EELs. Additionally, the divergence angles of beams are smaller, making it easier and more efficient to couple their light into optical fibers, or collimate it for free-space applications [69]. It comes as no surprise that there have already been reports of QKD modules equipped with VCSELs [71].

2.3 Polarization Switching in VCSELs

Although the benefits of VCSELs are clear for low-power applications, there exists a distinctive feature of surface-emitting lasers which can be a significant disadvantage. This phenomenon is *polarization switching* (PS), whereas the state of polarization of the emitted light can suddenly change (switch) between two orthogonal linear polarizations. If the application is polarization-sensitive—which is the case for single-photon qubit implementations of DV-QKD protocols—, a VCSEL with unstable polarization is certainly not suitable for the task.

PS is not a universal property of VCSELs; even devices from the same wafer can show differences. Nowadays, there are several methods to mass-produce polarization-stabilized versions. The main driving force behind this field of research was not the developent of QKD transmitters, but slightly more mundane devices, such as optical mice. Plenty of options have been analyzed in detail, such as providing external optical feedback for stabilization, using polarization-dependent mirrors or asymmetric resonators, making the optical gain polarization-dependent, etc. However, the one that eventually stood out and became widely accepted due to reliability is stabilization achieved by applying shallow surface gratings [72].

With all due respect to those involved in investigating polarization stabilization, I am not going to discuss those topics any further. For the purposes of this chapter, it is significantly more important to understand the reasons behind polarization switching, and discuss its consequences in DV-QKD protocols—both the drawbacks and the proposed benefits, when PS happens in a controlled manner. The latter may be called *polarization modulation*, and it will be explored in detail in Section 2.4.2.

2.3.1 Origins and Causes of PS in VCSELs

Polarization switching is unique to VCSELs, in the sense that it is not expected for properly designed edge-emitting laser diodes. The circular symmetry of the device alone would not be able to select preferred states of polarization. However, symmetry is never perfect due to manufacturing imperfections and the crystallographic structure of the materials. The directions of the crystallographic axes will eventually correspond to two *orthogonal* states of polarization, called polarization eigenmodes, denoted as $\hat{x}$- and $\hat{y}$-polarized modes from now on. As every VCSEL exhibits anisotropies—different properties in different directions—, the presence of linear phase anisotropy (γ_{p}), or birefringence, leads to a frequency split of $2\gamma_\mathrm{p}$ between the eigenmodes, on the order of several (tens) of GHz [73].

The PS mechanism can be roughly described as follows. When injection current is increased above threshold, one of the eigenmodes becomes dominant with the other being suppressed. Note that as per the convention in Ref. [73], it is useful to define the injection current relative to threshold as

$$\mu = \frac{I}{I_\mathrm{th}}, \tag{2.1}$$

where I is the injection current and I_th is the threshold current. If one increases μ further, a sudden and abrupt switch can be experienced to the orthogonal eigenmode. Decreasing the current again, a back-switch is to be expected to the originally dominant mode. If a PS at increasing current occurs from the higher frequency mode to the lower frequency one, it is called a Type I PS; the low-to-high frequency change is a Type II PS. See Fig. 2.1 for a simple example.

A given VCSEL can have multiple switching points as well, e.g. a Type I switch followed by a Type II or vice versa, but it can also lack the PS mechanism totally. The first switch (closest to threshold) generally happens while only the fundamental transverse mode is lasing; for switches at higher currents, multiple transverse modes can be active simultaneously, generally leading to a decreased suppression ratio between the two polarizations. Another distinctive feature is polarization switching with hysteresis, also encountered in several devices. In such a case, the switching point for increasing and decreasing current is not the same, showing a "memory effect" included in the selection of polarization. Figure 2.2 illustrates a PS with hysteresis.

The polarization switching of VCSELs was studied in-depth during the '90s, with special emphasis placed on its origins. The earliest explanations by Choquette et al. suggested that the dominant effect was thermally induced [74]. The two eigenmodes are distinct in wavelength, and the material gain curve is frequency-dependent; thus, $\hat{x}$- and $\hat{y}$-polarized modes experience different gains. Around threshold, the one with higher gain starts lasing, suppressing the orthogonal mode. However, increasing the injection current heats the material and red-shifts the gain spectrum, so as the relative gain of the eigenmodes is reversed, resulting in a polarization switch. This reasoning explains Type I PS in gain-guided VCSELs, but not Type II [72], and also fails to account for the potential hysteresis associated with switching.

San Miguel, Feng and Moloney created a significantly more difficult model (SFM or spin-flip model), incorporating four magnetic sublevels [75] and their respective population dynamics, extending the set of rate equations governing the operation of semiconductor lasers. The parameters involved in the SFM model introduce fast,

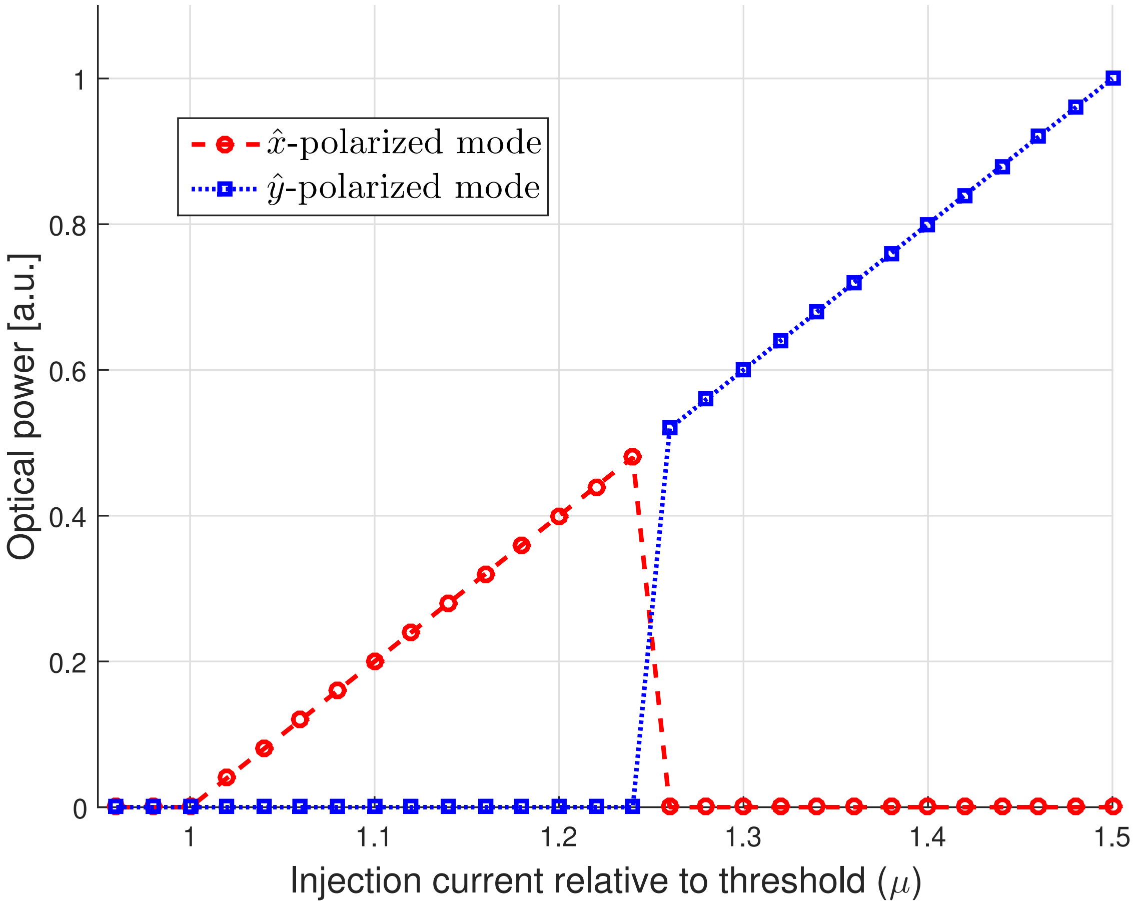

Figure 2.1: Example of PS in a polarization-resolved light-current characteristic. The $\hat{x}$-polarized eigenmode starts lasing as the current is increased beyond the threshold $\mu = 1$, and retains its polarization until the current reaches $\mu \approx 1.25$. Here, a polarization switch occurs, whereas the $\hat{x}$-polarized mode surrenders its stability conditions to the $\hat{y}$-polarized, but the total output power still increases linearly. Taken from Ref. [14]; based on figures in Ref. [73].

short time-scale mechanisms to the explanation of polarization switches. These are the decay rate of the electric field in the cavity and that of the total carrier number; the spin-flip relaxation rate γ_s describing the mixing of carriers with opposite values of angular momentum; the linewidth enhancement factor α; the frequency split between the modes; etc.

The model's validity was proven in the paper of Martín-Regalado et al. [73] via an extended numerical mode stability analysis. The parameter set was slightly redefined (but not changed in meaning) to include the phase anisotropy γ_p and the amplitude anisotropy γ_a. Amplitude anisotropy is the product of both gain and loss anisotropies—the latter is also known as dichroism. It has been found that polarization switching is indeed explained if $\gamma_\mathrm{p} \neq 0$ and $\alpha \neq 0$, while γ_a is also given a small but nonzero value. The switching type depends on the sign of the amplitude anisotropy: $\gamma_\mathrm{a} < 0$ corresponds to a Type II switch with increasing current, whereas

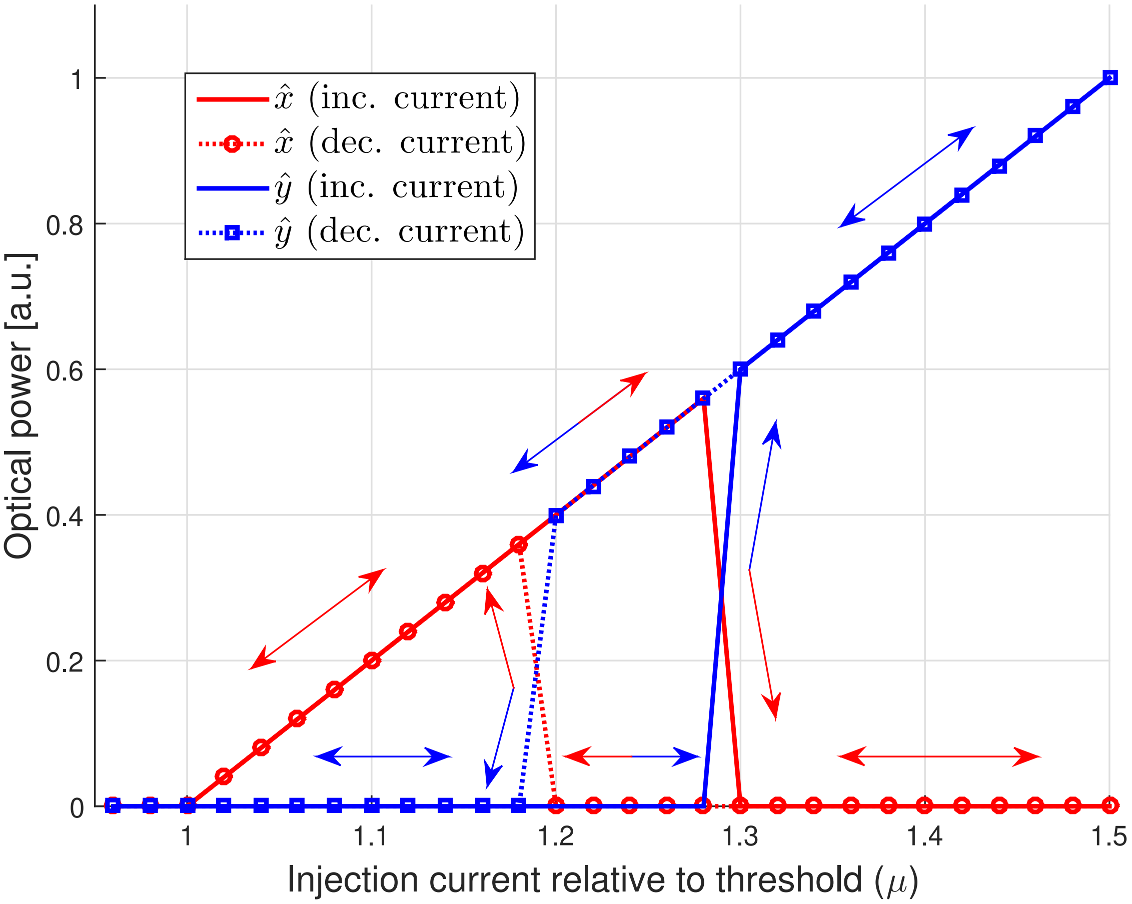

Figure 2.2: Example of PS with hysteresis in a polarization-resolved light-current characteristic. Once again, as current is increased above threshold, the $\hat{x}$-polarized mode becomes stable. With increasing current, switching to $\hat{y}$ occurs around $\mu = 1.29$; with decreasing current, $\hat{x}$ becomes dominant below $\mu = 1.19$. It can be concluded that the region $1.19 < \mu < 1.29$ is bistable. Arrows help determine the changes in current along different sections of the curves. Taken from Ref. [14]; based on figures in Ref. [73].

$\gamma_a > 0$ results in a Type I switch under the same conditions [76]. Type II switches are not abrupt: there are intermediary elliptic and unpolarized states near the switching point. Another notable result is the finding of large bistable regions of the parameter space also involving μ, where in principle both eigenmodes are stable. This allows for hysteresis cycles, since in these regions the already lasing state is retained, and a switch is only expected after crossing into a region where only the orthogonal state can be stable [77]. The SFM model's predictions generally show great agreement with experimental results.

As a sidenote, I need to mention that polarization switching is also strongly influenced by external optical feedback into the laser cavity, depending on both the angle and strength of the incoming light [78].

2.3.2 Increased QBER Induced by Polarization Switching

For DV-QKD protocols implemented with single photon polarization qubits, the unwanted switching of polarization to the orthogonal state will introduce errors in the raw keys of the participants. If the rate of switches is high enough, the key distribution will be aborted even in the absence of eavesdropping, due to high levels of QBER. This effect is best illustrated on basic QKD protocols: let us look at BB84 and B92.

BB84 Example. In case of BB84, if the basis choices of Alice and Bob are independent, then, on average, they agree 50% of the time. It is also safe to assume that polarization switches happen independently from basis choices. Since a PS happens between orthogonal states, it will certainly lead to a quantum bit error in the raw key if the chosen bases are the same. These key bits can cause two types of problems: a false alarm, if they are picked for the key sifting process, or a difference in true keys, leading to incorrect decryption. However, those switches in qubits, for which the choises differed, can be neglected, as these will not manifest in the creation of a key bit. Table 2.1 shows a simple key distribution scenario depicting all possible situations. Polarization switches and differing raw key bits are denoted by red numbers. Qubit #2 features an ultimately harmless switch, whereas qubit #4 contains a key bit mismatch. Altogether, every second PS on average is responsible for errors.

Table 2.1: Polarization switching effects in BB84. Taken from Ref. [11] and Ref. [14].

Qubit number	**1**	**2**	**3**	**4**	**5**	**6**
Intended state (s_A, m_A)	↑	↗	↗	→	↑	↘
Sent state ($\tilde{s}_\mathrm{A}$, m_A)	↑	↘	↗	↑	↑	↘
Measurement basis m_B	+	+	×	+	×	×
Alice's raw key s_A	1		0	0		1
Bob's raw key s_B	1		0	1		1

Model the polarization switch as a random variable, where the surviving/measured photon from the original wave packet has the correct polarization with probability $1 - p_\mathrm{S}$ and the orthogonal with p_S. If Eve is not present and the channel is assumed to be perfect—that is, transmitting quantum states without altering them—, this directly translates to a QBER value p_S, since switches are independent from basis choices, and each raw key bit born from a switched qubit will be erroneous.

If there is an eavesdropper, the QBER introduced by the PS depends on the strategy of Eve. During the Intercept-and-Resend (I&R) attack, Eve measures the polarization in a randomly chosen basis, then sends a new qubit prepared in the state she measured. The basis choices of Bob and Eve are independent. If Bob chose correctly, but Eve did not, there is a chance that the resulting raw key bits will differ. Including th PS effects into this framework, the analysis can be performed by a simple enumeration of all relevant possibilities; we need only look at cases where Bob chose correctly [12].

1. If Eve chose correctly as well (50% of examined cases), she measures the correct state with probability $1 - p_\mathrm{S}$ and the wrong state with p_S.

2. If Eve chose the conjugate basis (50% of examined cases), her measurement results are still completely random (with probabilities 0.5, since PS happens between orthogonal states.

Averaging these two, the total QBER is $0.25 + 0.5 \cdot p_\mathrm{S}$. Note that I&R in itself causes an error rate of 0.25 when everything else is thought of as ideal; therefore, the added QBER due to switches is halved compared to the situation where Eve is not present.

B92 example. B92 has three, rather than two different ways through which a polarization switch can affect the process. Following the previous model, let us suppose that a switch happens with probability p_S. If Bob measures the switched state in the correct basis—once again, using the independence of choices, this happens half the time—, the result of measurement would be surely $b = 1$, but there will be a certain mismatch between a and $\overline{a'}$, causing a raw key bit error. However, if Bob measured in the wrong basis, switches would not introduce any errors. A result $b = 0$ would lead to discarding both a and a', whereas in case of $b = 1$, $a = \overline{a'}$, which is not problematic.

Table 2.2: Polarization switching effects in B92. Taken from Ref. [14].

Bit number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Intended state a	↗	↗	→	→	→	↗	→	↗
Sent state $\tilde{a}$	↗	↘	→	→	↑	↘	→	↗
Measurement basis a'	+	+	×	+	+	+	+	×
Measurement result b	1	0	0	0	1	1	0	0
Alice's raw key a	1				0	1		
Bob's raw key $\overline{a'}$	1				1	1		

Since there is no direct analogy to an I&R attack as described for BB84 in case of B92, I am not discussing here the QBER resulting from PS and any type of eavesdropping.

2.4 A Proposed New Transmitter Structure for the BB84 Protocol

So far, the problems of unwanted polarization switches have been discussed, However, I also foreshadowed the potential benefits of controlled polarization switching in Sect. 2.3. My exact proposition, further elaborated in this section, is to take advantage of the PS mechanism in VCSELs by deliberately modulating them between two orthogonally polarized states and exploit it in BB84 implementations using single photon polarization qubits.

2.4.1 Trivial and Proposed Transmitter Designs

Assume an implementation of the BB84 protocol with quasi-single photon polarization qubits, built using attenuated lasers instead of true single-photon sources. Arguably, the trivial option to build a transmitter for this application is to take four distinct semiconductor lasers—either edge- or surface-emitting—with linearly polarized light, align them in such a fashion that their polarization is exactly one of the four BB84 states $|0\rangle$, $|1\rangle$, $|+\rangle$ and $|-\rangle$ with polarization angles 0°, 90°, +45°or −45°, respectively (Fig. 2.3). The first string of random numbers, the *selection bits* are responsible for basis selection; essentially, they pick one pair of lasers. The *key bits* then choose the desired laser from the pair. This configuration requires four individual semiconductor lasers in one-to-one correspondence with the four states, as seen in Ref. [30], for example.

One could think of plenty of different setups for achieving the same goal. Let us focus on a solution, which I first outlined in Ref. [11], and later expanded the analysis in Ref. [14]. This newly proposed transmitter design contains only two lasers; namely, VCSELs which exhibit PS and can be modulated on demand between the two eigenmodes. Since polarization switching happens between two orthogonally polarized states, one such laser is suitable, in theory, to transmit both states found within a certain basis. There is no one-to-one correspondence between light sources and polarization angles anymore, rather between light sources and bases: one VCSEL is oriented so that its eigenmodes lie along the directions of the diagonal basis, whereas the other represents states in the rectilinear basis. Selection bits choose

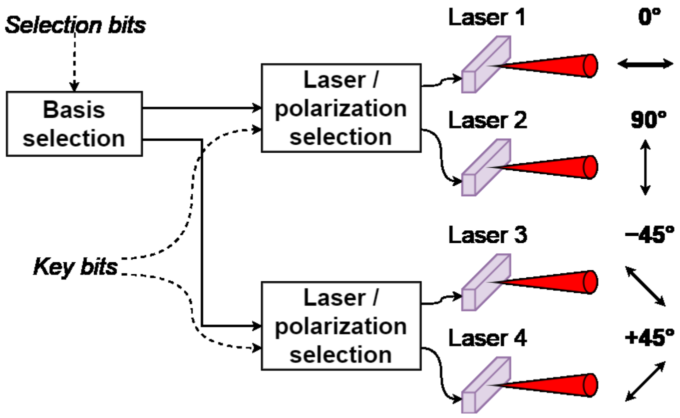

Figure 2.3: The trivial BB84 transmitter setup with four linearly polarized lasers. Taken from Ref. [11] and Ref. [14].

one specific VCSEL along with its basis, while key bits decide how that VCSEL is modulated to achieve lasing in the desired polarization (Fig. 2.4).

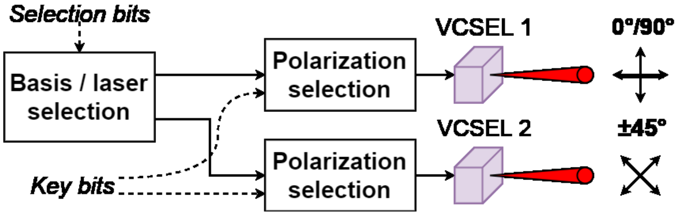

Figure 2.4: The proposed new BB84 transmitter setup with two polarization modulated VCSELs. Taken from Ref. [11] and Ref. [14].

The most obvious benefit of this setup is that it uses two lasers instead of four, which could lead to cost reductions compared to the trivial design. However, there are some difficulties that need to be addressed before implementation, among which the most significant is the following question: How does one achieve polarization modulation in practice, and which option is best for the examined application?

2.4.2 Polarization Modulation of VCSELs

To obtain an answer, I first gathered the possibilities that have either already been tried, or theoretically possible for the purpose of polarization modulation. Those which are either *impractical* for the BB84 protocol's implementation or seem *unrealistic* with respect to realization constraints, need to be excluded from the list of candidates. There are three main groups: external optical injection, modulation of a parameter playing a role in polarization switching, and the use of specially designed VCSELs. By injecting external, linearly polarized light into the laser cavity, the dominant polarization eigenmode can be switched on demand. If this is done with the help of additional "master" lasers, as seen in Ref. [79], the benefit of decreasing the number of lasers is lost. Such a solution is better suited and more prosperous for all-optical signal processing. External optical injection can also manifest as polarization self modulation (PSM), if part of the laser's own light is reflected back into the active region after 90°polarization rotation. Polarization self modulation means that the two orthogonal states of polarization are modulated periodically in opposite phase, with the modulation frequency depending on the effective cavity length. Although this phenomenon is not exclusive to VCSELs, their shorter cavities can result in higher frequencies of modulation; tens or hundreds of gigahertzes were estimated from a theoretical point of view for general semiconductor lasers [80], and 6 GHz was experimentally demonstrated in VCSELs as early as the first part of the 90's [81, 82]. The PSM operation differs from that expected in my case, rendering it useless for the transmitter design: light sources in a BB84 transmitter should not send periodic signals.

A second approach seems more straighforward: choose one of the parameters influencing the stability regions of each eigenmode and modulate it so that the laser can cross the boundaries between those regions. The parameter needs to be easily accessible for modulation; this excludes built-in parameters of the laser (α, γ_p, γ_a, $\ldots$), even if they are slightly current-dependent. Moreover, the related time constants should be short enough so that high-frequency modulation would be possible. Even if some switches are thermal in origin, the slow heating and cooling speeds prevent quick changes between polarization states. Since switches are also present at a constant active region temperature [83], temperature control is even desirable for repeatability. Altogether, current-induced polarization modulation is the only promising option both in terms of implementation and speed.

Most of the earlier studies focused on sinusoidal current modulation around a bias point close to the DC polarization switching current. Gain-guided circular VCSELs, where most of switching is attributed to thermal effects, are significantly limited in terms of polarization modulation. Choquette *et al.* reported in 1994 that the maximum frequency where a given laser exhibited switching was only 80 kHz [84]. Verschaffelt and colleagues analyzed a different VCSEL in 2002 [85], and defined polarization modulation as successful if switching happened at least in 80% of signal periods. The highest frequency fulfilling this criterion is 90 kHz—the same order of magnitude as in the earlier study. In case of index-guided VCSELs, however, switching can be mainly attributed to mechanisms much faster than heating or cooling [86]. As a result, one could expect faster polarization modulation frequencies for this type of device, with early results achieving 50 MHz [87]. The last two studies also briefly mentioned pulsed mode current-induced polarization modulation, when lasers are biased near threshold, the current is suddenly increased above the DC switching point, then decreased back again. In Ref. [86], the researchers did observe PS for 10 ns long current pulses, whereas in Ref. [87], the state of polarization turned out to be stable when the modulation current consisted of 22 ns long pulses at a repetition rate of 1 kHz. Altogether, pulsed mode PS modulation has not been extensively researched so far to the best of my knowledge.

A 2018 study biased a VCSEL with hysteresis-free PS at the switching current, and modulated it with 10 µs long pulse-per-second (PPS) signals featuring rapid rise and fall times [88]. With the help of polarization controllers, the PPS signal was then recovered after one-way and back-to-back propagation in a telecommunication fiber; the measured pulse width was 9.98 and 9.97 µs, respectively, with some timing jitter experienced between successive pulses.

Barve *et al.* conducted extensive research on current-induced polarization modulation of VCSELs. They have shown experimentally that biasing a laser with a DC shifted and gated RF signal (4 GHz carrier frequency), where the gating function is essentially the modulation content, a commercial-type VCSEL with known DC PS dynamics can be modulated at the very high frequency of 1.35 GHz [89, 90]. The research group also investigated asymmetric index-guided VCSELs in their multiple transverse mode regime, and achieved very high extinction ratios compared to earlier results [91]. As an extension to this work, they could produce a polarization modulation frequency around 300 MHz by periodically modulating the RF signal's carrier frequency and keeping its power constant, and 1.5 GHz by modulating its power and keeping the carrier frequency constant [92]. The asymmetry provided by an elliptical mesa in these devices may be thought of as a special VCSEL design, but less extreme as those introduced in the following paragraphs.

The third possibility is that of VCSELs designed specifically for exploiting the possibilities of using both polarizations. As long as the costs of the unique design do not exceed the price of another laser device, these solutions should also be interesting for application in BB84; the distinct features perhaps even make them more fitting.

Barve *et al.* continued their work by designing VCSELs for asymmetric current injection through two pairs of orthogonally placed electrodes. This way, the two orthogonal polarizations can be independently modulated, with a data rate of 4 Gbps having already been demonstrated [93, 94].

In another study, index-guided cruciform VCSELs (with cross-shaped transverse cavity geometries) have been shown to produce on-demand polarization switching if there is a small-signal current modulation around the DC switching point. Switching at a modulation frequency of 50 MHz was observed in 1994, limited by the modulation source [84]. Large-signal pulsed operation was also analyzed for the same devices; periodic current variations between treshold and slightly above the switching point caused the power in the eigenmode being stable at small currents to exhibit a double-frequency pulsing. Between pulses, the laser was either biased above the PS point or close to threshold, causing the intensity to drop. For frequencies above 10 MHz, however, the second pulse in a period gradually loses its power and ultimately disappears.

2.4.3 Realization Possibilities

Although the proposed design looks relatively simple in its block diagram form (Fig. 2.4), the true physical implementation requires careful considerations so that the transmitter is capable of the desired operation, without giving extra options of eavesdropping to adversaries. I introduce two proposals (options A and B) for current-induced polarization modulation, then I discuss the practical differences between them. During discussion, suppose that the given VCSEL operates around the lowest current PS point, below and above which $\hat{x}$- and $\hat{y}$-polarizations are dominant, respectively. The switching currents are μ_S if there is no hysteresis and $\mu_\mathrm{SL} < \mu_\mathrm{SH}$ if there is.

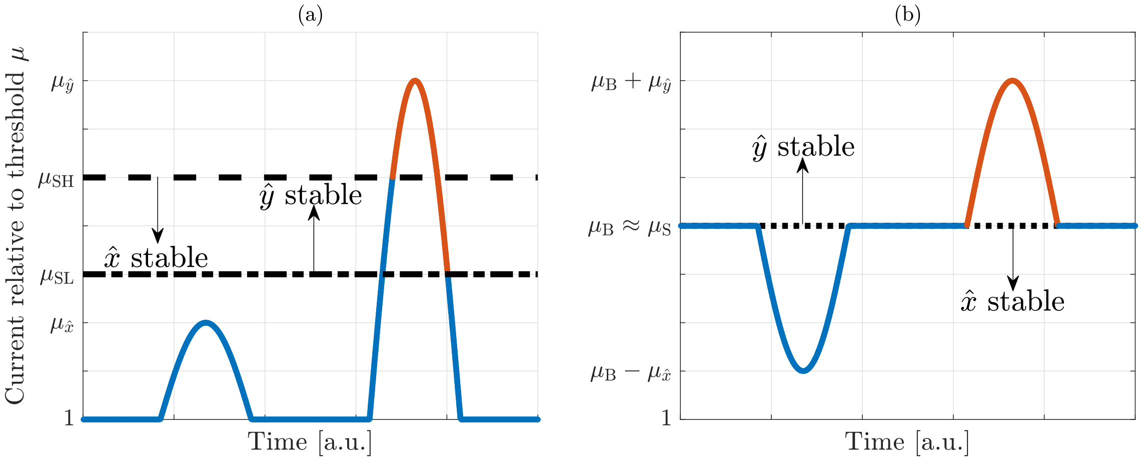

Figure 2.5: Polarization modulation options. (a) Option A: bias at the threshold and different current pulse amplitudes; (b) option B: small-signal modulation around the bias in the DC switching point. First pulse is $\hat{x}$-polarized, second pulse is $\hat{y}$-polarized for both subplots, whereas the common baseline is the threshold current $\mu = 1$. Note that as an arbitrary choice, hysteresis is present in (a) but not present in (b). Arrows denote stability regions; blue and red lines correspond to $\hat{x}$- and $\hat{y}$-polarized portions of the output light signal, respectively.

Option A (see Fig. 2.5(a) for an example) sets the DC bias current of the VCSEL close to threshold. For $\hat{x}$-polarized light pulses/photons, the laser is modulated with a current pulse of amplitude $\mu_{\hat{x}} < \mu_{\mathrm{S(L)}}$ which does not cross into the bistable/$\hat{y}$-stable regions, whereas a $\hat{y}$-polarized light pulse is obtained by a current pulse of amplitude $\mu_{\hat{y}} > \mu_{\mathrm{S(H)}}$. Option B (Fig. 2.5(b)), on the other hand, sets the bias μ_{B} at the switching point ($\mu_{\mathrm{B}} \approx \mu_{\mathrm{S}}$) or within the bistable region ($\mu_{\mathrm{SL}} < \mu_{\mathrm{B}} < \mu_{\mathrm{SH}}$). $\hat{x}$- and $\hat{y}$-polarized light pulses are achieved by superimposing a current pulse of *negative* amplitude $\mu_{\hat{x}}$ or *positive* amplitude $\mu_{\hat{y}}$ onto μ_{B}, respectively. Both options have benefits and disadvantages as well.

A common problem is that different bits (polarizations) are obtained with different current pulses. Since semiconductor lasers have an approximately linear current–optical power characteristic just above threshold, the intensity waveform of light pulses will closely resemble the shape of current pulses, and amplitude differences will result in proportional optical power differences. Additionally, the state of polarization may not be constant during the whole duration of a pulse. This is especially problematic for $\hat{y}$-polarized pulses in option A, which need to go through the $\hat{x}$-stable region of the parameter space to reach the intended polarization, then back again; therefore, the start and beginning of the light pulse might be orthogonally polarized to what is desired. VCSELs that are used in such a transmitter should be carefully analyzed, so that the temporal evolution of polarization within a pulse is well understood. Some of these issues may also arise in trivial transmitters, as no two lasers are perfectly identical.

Obviously, different pulse shapes, sizes and incorrectly polarized fractions of pulses are not to be allowed. A high degree of temporal overlap between signal shapes corresponding to different bits is necessary, otherwise eavesdroppers might gain information from the time of arrival. Differing optical power levels require differing attenuation to reach the same mean photon number. Wrongly polarized photons could lead to an increased QBER even in the absence of eavesdropping. Altogether, the following features are necessary: pulse shaping, which uniformizes the optical powers as a function of time, and acts as gating to block incorrectly polarized portions; and fast variable attenuation to bring the average power level to a common value.

Both of these can be achieved by employing a lossy optical modulator. Electro-absorption modulators (EAMs) are a formidable choice: these change their absorption coefficients as a function of an externally applied electrical field/voltage [95]. 3 dB modulation bandwidths can be really wide—values between 36 and 55 GHz are reported in several exemplary studies [96–98]—, which makes them suitable for e.g. 56 Gbps data rate in classical optical communication systems, and also for high key transmission rates in BB84. High-speed devices sometimes exhibit low extinction ratios, but the extinction ratio they provide (2.7–4.8 dB [96, 98]) is enough to cover the differences between different pulse powers, since the first DC PS points are generally close to threshold. Note that the attenuation of a single EAM is not enough to reach the quasi-single photon power level; therefore, the modulator should be cascaded with an optical attenuator, responsible for the bulk of necessary losses.

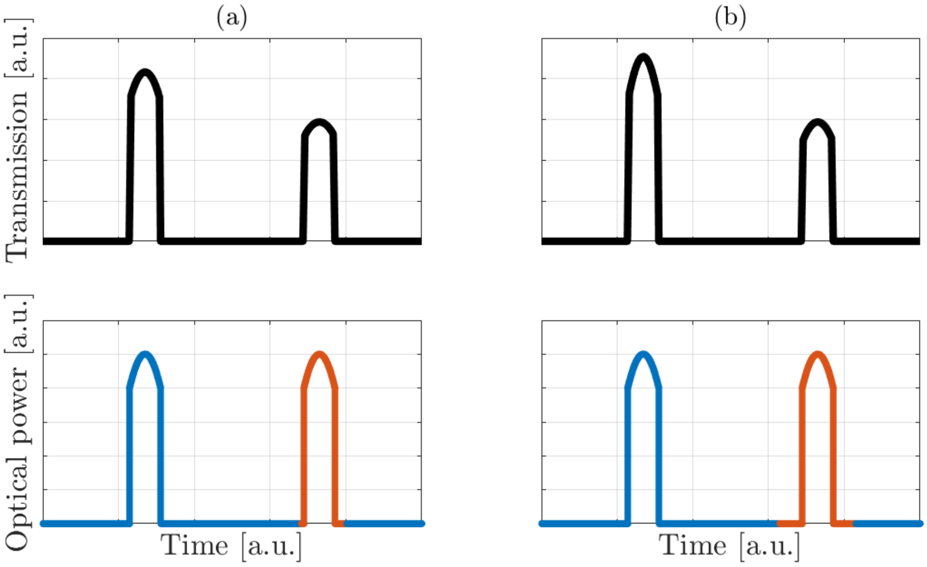

Figure 2.6: Top row: EAM transmission T_{EAM} as a function of time. Bottom row: corresponding output optical power signals. (a) and (b) columns refer to options A and B, respectively, with the input power signals being proportional to the current signals above threshold, as seen in Fig. 2.5. Blue and red lines correspond to $\hat{x}$- and $\hat{y}$-polarized portions of the output light signal, respectively. The baseline is zero for all figures.

The main differences between options A and B are also important to notice. As mentioned in the previous section, current-induced polarization modulation of VCSELs is significantly better-known for small-signal modulation around the switching point than pulsed-mode operation, making option B more promising at first glance. Moreover, one could expect that pulses in option B have a higher polarization extinction ratio, since there is no need to cross a region in which either of the polarizations is instable before reaching the intended current values. However, the negated pulse shapes of B require two distinct pulse shaping voltage signals driving the EAM, whereas the shapes in A only differ in amplitude. Also, the average output optical power is non-zero for option B, meaning that the modulator must block any light leaving the transmitter between pulses; thus, B features higher losses and it is less energy-efficient than A.

Figure 2.6 shows the ideal transmission T_{EAM} of the modulator as a function of time, along with the target, identical optical power signals for both polarizations. Note the difference between transmission signal shapes of options A and B. In practice, a transmission of identically 0 is not possible, and bias optical power suppression is

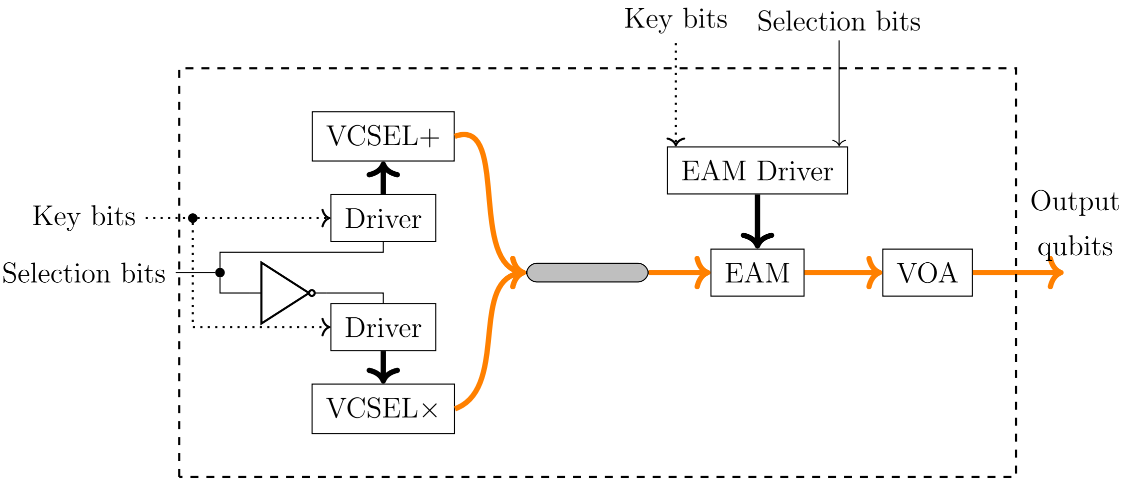

Figure 2.7: Complete functional diagram of the proposed transmitter structure. VCSEL+ and VCSEL× are the lasers emitting in the rectilinear and diagonal bases, respectively. Selection bits enable one laser driver, disabling the other, while key bits control the output light pulses. Both sequences are used to control the driver of the modulator, to ensure flexible pulse shaping. EAM: electro-absorption modulator; VOA: variable optical attenuator.

significantly more important for option B. The transmission is defined as

$$T_{\mathrm{EAM}}\left(t\right) = \frac{P_{\mathrm{in,EAM}}\left(t\right)}{P_{\mathrm{out,EAM}}\left(t\right)}, \tag{2.2}$$

where $P_{\mathrm{in,EAM}}$ and $P_{\mathrm{out,EAM}}$ are the modulator's input and output optical powers, respectively. The attenuation $a_{\mathrm{EAM}}^{\mathrm{dB}}$, measured in dBs, is related to the transmission as

$$a_{\mathrm{EAM}}^{\mathrm{dB}}\left(t\right) = -10 \cdot \log_{10}\left(T_{\mathrm{EAM}}\left(t\right)\right). \tag{2.3}$$

The driver of the modulator should be able to select the given voltage waveform based on the key and basis selection bits at its input. The desired change in attenuation needs to be properly translated into changes in voltage, as the relationship between the two is non-linear for EAMs.

Figure 2.7 shows the detailed functional diagram of the proposed transmitter: two lasers emitting in different bases driven by differentially enabled laser drivers, which provide the necessary current signals based on the key bits. The VCSELs' output is connected to optical fibers, which are then coupled using a symmetric coupler, leading into the modulator. The EAM is controlled by a driver, whereas a final variable optical attenuator is responsible for decreasing the average power so that the mean photon number of output qubits is around 0.1.

2.4.4 Spectral Attacks

In case BB84 (or a decoy state based on BB84) is implemented with polarization encoded single-photon qubits, it is vital to minimize information leakage through different degrees of freedom. Such possibilities include spatial, temporal and spectral differences between different states [99]. In the proposed design, mutual information between bits and spatial modes can be all but eliminated using a single-mode fiber section, while temporal differences are minimized by the modulator's pulse shaping function.

Spectral information leakage allows Eve to perform a *spectral attack*. A spectral attack consists of performing a non-destructive frequency measurement on the photon, which leaves its polarization properties intact, then sending it towards Bob. In the trivial design, spectral attacks can be avoided by using lasers with largely overlapping spectra. The leakage can be made arbitrarily small; one article reports an example, where the mutual information between frequency and key bit value is, on average, $6.6 \cdot 10^{-3}$ bits per state. However, due to the frequency split between eigenmodes of a VCSEL, the new design has an inherent weakness against spectral attacks. Since frequency and polarization are perfectly correlated, Eve can deduce the value of the qubit without introducing quantum bit errors, thus remaining undetected. The frequency split can be several tens of GHz [73], which is resolved within the spectral resolution of current technology.

Let us consider first a worst-case scenario, when all four states are spectrally distinguishable, with central frequencies $f_{\mathrm{RL}} < f_{\mathrm{RH}}$ and $f_{\mathrm{DL}} < f_{\mathrm{DH}}$. R and D denote the rectilinear and diagonal bases, while L and H stand for low and high As a first step, Eve gathers information about the frequencies and guesses the bijective function from the set of polarizations to the set of possible frequencies. There are $4! = 24$ such permutations. If she guesses wrong, Alice and Bob abort the protocol due to high QBER levels, and restart it. Eve is alerted about this via a public channel announcement of the legitimate parties, and changes her guess. The eavesdropper will eventually stumble upon the solution in at most 24 tries with a mean trial number of 12 (assuming that she makes her guesses at random), allowing her to gain full information without getting noticed.

The new design can only be of interest in a practical point of view, if the possibility of successful spectral attacks can be averted. One solution, which is not provably secure, but potentially useful in real-life situations with loosened assumptions about Eve's capabilities, is to use VCSELs with a frequency split that is small enough not to be resolved experimentally with current technology. Even so, the correlation between frequencies and bases should also be avoided—or else, Eve can first deduce the basis, then perform a polarization measurement in that basis. Therefore, the two VCSELs should also emit in the same band.

There also exists a method of protection that does not rely on the advances of technology, but requires two very similar VCSELs. Their respective low and high frequency eigenmode spectra should overlap, so that the following relationship stands between central frequencies:

$$f_{\mathrm{DL}} \approx f_{\mathrm{RL}} \tag{2.4}$$

$$f_{\mathrm{DH}} \approx f_{\mathrm{RH}}. \tag{2.5}$$

The split between low and high frequencies can be arbitrarily high. Now, the VCSELs should be oriented in such a way that eigenmodes with frequencies f_{DL} and f_{RH} correspond to a bit value 1 (or 0), whereas those with f_{DH} and f_{RL} correspond to 0 (or 1). Namely, lower and higher frequencies should carry opposite values of bits in case of the two lasers [14].

This way, the correlation between frequencies and both bits and bases can be decreased very close to zero, with some residual correlation due to the non-perfect spectrum overlaps. The proposed transmitter design can thus be prepared to be safe against potential spectral attacks as well.

2.5 Conclusion

I have described the general benefits of using VCSELs in low-power applications, focusing on DV-QKD protocols, most notably BB84. For this protocol, I also proposed a new transmitter design, which exploits polarization switching found in some surface-emitting lasers. Modulating the polarization on-demand, two VCSELs are enough to emit all four qubit states of BB84, leading to potential cost and size reduction.

I have also carefully investigated the problems and concerns regarding this new design. Two distinct current-driven polarization modulation options have been outlined, as well as the suggestion of electro-absorption modulators for proper attenuation and pulse shaping, to avoid leaking any information to eavesdroppers due to varying power levels or temporal differences. A potential spectral attack has also been discussed, when Eve uses the frequency split between polarization eigenmodes to distinguish between all states, along with defence methods against it. This chapter forms the basis of Thesis I.

Chapter 3

Quantum Random Number Generation Based on the Time Differences between Photon Detections

3.1 Introduction

The bulk of my research has been focused on quantum random number generation, a topic briefly introduced in Section 1.4. This and the following chapter are describing distinct, but very closely related QRNG architectures both theoretically and experimentally.

A certain group of optical quantum random number generators extracts randomness from the timing information of photon detections. Recall that following the convention in Ref. [38], these QRNGs may be called *time-of-arrival generators*. First of all, I am shortly introducing a selection of such devices.

Some ToA methods utilize quasi-single photon sources similar to those encountered in practical DV-QKD implementations. The generator in Ref. [100] consists of a heavily attenuated pulsed laser with mean photon numbers around 0.1, and the random bits are extracted from the time elapsed between pulses that triggered a detection event. Von Neumann extraction is applied to the bit sequence in order to reduce the significant bias. A proof-of-concept device was based on a similar principle, but for the purpose of bit generation, the laser pulses were grouped in a more sophisticated way [101]. Another article reports bit generation by detecting long coherence time photons, whose wave function overlaps many gating cycles of a SPAD [102]. The bits are assigned based on the parity of the clock cycle where the detection happened.

Other researchers directly sampled the elapsed times between photon detections from a light source (laser diode or LED) operating in CW mode. The exponential waiting-time distribution was measured with good precision in the generator of Ref. [103]. The non-uniformity was decreased by only keeping a certain number of LSBs and applying hash functions for data whitening, with the remaining bits fitting the min-entropy calculated in advance. The group of Wahl et al. used a high-resolution (1 ps) time-to-digital converter and digitized the time differences at 16 bits. Resilient functions helped remove the residual bias present at the MSBs [104]. The generator was later reported to have had been operating in stable conditions for 20 months continuously [105]. An interesting hybrid is also reported [106], where the time between photon detections is digitized and assigned symbols from a four-letter alphabet. Post-processing is then performed before converting letters to bits. CW light is also the photon source in Ref. [107], where bits are generated after setting up a coincidence window between two single-photon counting modules.

A prosperous idea is to use integrated designs, such as arrays of single-photon detectors for photon detection, and thus increase the bit generation rates using parallelization. In Ref. [108] detector pairs are chosen from an array, and arrival times are compared between each pair. A single-pixel generator is shown in Ref. [109], where successive detection times are measured with respect to a reference clock signal, then compared to each other. A potential prototype of a commercial QRNG is demonstrated in Ref. [110], where the SPAD array and a low-efficiency LED are integrated in a 3D chip packaging.

ToA QRNGs typically generate bits at rates in the order of 1–10 Mbps [102, 107, 109–112] or slightly higher, e.g. 40 Mbps in Ref. [103]. Some proof-of-concept devices are slower, only achieving sub-Mbps speed [100, 101, 110], but there also exist high-end devices, one of them reaching 152 Mbps [104], while that in Ref. [108] could potentially achieve 128 Mbps through parallelization of 256 SPAD pixels. These values are still lagging behind the capabilities of ASE generators, but with a significant benefit: the source of randomness is better understood.

In this chapter, a ToA generator is discussed. Building on an idea introduced in a previous publication, I created the mathematical model of the method, theoretically deriving its bit generation efficiency and rate, among other characteristics, as a function of input parameters.

3.2 Method of Random Number Generation

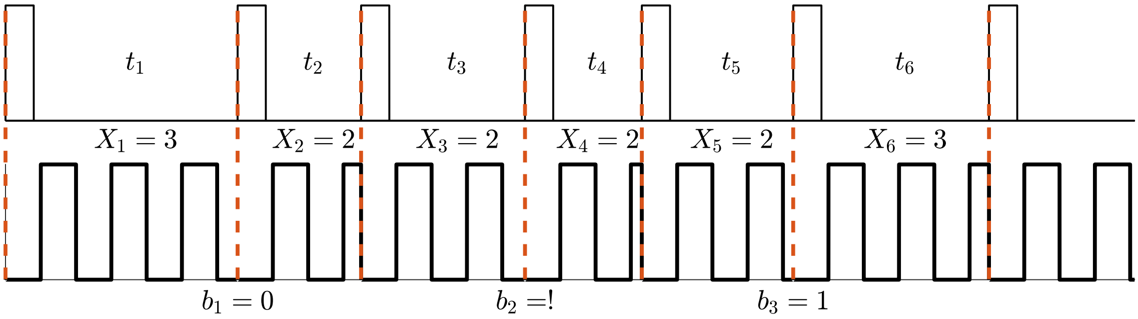

Figure 3.1: Timing diagram for the random number generation method. Top row: output pulses from the single photon detector, bottom row: restartable clock signal; red dashed lines denote the boundaries of intervals. X_j denotes the count of rising clock edges within the time interval described by the random variable $T_j = t_j$. "!" denotes a case of equality, where no bit is generated. Taken from Ref. [113].

The investigated method of random number generation was first implemented based on the detection of radioactive decay [41], but Mario Stipčević et al. replaced radioactive materials with a continuous-wave light source, emitting photons at an approximately constant rate [111]. Although they used an attenuated light emitting diode (LED), it can be swapped for a semiconductor laser. It is known that the arrivals of photons from an attenuated laser form a Poisson point process [26, 114] with a parameter λ describing the average number of photons per unit time. My model was composed with laser diodes in mind, but in pracical situations, it can also be applied to a LED-based setup (see Sect. 3.3.2).

A single-photon detector, either a single-photon avalanche detector (SPAD) or a photomultiplier tube (PMT) detects incoming photons. The time elapsed between two detections is then measured and registered on a signal processing unit. Two consecutive time measurements are compared to each other, and a bit is assigned to this pair based on whether the earlier or the later was longer. Every measurement is used in only one comparison, otherwise subsequent bits would become highly (anti)correlated. If intervals could be measured with infinite precision, the probability of equality in the comparison would vanish. However, in practice, only an imperfect, finite resolution measurement is possible. As an example, one may count the rising edges of a clock signal between detections as a discrete approximation of the analog interval length. This discretization causes the probability of equality to be greater than zero. Equalities must be discarded to keep the distribution uniform, limiting both the available bit generation efficiency and rate. The method is visualized in Fig. 3.1.

It has been proven in Ref. [111] that it is advantageous to restart the clock signal at each detection, rather than letting it run continuously (see Fig. 3.2). This is also eliminating correlations between subsequent time measurements, which ideally should be independent and uncorrelated. The authors operated the generator close to the *fast-clock limit*, where the average number of rising edges between detections is large, as this leads to the highest possible efficiency. I will show that this is, however, suboptimal in terms of the bit generation rate for any configuration of parameters.

This random number generation method has been influential, the generator reported in Ref. [112] uses the same principle completed by post-processing; however, the hardware is integrated, not directly assembled from discrete components.

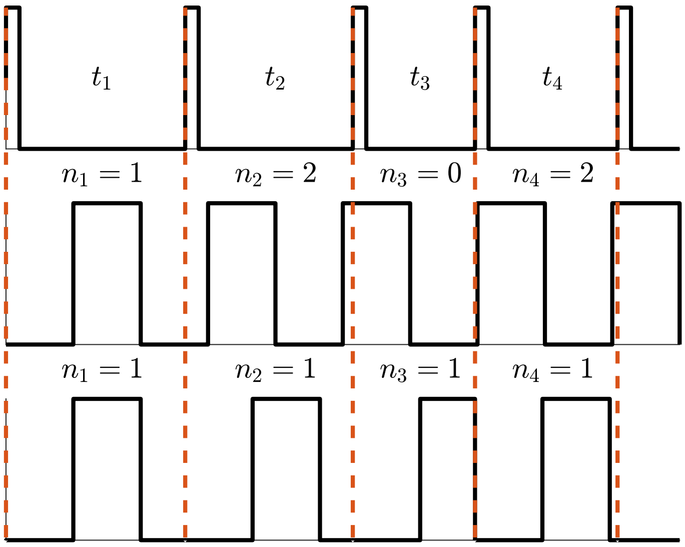

Figure 3.2: An example of time measurement of intervals (top row) with continuous/non-restartable (middle row) and restartable (bottom row) clock signals. t_j denotes the length of interval j, n_j is the number of rising edges counted within interval j. Note the significantly different values measured by the different clock signals; especially that $n_1 < n_2$ for a continuous clock, even though $t_1 > t_2$. Taken from Ref. [1].

3.3 Mathematical Model

A good model is preferably able to explain as many of the underlying phenomena as possible, all while remaining relatively simple and incorporating only a "few" parameters. What "few" means is also relative: the ultimate measure of quality is that the quantities or results predicted by the model shall be in a great agreement with the experimental findings. Table 3.1 lists all parameters important to the method, and whether they have been included in the final model. All noise parameters—$\lambda_{\mathrm{ambient}}$, λ_{dark} and p_{after}—are excluded from the model by setting them to zero. The validity of these choices will be explained in detail in Section 3.3, but as a general idea, their values can be close to zero or at least—if applicable—significantly smaller than λ with carefully chosen equipment. Thus, their effect on the random bit sequences can be reduced to almost negligible levels.

Finally, three parameters are left: the input photon rate λ, which is defined so that it already includes all propagation losses as well as the non-unit quantum efficiency of the detector; the clock signal's time period τ; and the dead time τ_{d} of the detection system, during which it is impossible to register any further events. Two basic types of dead time can be distinguished: the extendable (or paralyzable) case, where detections during a dead time increase its length, ultimately paralyzing the detector if the input count rate is too high, and the non-extendable (or non-paralyzable) case, where the detector is totally insensitive during a dead time and all incoming counts are rejected.

These are both idealizations: a real device might exhibit a behaviour in between these two extrema [115, 116]. In the current model, I take τ_d to be non-extendable, an assumption that will be explained while discussing the experiments.

As a further simplification, both λ and τ_d are taken to be constant. In reality, the photon rate is never truly constant, owing to e.g. thermal effects, while the dead time changes stochastically between a lower and a higher boundary—the constant in the model can be thought of as a mean value.

Table 3.1: List of possible relevant parameters of the mathematical model.

Parameter	**Description**	**Unit**	**Inclusion**
λ	Input photon rate of the detector	$\frac{1}{\mathrm{s}}$	✓
λ_ambient	Rate of ambient noise photons hitting the detector	$\frac{1}{\mathrm{s}}$	✗
λ_dark	Dark count rate of the detector	$\frac{1}{\mathrm{s}}$	✗
p_after	Afterpulsing probability of the detector	–	✗
τ	Period of the restartable clock signal	s	✓
τ_d	Dead time (non-paralyzable) of the detection system	s	✓

The random variables governing the bit generation scheme are the (analog) time intervals T_j and their discretized pairs X_j obtained from counting the rising clock edges within the j^th interval. Since λ is assumed to be constant, the underlying Poisson point process is homogeneous, therefore all variables $\{\,T_j\,\}_{j\in\mathbb{Z}^+}$ are identically distributed. The same holds for the random variables $\{\,X_j\,\}_{j\in\mathbb{Z}^+}$. Together with the independence of variables (see the explanation for the restartable clock), all T_j (and X_j) are independent and identically distributed (i.i.d.). This will be exploited in notation: if the index is unimportant, it is simply omitted, and only T or X is written.

If X_j are i.i.d., then the difference between two successive discretized time measurements $Y_i = X_{2i} - X_{2i-1}$—which is itself a discrete random variable—has a probability mass function (PMF) symmetric around zero. The bits are assigned to the variables based on the value of $W_i = \mathrm{sgn}\,(Y_i)$, where sgn() denotes the sign function. $P\,[W_i = 1] = P\,[W_i = -1]$ is a consequence of the symmetry of Y_i; therefore, the resulting bit sequence will follow a uniform distribution if the cases $W_i = 0$ are discarded.

I decided to focus on the calculation of two main quantities. The first is the bit generation efficiency η_R, defined as the average number of bits generated per random event; the second is the bit generation rate R, the average number of bits generated per unit time. The two are related by the formula

$$R = \frac{\eta_{\mathrm{R}}}{\mathbb{E}[T]}, \tag{3.1}$$

where $\mathbb{E}[T]$ is the expectation value of the random variable T; that is, the average duration of a random event.

For the given method, the efficiency is simply half the probability that two successive time intervals are not measured to be equal. The factor of $1/2$ comes from the fact that two random events are used for the generation of each bit.

$$\eta_{\mathrm{R}} = \frac{P\left[X_{2i} \neq X_{2i-1}\right]}{2} = \frac{1 - P\left[X_{2i} = X_{2i-1}\right]}{2} = \frac{1 - P_{\mathrm{eq}}}{2} \tag{3.2}$$

It is trivial to see from (3.2) that $0 \leq \eta_{\mathrm{R}} < 0.5$. Moreover, knowledge of the exact probability distribution governing the random variables T and X is enough to calculate both the bit generation efficiency and rate.

During the analysis, I start with excluding the dead time effects. This is a wildly idealistic approach, and I will refer to it as the ideal case; however, it provides an insight into the mathematical methods, and the results take simple forms. It also serves as a baseline, a basis of comparison after the results are also obtained with the dead time included.

3.3.1 The Ideal Case of Zero Dead Time

If there is no dead time, all detections must be modelled by zero-width pulses, so that no matter how close they are to each other in time, their exact number can be resolved. The time intervals between dead time free photon detections from a Poisson point process follow an exponential distribution with parameter λ. The probability density function (PDF) f_T and cumulative distribution function (CDF) F_T are given in Eqs. 3.3 and 3.4.

$$f_T(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{3.3}$$

$$F_T(t) = \int_{-\infty}^{t} f_T\left(t'\right)\, \mathrm{d}t' = \begin{cases} 1 - \mathrm{e}^{-\lambda t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{3.4}$$

The PMF of the discrete variable X, $P\left[X=n\right]$, can be calculated using the definition of cumulative distribution functions:

$$\begin{aligned}
P\left[X=n\right] &= P\Big[n\tau \leq T < (n+1)\tau\Big] && (3.5)\\
&= F_T\Big[(n+1)\,\tau\Big] - F_T\left(n\tau\right) && (3.6)\\
&= \left(1-\mathrm{e}^{-\lambda(n+1)\tau}\right) - \left(1-\mathrm{e}^{-\lambda n\tau}\right) && (3.7)\\
&= \mathrm{e}^{-n\lambda\tau} - \mathrm{e}^{-(n+1)\lambda\tau} && (3.8)\\
&= \mathrm{e}^{-n\lambda\tau}\left(1-\mathrm{e}^{-\lambda\tau}\right), && (3.9)
\end{aligned}$$

where $n \in \mathbb{N}$; the PMF is zero for any $n<0$. Using the substitution $p = 1-\mathrm{e}^{-\lambda\tau}$ one finds that this PMF has the form of $P\left[X=n\right] = (1-p)^n p$, which is a failure counting geometric distribution with parameter p. This is an interesting observation, since the geometric distribution is the discrete equivalent of an exponential distribution, in the sense that these are the only probability distributions holding the memoryless property [117].

From the PMF, P_{eq} may be calculated easily using the fact that subsequent (and all other) measurement results are independent and identically distributed (i.i.d.). Remember that independency arises as a consequence of a restartable clock signal.

$$\begin{aligned}
P_{\mathrm{eq}} = P\left[X_{2i}=X_{2i-1}\right] &= \sum_{n=0}^{\infty} P\left[X_{2i}=n,\ X_{2i-1}=n\right] && (3.10)\\
&= \sum_{n=0}^{\infty} P\left[X=n\right]^2 && (3.11)\\
&= \left(1-\mathrm{e}^{-\lambda\tau}\right)^2 \sum_{n=0}^{\infty} \mathrm{e}^{-2n\lambda\tau} && (3.12)\\
&= \left(1-\mathrm{e}^{-\lambda\tau}\right)^2 \cdot \frac{1}{1-\mathrm{e}^{-2\lambda\tau}} && (3.13)\\
&= \frac{\mathrm{e}^{2\lambda\tau}-2\mathrm{e}^{\lambda\tau}+1}{\mathrm{e}^{2\lambda\tau}-1} && (3.14)\\
&= \frac{\mathrm{e}^{\lambda\tau}-1}{\mathrm{e}^{\lambda\tau}+1} && (3.15)
\end{aligned}$$

The equality between Eqs. 3.10 and 3.11 is the consequence of the i.i.d. relationship between X_{2i} and X_{2i-1}. Moreover, the ordinary summation formula of a geometric series can be used between Eqs. 3.12 and 3.13, exploiting that $\lambda\tau$ is strictly positive, thus $\left|\mathrm{e}^{-2\lambda\tau}\right| < 1\ \forall\lambda,\ \forall\tau$; therefore, the series always converges.

The bit generation efficiency takes the remarkably simple form

$$\eta_{\mathrm{R}}\left(\lambda,\tau\right) = \frac{1-P_{\mathrm{eq}}}{2} = \frac{\mathrm{e}^{\lambda\tau}-1}{\mathrm{e}^{2\lambda\tau}-1} = \frac{1}{\mathrm{e}^{\lambda\tau}+1}, \qquad (3.16)$$

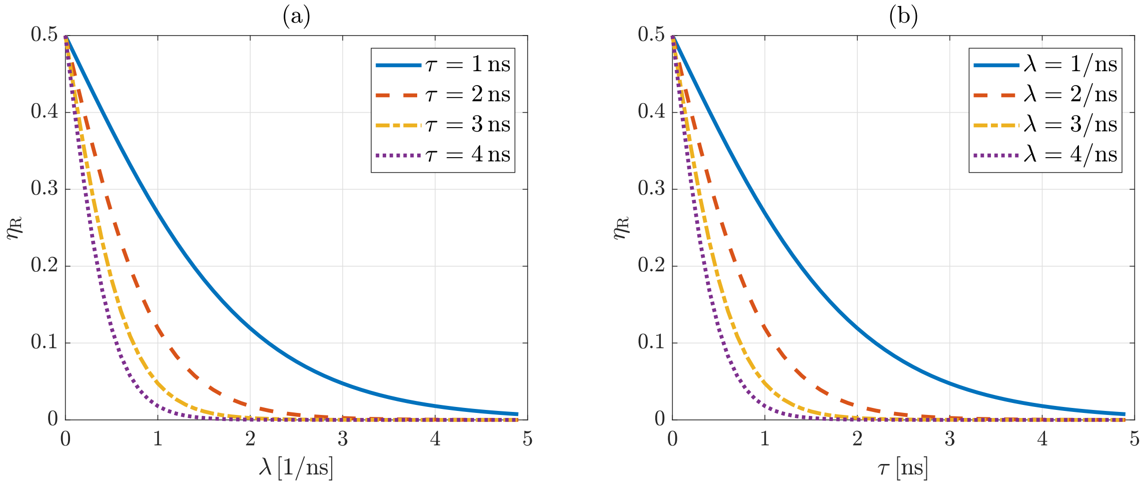

Figure 3.3: Bit generation efficiency $\eta_{\rm R}$ in the ideal case as a function of (a) λ, for constant values of τ; (b) τ, for constant values of λ. The function is symmetric in its arguments, and monotonically decays to zero as either of them is increased.

which is symmetric in the variables λ and τ; specifically, it only depends on their product. The efficiency approaches its maximum of 0.5 as $\lambda\tau \to 0$—this is the so called fast-clock limit—, and monotonically decays to zero as the product goes to infinity (slow-clock limit). Figure 3.3 shows the efficiency as a function of purely λ and τ, for different constant values of τ and λ, respectively, while Fig. 3.5(a) depicts the two-dimensional surface plot along both variables simultaneously.

The bit generation rate R also involves the expectation value of time differences between detections, which turns out to be the reciprocal of the input photon rate:

$$\mathbb{E}\left[T\right] = \int_{-\infty}^{\infty} t \cdot f_T(t)\ \mathrm{d}t = \int_{0}^{\infty} \lambda t \cdot \mathrm{e}^{-\lambda t}\ \mathrm{d}t = \frac{1}{\lambda}, \tag{3.17}$$

meaning that R is not symmetric in its variables, but has somewhat more interesting properties:

$$R\left(\lambda, \tau\right) = \frac{\eta_{\rm R}\left(\lambda, \tau\right)}{\mathbb{E}\left[T\right]} = \frac{\lambda}{\mathrm{e}^{\lambda\tau} + 1}. \tag{3.18}$$

Since $\mathbb{E}\left[T\right]$ is independent of the clock period, for a constant λ, R is just a monotonically decreasing function of τ, approaching its maximum of $\lambda/2$ as $\tau \to 0$. This is logical, since with ever increasing time measurement precision, the probability of equality decreases to zero, no comparisons are thrown away, and asymptotically each two detections result in the creation of one bit.

Now, in a practical approach, it is more sensible to imagine a fixed clock period and the ability to change the light source's intensity—thus, the input photon rate—in a continuous fashion. The qualitative behaviour of $R\left(\lambda, \tau = \tau_0\right)$ is quite simple to reveal without looking at the function itself. As $\lambda \to 0$, the relative measurement precision

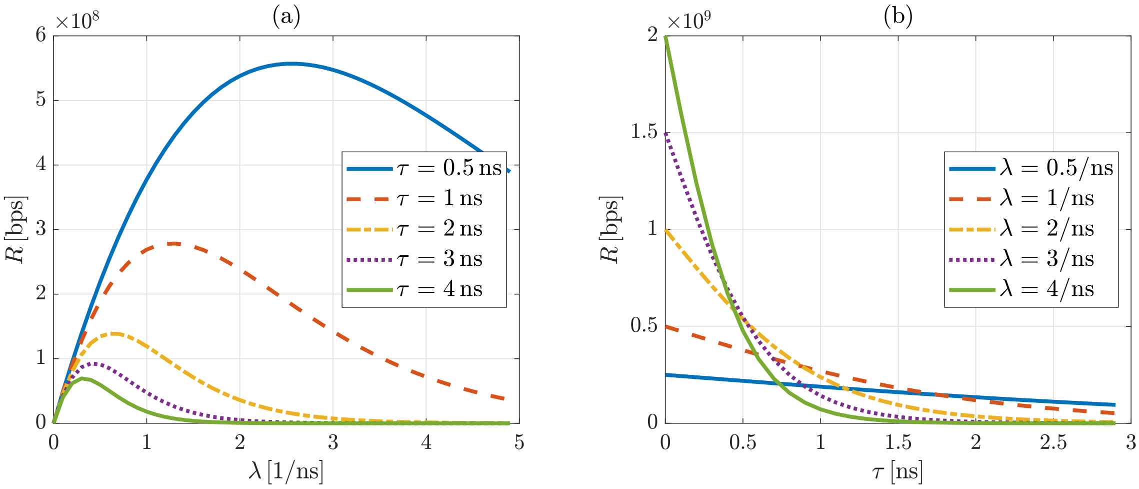

Figure 3.4: Bit generation rate R in the ideal case as a function of (a) λ, for constant values of τ; (b) τ, for constant values of λ. Taken from Ref. [1].

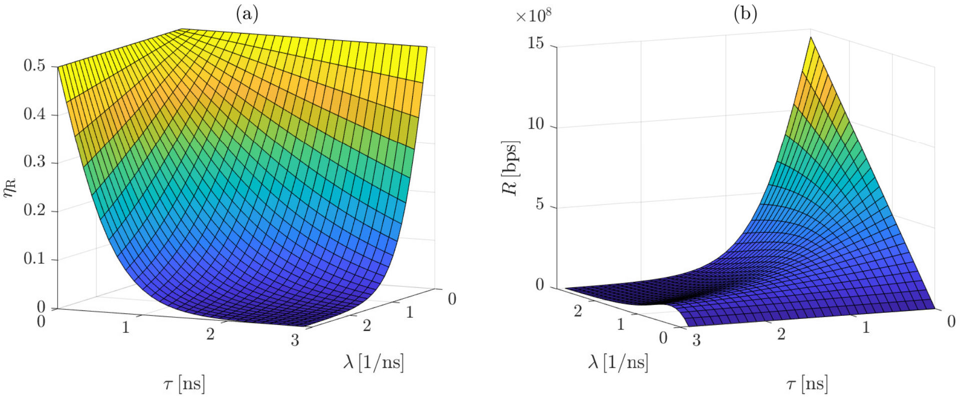

Figure 3.5: (a) Bit generation efficiency η_{R} and (b) bit generation rate R in the ideal case as a function of both λ and τ.

and the efficiency increases, but there are fewer and fewer detections; therefore, R decays to zero. On the other hand, if $\lambda \to \infty$, almost every interval is measured to be equal to all the others ($X_j = 0$ almost $\forall j$), and as the efficiency drops to zero, so does the generation rate. In between, it is logical to assume that there is a certain value of λ corresponding to a maximal bit generation rate, since R is non-negative and not identically zero. Figure 3.4 shows that the intuition behind this reasoning is right. The 3 dimensional surface plot is shown in Fig. 3.5(b).

The optimal photon rate that results in the maximum bit generation rate for a given clock period,

$$\lambda_{\mathrm{opt}}(\tau) = \arg\max_{\lambda} R(\lambda, \tau), \tag{3.19}$$

can be calculated by setting the partial derivative $\partial R(\lambda, \tau) / \partial \lambda$ to zero and solving it for λ:

$$\frac{\partial R(\lambda, \tau)}{\partial \lambda} = \frac{\mathrm{e}^{\lambda\tau}(1-\lambda\tau)+1}{\left(\mathrm{e}^{\lambda\tau}+1\right)^2} = 0 \tag{3.20}$$

$$\Longrightarrow \mathrm{e}^{\lambda_{\mathrm{opt}}\tau}\left(1-\lambda_{\mathrm{opt}}\tau\right)+1 = 0 \tag{3.21}$$

$$\Longrightarrow \lambda_{\mathrm{opt}}(\tau) = \frac{W_0\left(\mathrm{e}^{-1}\right)+1}{\tau} \approx \frac{1.27846}{\tau}. \tag{3.22}$$

Here $W_0(x)$ is the principal branch of the Lambert W function. The optimal photon rate is thus inversely proportional to the clock period. There are two interesting consequences regarding the efficiency. First, in order to maximize the bit generation rate, some of the efficiency needs to be sacrificed: it will be significantly less than 0.5. Second, in the ideal, dead time free case, the efficiency corresponding to λ_{opt} and $R_{\max}$ is a constant value, independent from τ:

$$\eta_{\mathrm{R,opt}} = \eta_{\mathrm{R}}\left(\lambda_{\mathrm{opt}}, \tau\right) = \frac{W_0\left(\mathrm{e}^{-1}\right)}{W_0\left(\mathrm{e}^{-1}\right)+1} \approx 0.217812. \tag{3.23}$$

This means that on average, a little less than five time intervals between detections are needed to generate one bit, but surprisingly, this leads to the most bits generated per unit time.

Similarly to λ_{opt}, the maximal bit generation rate $R_{\max}$ is also proportional to the reciprocal clock period (Eq. 3.24), meaning that with better time measurement resolution (smaller τ), it becomes possible to generate more and more bits per unit time, given that our single-photon detector is safe to operate under the given intensity value.

$$R_{\max}(\tau) = \max_{\lambda} R(\lambda, \tau) = R\left(\lambda_{\mathrm{opt}}, \tau\right) = \frac{W_0\left(\mathrm{e}^{-1}\right)}{\tau} \approx \frac{0.27846}{\tau} \tag{3.24}$$

3.3.2 The Practical Case of Non-Zero Dead Time

To make the model more realistic, a dead time τ_{d} must be included, which is a property of the single-photon detector and the time-tagging electronics. As mentioned in Sect. 3.3, τ_{d} is taken to be non-extendable and constant. By these assumptions, the random variable T can never be shorter than τ_{d}, and it follows a shifted exponential distribution [104], its PDF and CDF taking the following forms:

$$f_T(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda(t-\tau_{\mathrm{d}})}, & t \geq \tau_{\mathrm{d}} \\ 0, & t < \tau_{\mathrm{d}}; \end{cases} \tag{3.25}$$

$$F_T(t) = \begin{cases} 1-\mathrm{e}^{-\lambda(t-\tau_{\mathrm{d}})}, & t \geq \tau_{\mathrm{d}} \\ 0, & t < \tau_{\mathrm{d}}. \end{cases} \tag{3.26}$$

The average time between detections is increased by exactly the length of dead time compared to the ideal case.

$$\mathbb{E}[T] = \int_{-\infty}^{\infty} t \cdot f_T(t)\ \mathrm{d}t = \int_{\tau_\mathrm{d}}^{\infty} \lambda t \cdot \mathrm{e}^{-\lambda(t-\tau_\mathrm{d})}\ \mathrm{d}t = \tau_\mathrm{d} + \frac{1}{\lambda} = \frac{1+\lambda\tau_\mathrm{d}}{\lambda} \tag{3.27}$$

As we have seen in the ideal case, the expectation value was the inverse of the count rate. Without dead time, the input and output count rates are the same, since the effects of quantum efficiency have already been accounted for in λ. In the non-ideal case it is therefore reasonable to think of $\mathbb{E}[T]$ as the reciprocal of the output count rate [116]

$$\lambda_\mathrm{out} = \frac{\lambda}{1+\lambda\tau_\mathrm{d}}, \tag{3.28}$$

which is less than or equal to λ for every combination of non-negative λ and τ_d, and has the limit of $1/\tau_\mathrm{d}$ as $\lambda \to \infty$.

The presence of $\tau_\mathrm{d} \neq 0$ is one of the reasons why LEDs can be substituted for laser diodes as well: the dead time erases the photon bunching, which is a characteristic of thermal light sources [103]. Additionally, the strong attenuation can be thought of as high-probability random (Bernoulli) deletion of photons, which has been shown to bring the photon number distributions close to Poissonian [118].

The dead time can be rewritten to the form

$$\tau_\mathrm{d} = k\tau + \Delta\tau = \left(k + \frac{\Delta\tau}{\tau}\right)\tau, \tag{3.29}$$

where $k = \lfloor \tau_\mathrm{d}/\tau \rfloor$ is a non-negative integer and $0 \leq \Delta\tau < \tau$. The ratio $\Delta\tau/\tau$ is referred to as the "fractional dead time" in the following. At first, this form might seem arbitrary, but it greatly simplifies the analysis. If the fractional dead time is fixed, changing k only shifts the PMF $P[X=n]$ to $P[X=n+k]$: measuring a time interval to be shorter than k times the clock period is impossible. It can be quickly realized that both P_eq and (consequently) η_R are shift-invariant, since the infinite summation has the same terms with k zeros padded to the front. However, the bit generation rate varies with k as well; Eq. 3.27 shows that the expectation value of a time interval's length depends on the entire value of the dead time.

Although it is possible to perform a unified analysis on how the distribution, η_R and R change as k and $\Delta\tau/\tau$ are altered, it is interesting to treat the case of zero fractional dead time ($\Delta\tau = 0$) separately.

Zero fractional dead time. The reason behind this separation is that due to the shift-invariance discussed above, $\Delta\tau = 0$ for any k results in a shifted geometric

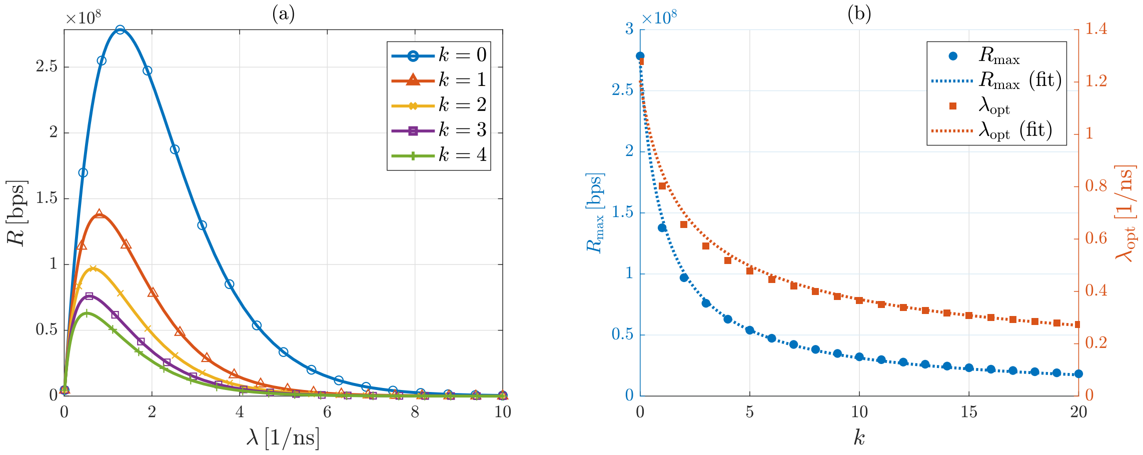

Figure 3.6: Effects of increasing k for $\tau = 1\,\text{ns}$ and zero fractional dead time. (a) Bit generation rate R as a function of λ; (b) peak bit generation rate R_{max} and optimal input photon rate λ_{opt} as a function of k along with best-fit model lines. Increasing k decreases both parameters.

distribution of X (Eq. 3.30) and a bit generation efficiency unchanged from the ideal situation (Eq. 3.23).

$$P\left[X = n\right] = \begin{cases} 0, & n < k \\ \mathrm{e}^{-n\lambda\tau}\left(1 - \mathrm{e}^{-\lambda\tau}\right), & n \geq k \end{cases} \tag{3.30}$$

The bit generation rate, on the other hand, is affected negatively by a higher k value, since the average length of random events is increased:

$$R\left(\lambda, \tau, \tau_{\text{d}} = k\tau\right) = \frac{\lambda}{\left(1 + k\lambda\tau\right)\left(\mathrm{e}^{\lambda\tau} + 1\right)}. \tag{3.31}$$

However, in a qualitative aspect, for a constant τ (and τ_{d}), R is still a simple, single-peaked function that vanishes as $\lambda \to 0$ or $\lambda \to \infty$. Its maximum value is decreased compared to the ideal case, while the maximum's location λ_{opt} is shifted to the left. Its exact value can only be calculated numerically, since there is no analytical solution for λ to the equation $\mathrm{e}^{\lambda\tau}\left(\lambda^2\tau\tau_{\text{d}} + \lambda\tau - 1\right) = \mathrm{e}^{\lambda\tau}\left(\lambda^2 k\tau^2 + \lambda\tau - 1\right) = 1$. Figure 3.6(a) shows several bit generation rate curves for $\tau = 1\,\text{ns}$ as a function of τ with different k values. Subfigure (b) depicts how λ_{opt} and R_{max} decrease as k increases, along with best fit curves of the form $a \cdot x^b + c$. Curve fitting was done using all samples from $k = 0$ to 499, yielding the approximations

$$\lambda_{\text{opt,fit}} = 1.195 \cdot (k+1)^{-0.5018} + 0.01102, \text{ and} \tag{3.32}$$

$$R_{\text{max,fit}} = 2.737 \cdot 10^8 \cdot (k+1)^{-0.9069} + 8.38 \cdot 10^4. \tag{3.33}$$

The respective R-squared values are 0.9966 and 0.9989. These can be briefly summarized as $\lambda_{\text{opt}} \underset{\sim}{\propto} (k+1)^{-0.5}$ and $R_{\text{max}} \underset{\sim}{\propto} (k+1)^{-0.9}$, where $\underset{\sim}{\propto}$ denotes approximate proportionality.

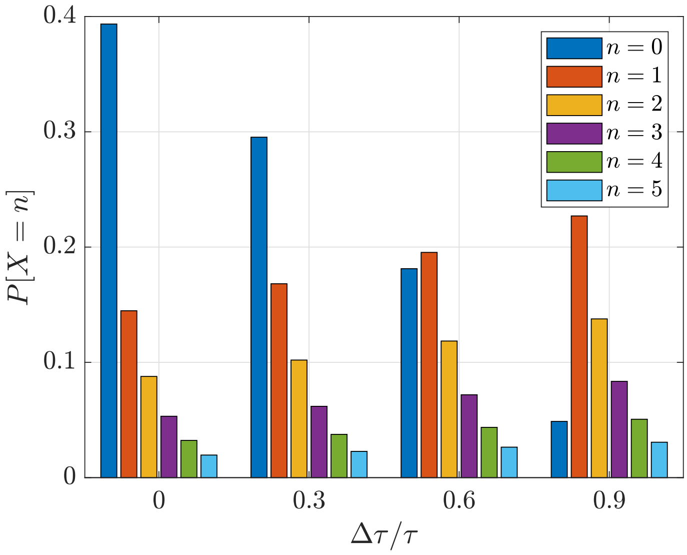

Figure 3.7: The dead time distorted geometric distribution of X for $k=0$, $\lambda=0.5$, $\tau=1$ (in arbitrary units), and different values of $\Delta\tau/\tau$. Probabilities for $n>5$ are not shown. Taken from Ref. [1].

Non-zero fractional dead time. For any non-zero value of $\Delta\tau$, X follows a *dead time distorted geometric distribution*, for which the PMF (Eq. 3.34) can be obtained from the CDF in Eq. 3.26:

$$P\left[X=n\right]=\begin{cases}0, & n<k\\ 1-\mathrm{e}^{-\lambda(\tau-\Delta\tau)}, & n=k\\ \mathrm{e}^{-\lambda\left[(n-k)\tau-\Delta\tau\right]}\left(1-\mathrm{e}^{-\lambda\tau}\right), & n>k.\end{cases} \tag{3.34}$$

This distribution, once again, only depends on k as a shifting factor, but it is heavily altered by changing the fractional dead time. For a given λ and τ, $P\left[X=k\right]$ is gradually decreasing with increasing $\Delta\tau/\tau$, and the probability is redistributed among the other outcomes (see Fig. 3.7 for an example). At a certain value, depending on all the aforementioned factors, the outcome $X=k+1$ becomes the most probable. From the distribution, the formula for the probability of equality P_{eq} can be calculated using the same methods as in the ideal case, although the non-zero fractional dead time made it significantly complicated:

$$P_{\mathrm{eq}}=\underbrace{\left[1-\mathrm{e}^{-\lambda(\tau-\Delta\tau)}\right]^2}_{P_{\mathrm{eq}}^0}+\underbrace{\mathrm{e}^{2\lambda\Delta\tau}\frac{\left(1-\mathrm{e}^{-\lambda\tau}\right)^2}{\mathrm{e}^{2\lambda\tau}-1}}_{P_{\mathrm{eq}}^+}. \tag{3.35}$$

P_{eq}^0 is the probability that both X_{2i} and X_{2i-1} have the value k—the least possible for a dead time between $k\tau$ and $(k+1)\,\tau$—, while P_{eq}^+ is the sum of the contributions of all other possible equalities. The formulae for the bit generation efficiency and rate follow from simply substituting P_{eq} and $\mathbb{E}\left[T\right]$ into Eqs. 3.2 and 3.1, respectively.

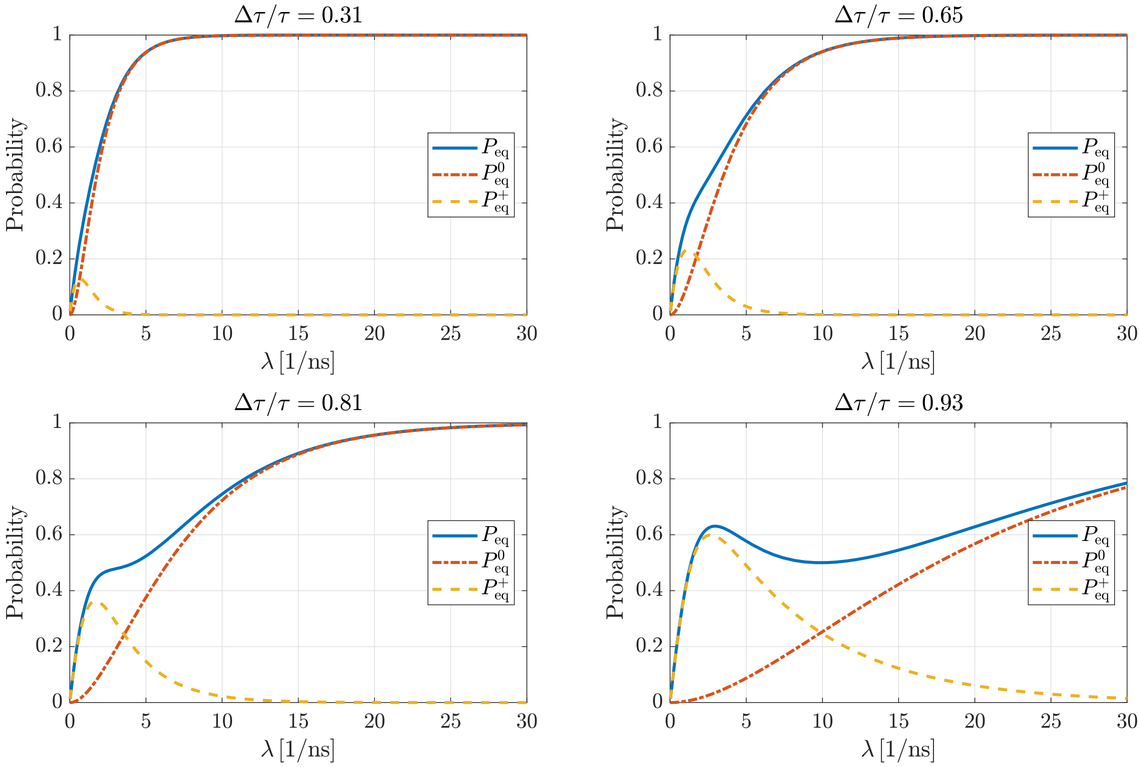

Figure 3.8: The contributions of P^0_{eq} and P^+_{eq} to the total probability of equality P_{eq} as a function of λ, for $\tau = 1\,\mathrm{ns}$ and four different values of the fractional dead time.

$$\eta_{\mathrm{R}}\left(\lambda,\tau,\tau_{\mathrm{d}} = k\tau + \Delta\tau\right) = \frac{1}{2}\cdot\left\{1 - \left[1 - \mathrm{e}^{-\lambda(\tau-\Delta\tau)}\right]^2 - \frac{\mathrm{e}^{2\lambda\Delta\tau}\left(1-\mathrm{e}^{-\lambda\tau}\right)^2}{\mathrm{e}^{2\lambda\tau}-1}\right\} \tag{3.36}$$

$$R\left(\lambda,\tau,\tau_{\mathrm{d}} = k\tau + \Delta\tau\right) = \frac{1}{2}\cdot\frac{\lambda}{1+\lambda\tau_{\mathrm{d}}}\cdot\left\{1 - \left[1 - \mathrm{e}^{-\lambda(\tau-\Delta\tau)}\right]^2 - \frac{\mathrm{e}^{2\lambda\Delta\tau}\left(1-\mathrm{e}^{-\lambda\tau}\right)^2}{\mathrm{e}^{2\lambda\tau}-1}\right\} \tag{3.37}$$

Once again, both P_{eq} and the efficiency are independent of k. They have essentially only two parameters: in one formulation, the products $\lambda\tau$ and $\lambda\Delta\tau$; in another, slightly more intuitive one, the product $\lambda\tau$ (the average number of incoming photons within a clock period), and the ratio $\Delta\tau/\tau$ (the fractional dead time). However, the bit generation rate is a function of the entire length of τ_{d}, and it depends on four parameters: λ, τ, k and $\Delta\tau$ (or $\Delta\tau/\tau$).

Qualitatively, R behaves differently from the previous cases if we fix the last three parameters and only look at it in terms of λ (see Fig. 3.11 in the next section for examples). If the fractional dead time is zero, the bit generation rate takes the form of a single-peaked function that vanishes as $\lambda \to 0$ or $\lambda \to \infty$, regardless of the other

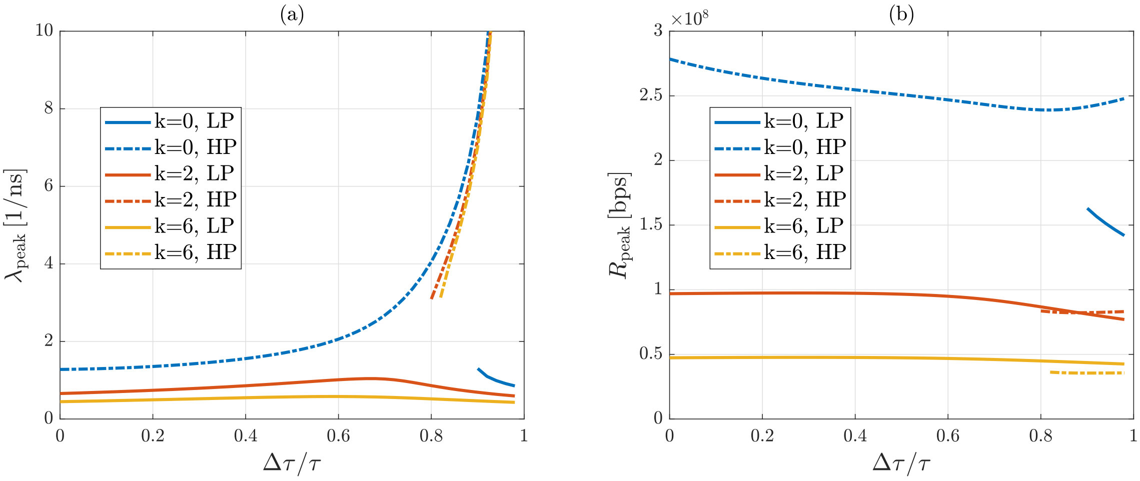

Figure 3.9: (a) λ_{peak} and (b) R_{peak} for $\tau = 1\,\text{ns}$ as a function of $\Delta\tau/\tau$ for $k = 0$, 2 and 6. Same colors denote the same k; continuous lines correspond to the lower-located peak (LP), dash-dotted lines correspond to the higher-located peak (HP).

parameters. First, it should be noted that as $\Delta\tau/\tau$ is increased for a certain $\lambda\tau$, the relative contributions of P_{eq}^{0} and P_{eq}^{+} to the total probability of equality change significantly. For a fixed fractional dead time, the first is always single-peaked, while the latter is monotonically increasing from 0 to 1 as a function of λ (or $\lambda\tau$). Figure 3.8 shows the plots along with their sum for four increasing values of $\Delta\tau/\tau$.

Correspondingly, the bit generation rate has distinct features as well. For a given τ and k, if we start to increase the fractional dead time from $\Delta\tau = 0$, the region around the *primary peak*—the one that occurs for all parameter combinations—flattens out, and eventually, beyond an appearance point $\left(\Delta\tau/\tau\right)_{\text{app}}$ a *secondary peak* emerges. Additionally, let the phrase *lower-located peak* (LP) denote the peak that appears at a lower λ, whereas *higher-located peak* (HP) denote the one appearing at a higher photon rate. For $k = 0$ and 1, HP is the primary, and it has a higher associated R_{peak} value—thus, it is a global maximum. For $k \geq 2$, LP is the primary, and usually it remains the global maximum. The only exception comes for $k = 2$, where the relative magnitude of the primary and secondary peaks vary with the fractional dead time. (Note that the definition of the primary peak is based on the fact that its location has a continuous curve as a function of the fractional dead time.) Figure 3.9 shows how the peak locations λ_{peak} and the peak generation rates R_{peak} (the local maxima) change for different values of $\Delta\tau/\tau$ in three distinct cases ($k = 0$, 2 and 6). Figure 3.10 depicts how the appearance point $\left(\Delta\tau/\tau\right)_{\text{app}}$ changes for different k values.

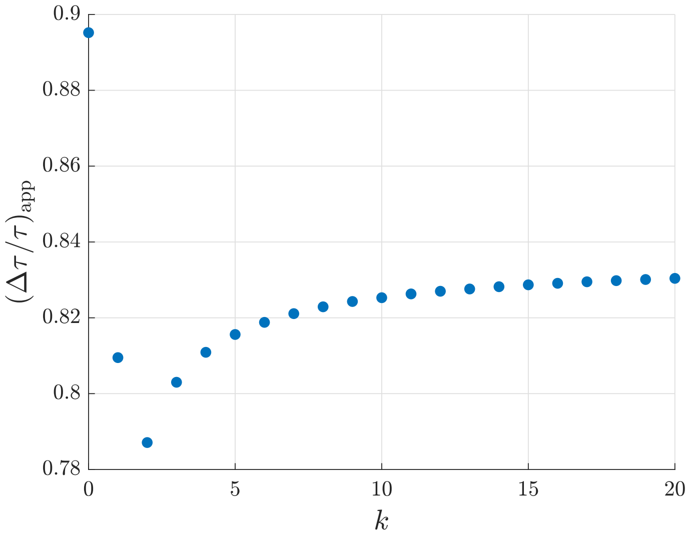

Figure 3.10: The appearance point $(\Delta\tau/\tau)_{\text{app}}$ of the second peak as a function of k, $\tau = 1\,\text{ns}$.

3.3.3 Simulation and Prior Evaluation

The validity and applicability of the model needs to be assessed carefully, since it is based on several approximations and simplifying assumptions. Before going on to my own experimental results in Sec. 3.4, I present some tools of prior evaluation; namely, comparison to the data found in the original paper describing the model [111], and simulation results obtained by MATLAB, which both address the correctness of the R-curves and the "true" randomness of the generated bits. MATLAB has a built-in method to generate pseudorandom numbers that are exponentially distributed. Although these values are deterministic, the PRNG has a high enough quality (and a long enough period) that the sequences are expected to pass the statistical tests. The tests are unable to detect determinacy; therefore, it can only reveal possible problems that are inherent to the bit generation method (not the PRNG), provided we do not surpass the period length.

Comparison to existing data. In Ref. [111], Stipčević and Rogina reported to have achieved a bit generation efficiency $\eta_{\mathrm{R}} = 0.487 \pm 0.02$ and a bit generation rate $R \approx 1\,\text{Mbps}$. The underlying parameters were $\lambda \approx 2\,/\mu\text{s}$, $\tau = 1/48\,\mu\text{s}$ and $\tau_{\mathrm{d}} = 25\,\text{ns}$. (The latter corresponds to $k = 1$ and $\Delta\tau/\tau = 0.2$ in my formalism.) Substituting these into the derived theoretical functions yield the results $\eta_{\mathrm{R}} \approx 0.4897$ and $R \approx 0.9328\,\text{Mbps}$. These values suggest that the model is, indeed, a good description of the physical device; the differences could be attributed to the fact that λ is only given approximately and/or statistical fluctuations which had not yet averaged out by the time the experimental results were recorded.

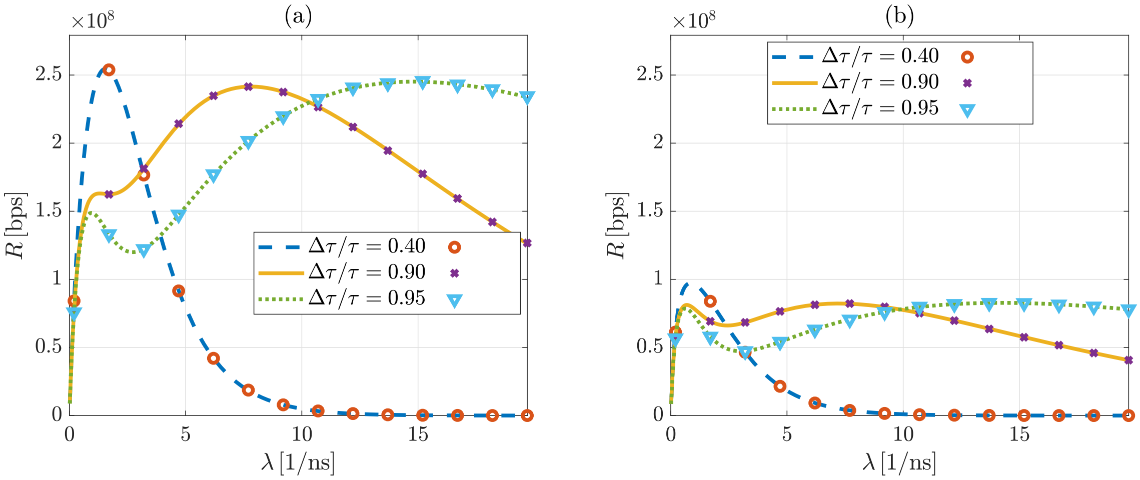

Figure 3.11: Bit generation rate R as a function of λ for $\tau = 1\,\mathrm{ns}$ and different values of $\Delta\tau/\tau$. (a) $k = 0$; (b) $k = 2$. Lines denote theoretical values, markers denote simulation results calculated from 10^6 time intervals. Taken from Ref. [1].

Simulations obtaining bit generation rate curves. To see whether the mathematical results are correct, we fixed the values of τ and τ_d and simulated time intervals following a shifted exponential distribution for different input photon rates. The simulated values are very much in agreement with the predictions for as few as 10^3 generated intervals for each λ; however, if we increase the number of intervals to 10^6, the conformity between the two is almost perfect. Figure 3.11 shows several different examples: single-peaked cases ($\tau_\mathrm{d}/\tau = 0.4$ and 2.4), a function where the secondary peak has barely emerged ($\tau_\mathrm{d}/\tau = 0.9$), and situations with two prominent peaks ($\tau_\mathrm{d}/\tau = 0.95$, 2.9 and 2.95). The simulations even show continuity for very high λ, where MATLAB could not plot the theoretical curves anymore due to the limited magnitude of arguments its built-in exponential function can take.

Randomness testing. Randomness testing was conducted using the Statistical Test Suite of the National Institute of Standards and Technology (NIST STS) [65]. In all simulations, τ and τ_d were kept constant at 1 and 0.4 ns, respectively. At first, the most basic uniformity criteria was analysed: the relative frequency of ones/zeros, that should be 0.5 ideally. Sequences were generated for 20 different values of λ from 10^6 time intervals, all passing the NIST Monobit test, proving that they are unbiased. Figure 3.12(a) shows one particular result for the relative frequencies with the bounds of acceptability. It can be seen that the bounds diverge as the input photon rate increases: this is due to the fact that the bit generation efficiency decreases, the bit sequences obtained from the same number of intervals are shorter and shorter, and the requirements for accepting the null hypothesis become looser. Figure 3.12(b), on the other hand, depicts the ratio of accepted sequences based on 1000 simulations

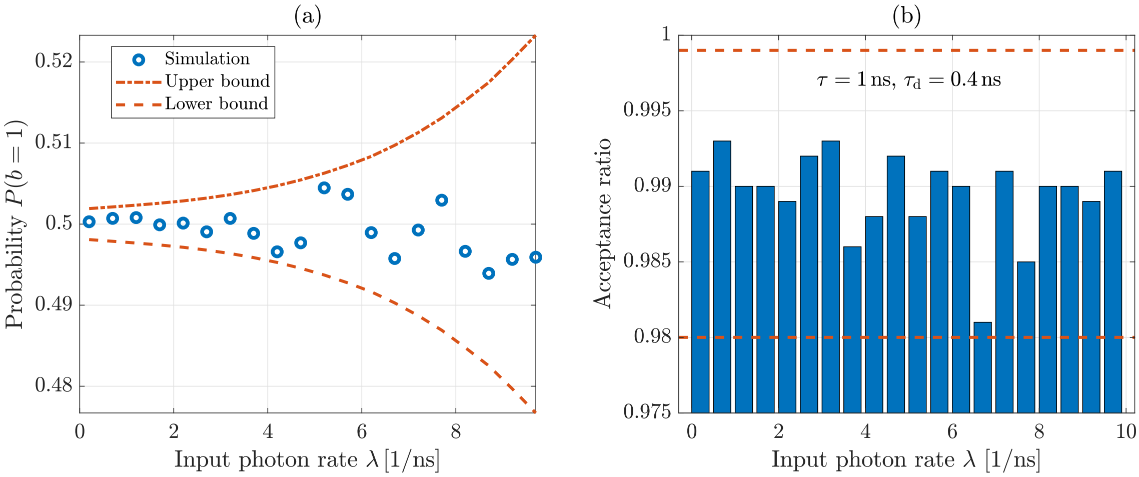

Figure 3.12: Bias testing of simulated bit sequences for $\tau = 1\,\mathrm{ns}$ and $\tau_{\mathrm{d}} = 0.4\,\mathrm{ns}$ as a function of λ. (a) Proportion (relative frequency) of ones during the simulation of 10^6 intervals. Red lines represent the bounds between which the NIST STS accepts the sequence as random. (b) Ratio of successful sequences out of 1000 in total (for each λ separately), with sequences being generated using 10^9 intervals. Red lines show the bounds of acceptability for the ratio test at a significance level of $\alpha = 0.01$. Taken from Ref. [113].

with each λ. Irrespectively of the input photon rate, the method was proven to produce bias-free bit sequences.

A 10^9 long sequence was also generated at the specific value of $\lambda = 0.4\,\mathrm{ns}$. In my experience, this is the shortest length sequence, which is worthwhile to be subjected to the entire NIST test suite, based on the different minimum recommendations of the individual tests. The bit sequence passed the whole test battery at a significance level of $\alpha = 0.01$, showing no significant deviations from what is expected from a perfectly random RNG. As a last step, the normalized autocorrelation function (ACF) of a 10^8 long subsequence was also calculated. All coefficients a_i corresponding to a time lag i (apart from a_0, which is 1 by definition) have small enough absolute values. Notably, the correlation between successive bits is small ($a_1 = -2.75 \cdot 10^{-4}$), further supporting the claim that the method provides a way to generate uniformly distributed bits, even without post-processing.

3.3.4 Information-Theoretical Considerations

Before heading onto the experimental realization of the random number generator, let us take a short digression towards the further possibilities the current time-measurement method can offer. It is possible to use the (dead time distorted) geometric random variables with other bit generation methods. These may provide a higher efficiency and bit generation rate, but might be less robust than the current solution. To analyze the amount of available randomness, we should reach out to the notions of information-theoretical entropies.

If a variable Y has possible outcomes $\{ y_n \}$—where $n \in \mathcal{J}$, some countable index set—, and the probabilities of outcomes are denoted as $P[Y = y_n] = p_n$, the *Rényi entropy of order* α is defined as [119]

$$H_\alpha (Y) = \frac{1}{1-\alpha} \log_2 \left(\sum_{n \in \mathcal{J}} p_n^\alpha \right) . \tag{3.38}$$

The base of the logarithm only changes the units in which the randomness is measured. For the choice of 2, randomness is measured in bits, in line with all previous discussions in this chapter.

Two specific types of the Rényi entropy are discussed for the geometric and dead time distorted geometric distributions: the min-entropy H_∞ and the Shannon entropy H_1 or simply H. The first is a more conservative measure of surprise, since $H_\infty (Y) \leq H_1 (Y)$ for any discrete random variable Y; the equality holds in case Y is uniformly distributed. More generally, the Rényi entropies of a given distribution are non-increasing in α [119].

The Rényi entropies share similarities with the bit generation efficiency η_R, since both quantify the number of bits that are (or can potentially be) generated by a given random variable. Analogously, we can define *entropies per unit time* as

$$h_\alpha (Y, S) = \frac{H_\alpha (Y)}{\mathbb{E}[S]}, \tag{3.39}$$

where S is an underlying continuous random variable, describing the duration necessary to measure Y, its discretized approximation. These quantities can be directly compared to the bit generation rate; all of them are actually upper bounds for R. h_∞ is the strictest bound, since the same inequalities hold as for the simple entropies.

Min-entropy. The min-entropy H_∞ of a discrete random variable quantifies the maximum amount of independent, uniformly distributed random bits that can be generated from the process [38, 120]. Using the notation introduced above, H_∞ can be defined as

$$H_\infty (Y) = \min_{n \in \mathcal{J}} (- \log_2 p_n) = - \log_2 \max_n p_n . \tag{3.40}$$

For the geometric (Eq. 3.9) and dead time distorted geometric distributions (Eq. 3.34), described by the variable X, the set of outcomes $\{ y_n \}$ is the set of all non-negative integers equal to or larger than $k = \lfloor \tau_\mathrm{d} / \tau \rfloor$; the notation can therefore be simplified as $y_n = n,\ n \in \{k, k+1, k+2, \dots\}$. For $k = 0$, this coincides with the natural numbers $\mathbb{N}$ (with 0 included). Let us use the notation

$$n_{\mathrm{m}} = \arg\max_n p_n \tag{3.41}$$

for the most probable outcome, so that $H_\infty(X) = -\log_2 p_{n_{\mathrm{m}}}$. It is clear that for the (shifted) geometric distribution, n_{m} is always equal to k. However, it has been shown that for the dead time distorted geometric distribution, n_m can be either k or $k+1$, depending on the values of $\lambda\tau$ and $\Delta\tau/\tau$. As the fractional dead time increases, the probability is being redistributed from k to $k+1$, and eventually the latter surpasses the former. The min-entropy is maximized in the intermediate case when the exact equality $p_k = p_{k+1}$ holds; thus, when

$$1 - \mathrm{e}^{-\lambda(\tau-\Delta\tau)} = \mathrm{e}^{-\lambda(\tau-\Delta\tau)}\left(1 - \mathrm{e}^{-\lambda\tau}\right). \tag{3.42}$$

This equation may be rewritten in the $[\lambda\tau]$–$\big[\Delta\tau/\tau\big]$ formulation. It can be solved for the fractional dead time that maximizes H_∞ for a given $\lambda\tau$, yielding the optimum

$$\left(\frac{\Delta\tau}{\tau}\right)_{\mathrm{opt}} = \frac{-\ln\left(2\mathrm{e}^{-\lambda\tau} - \mathrm{e}^{-2\lambda\tau}\right)}{\lambda\tau}. \tag{3.43}$$

If $\Delta\tau/\tau$ is smaller than this value, $p_k > p_{k+1}$, and vice versa. The corresponding maximal min-entropy for a certain product $\lambda\tau$ may now be calculated from either p_k or p_{k+1} by substituting the optimum in Eq. 3.43:

$$\begin{aligned} H_\infty^{\max}(X;\lambda\tau) &= -\log_2(p_k)\Big|_{\left(\frac{\Delta\tau}{\tau}\right)_{\mathrm{opt}}} \\ &= -\log_2\left(1 - \mathrm{e}^{-\lambda(\tau-\Delta\tau)}\right)\Big|_{\left(\frac{\Delta\tau}{\tau}\right)_{\mathrm{opt}}} \\ &= -\log_2\left(1 - \frac{1}{2 - \mathrm{e}^{-\lambda\tau}}\right). \end{aligned} \tag{3.44}$$

In the slow-clock limit, as $\lambda\tau \to \infty$, this maximum decreases to $-\log_2(0.5) = 1$, whereas in the fast-clock limit, the min-entropy grows without bounds for any value of $\Delta\tau/\tau$. Figure 3.13 shows the min-entropy in terms of $\lambda\tau$ and $\Delta\tau/\tau$. On the 3D plot (left), a "ridge" becomes visible, along which $H_{\min}$ is maximized for the given $\lambda\tau$ product. The function for $\big(\Delta\tau/\tau\big)_{\mathrm{opt}}$ (Eq. 3.43) is also plotted onto the top view (right). As the ridge perfectly overlaps with the dotted function, the calculations are proven to be correct. Moreover, the marginal view from the $\lambda\tau$ axis would reveal the top contour of the plot to be equal to the value in Eq. 3.44.

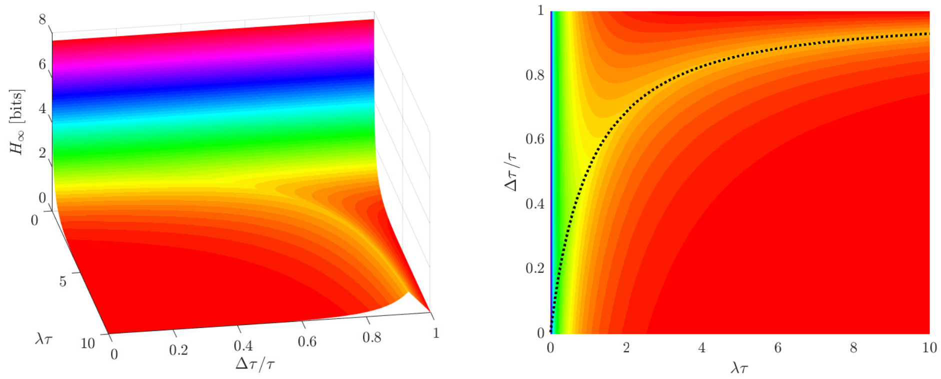

Figure 3.13: Min-entropy $H_\infty(X)$ as a function of $\lambda\tau$ and $\Delta\tau/\tau$. (a) 3D plot of the function showing a "ridge" in yellow; (b) top view with $\left(\frac{\Delta\tau}{\tau}\right)_{\text{opt}}$ highlighted by a dashed black line.

Shannon entropy. The Shannon entropy, widely used in classical areas, e.g. in source coding and lossless compression, defines probability-weighted average randomness of the possible outcomes. The index $\alpha = 1$ is usually dropped, leading to the definition and notation

$$H_1(Y) = H(Y) = -\sum_{n\in\mathcal{J}} p_n \cdot \log_2(p_n). \tag{3.45}$$

The geometric distribution with parameter p has the entropy

$$H^{\text{geom}}(X;p) = \frac{-(1-p)\cdot\log_2(1-p) - p\cdot\log_2(p)}{p}. \tag{3.46}$$

Substituting $p = \left(1-\mathrm{e}^{-\lambda\tau}\right)$, it can rather be expressed in terms of $\lambda\tau$:

$$H^{\text{geom}}(X;\lambda\tau) = \frac{\lambda\tau\cdot\log_2(\mathrm{e})\cdot\mathrm{e}^{-\lambda\tau} - \log_2\left(1-\mathrm{e}^{-\lambda\tau}\right)\cdot\left(1-\mathrm{e}^{-\lambda\tau}\right)}{1-\mathrm{e}^{-\lambda\tau}}. \tag{3.47}$$

The Shannon entropy of the dead-time distorted geometric distribution (Eq.3.34) can be phrased in terms of Eq. 3.47, yielding a somewhat less elegant closed-form solution:

$$\begin{aligned} H^{\text{dead}}(X;\lambda\tau,\lambda\Delta\tau) = {} & \mathrm{e}^{\lambda\Delta\tau}\cdot H^{\text{geom}}(X;\lambda\tau) - \lambda\Delta\tau\cdot\log_2(\mathrm{e})\cdot\mathrm{e}^{-\lambda(\tau-\Delta\tau)} \\ & + \log_2\left(1-\mathrm{e}^{-\lambda\tau}\right)\cdot\left(1-\mathrm{e}^{-\lambda\tau}\right)\cdot\mathrm{e}^{\lambda\Delta\tau} \\ & - \log_2\left[1-\mathrm{e}^{-\lambda(\tau-\Delta\tau)}\right]\cdot\left[1-\mathrm{e}^{-\lambda(\tau-\Delta\tau)}\right]. \end{aligned} \tag{3.48}$$

Just as in case of η_R, the Shannon-entropy is also calculated using an infinite sum for the distributions in question. Therefore, its value is translation-invariant—independent of k, but depending on $\Delta\tau/\tau$. Note that the two entropies are in the same family, as

$$\lim_{\Delta\tau\to 0} H^{\text{dead}}(X;\lambda\tau,\lambda\Delta\tau) = H^{\text{geom}}(X;\lambda\tau). \tag{3.49}$$

For both distributions, the entropy decays to zero in the slow-clock limit and diverges towards infinity in the fast-clock limit. See Fig. 3.14(a) for details. The corresponding Shannon entropies per unit time are derived from Shannon entropies just as R descends from η_{R}:

$$h^{i}\left(X\right)=\frac{H^{i}\left(Y\right)}{\mathbb{E}\left[T\right]}=\frac{\lambda}{1+\lambda\tau_{\mathrm{d}}}H^{i}\left(X\right), \tag{3.50}$$

where $i\in\{\,\mathrm{dead},\mathrm{geom}\,\}$. Figure 3.14(b) shows the entropy per unit time as a function of λ for different scenarios. The plot uses the same sets of parameters as Fig. 3.11(a) for the sake of a quick visual comparison. The bit generation rate and h^{i} are similar qualitatively, but quantitatively the latter is significantly higher than the former. R never exceeds 0.3 Gbps for the given parameter set, while the entropy per unit time may even exceed 1.5 Gbps. Altogether, it can be concluded that there could indeed exist methods, which can exploit more of the available randomness from the variables $\{\,X_j\,\}$ in case of time measurement with a restartable clock signal.

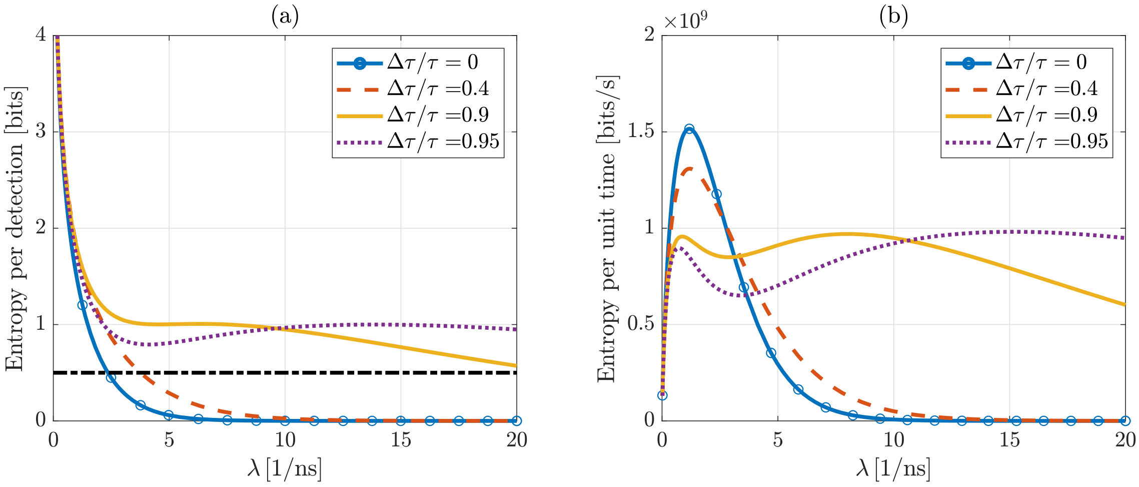

Figure 3.14: (a) Shannon entropy and (b) Shannon entropy per unit time of the geometric/dead time distorted geometric distributions for $k=0$ and $\tau=1\,\mathrm{ns}$ and different values of $\Delta\tau/\tau$. The dashed line on the left denotes 0.5, the theoretical maximal bit generation efficiency of the current method. Taken from Ref. [1].

3.4 Experimental Verification and Bit Generation

After the convincing simulation results, the method and its descriptive model was checked in practice as well. This consisted of building an experimental setup, finding the limitations imposed by the devices, then comparing theoretical curves to measurement data, and finally subjecting the generated bit sequences to randomness testing.

3.4.1 Setup of the QRNG

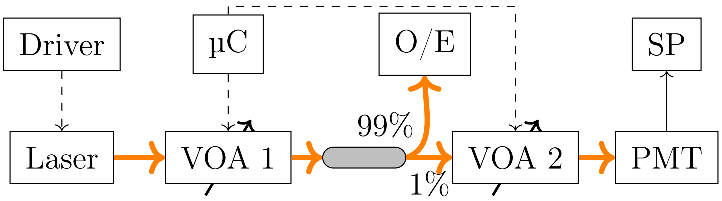

Figure 3.15: Experimental setup of the time-of-arrival QRNG scheme. Driver: laser driver; VOA: variable optical attenuator; µC: microcontroller; O/E: photodiode; PMT: photomultiplier tube; SP: signal processing. Taken from Ref. [113].

The physical setup (Figs. 3.15 and 3.16) consists of the following elements. The photon source is a semiconductor laser (*Thorlabs LP520-SF15*) emitting around a central wavelength 519.9 nm. Biasing is managed by a driver circuit controlling the current to maintain a predefined value. All optical fibers (*Thorlabs 460HP*) are specifically designed to aid single transverse mode propagation in the wavelength band of interest, using a core diameter of 2.5 µm. The light intensity is attenuated using two successive voltage controlled variable optical attenuators (VOA, *Thorlabs V450F*), which are set by a designated microcontroller-based board, and an optical splitter (*Thorlabs TW560R1F1*). The 99% port of the splitter is used for monitoring, measured by an amplified photodetector (*Thorlabs PDA10A2*), while the 1% port transmits the signal, providing 20 dB of attenuation. Attenuators serve a dual purpose: they protect the single-photon detector (a photomultiplier tube, *PicoQuant PMA-175 NANO*) against high levels of input power, and they are responsible for tuning the photon rate so that the operation can be evaluated on a wide range of parameter values. Finally, the PMT's output voltage pulses are time-tagged by a *PicoQuant TimeHarp 260* time-to-digital converter (TDC) with a base resolution of $\tau_\mathrm{b} = 250\,\mathrm{ps}$. The comparison of time intervals and the corresponding bit assignment is software-based, running on a computer directly accessing the TDC.

The exclusion of noise from the model can be explained by the parameters of the PMT in question. Owing to the fact that it is sensitive towards the blue end of the visible spectrum, the afterpulsing probability is zero, while the rate of dark counts—which are thermal in origin—is specified to be smaller than 50 counts per second at room temperature. When the whole setup was covered by a box, ensuring that no external light is coupled into the fibers or directly into the PMT, the measured total amount of dark and ambient counts was below 1 cps. This is significantly lower than any input photon rate of practical interest. Thus, the randomness is almost fully extracted from purely quantum effects. Further notable detector properties include

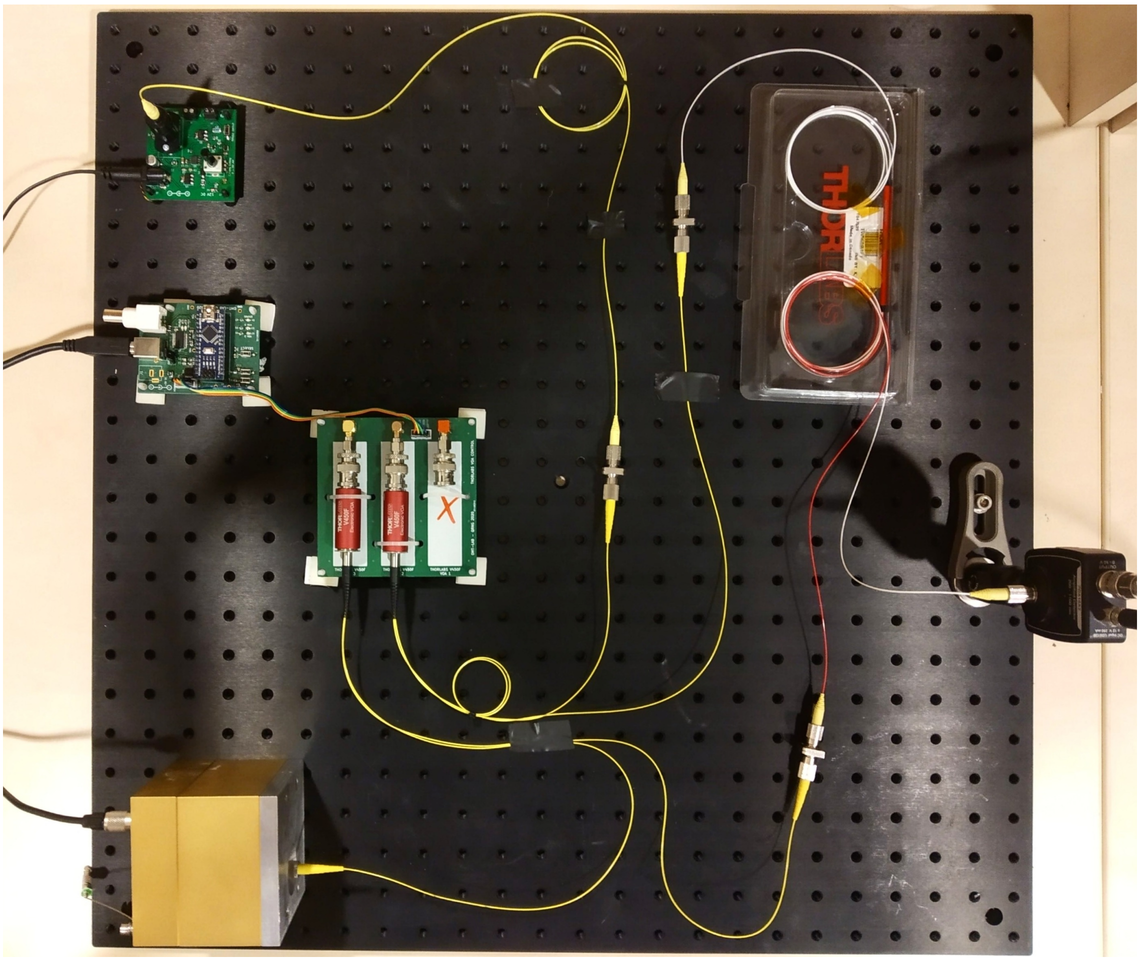

Figure 3.16: Photograph of the time-of-arrival QRNG. Top left: laser and laser driver; middle left: VOA driver and VOAs; bottom left: PMT; top right: splitter; middle right: photodiode. The TDC and the compueter are not visible.

the maximum allowed output photon rate $\lambda_{\mathrm{out,max}} = 5 \cdot 10^6$ counts per second, the quantum efficiency, which is approximately 21% at 520 nm, and the 180 ps FWHM transit time spread.

The dead time of the detection system is dominated by that of the TDC with an average value of $\tau_{\mathrm{d}} = 2\,\mathrm{ns}$. The PMT has no direct dead time—that is, it remains sensitive after detection, but detections happening sooner than 1.5 ns, the average output pulse width, cannot be discriminated. τ_{d} is not constant, as it could be seen experimentally. For the comparison of experimental and theoretical results, the latter are still calculated assuming 2 ns long constant dead time. This simplification does not bring forth significant differences. Although it is not explicitly stated anywhere that the dead time is indeed non-extendable, but we can argue in favor of this statement. Assume that the detection system is paralyzable. If τ_{d} is significantly shorter than the mean time between photon arrivals, such that

$$\tau_{\mathrm{d}} \ll \frac{1}{\lambda} \Longleftrightarrow \lambda\tau_{\mathrm{d}} \ll 1, \tag{3.51}$$

the probability of a new arrival during the dead time duration is small. Therefore, the insensitive period is unlikely to be extended, and the detector behaves approximately as non-paralyzable. This condition holds for the experiments, as it will be shown later.

The most significant deviation from the mathematical model is, however, something else. The TDC operates with a continuous (non-restartable) clock signal, which introduces correlations between successive bits and might change the bit generation efficiency and rate drastically for the same parameter set. In the next section, I introduce the *high-precision regime*, a domain of operation where differences between continuous and restartable clocks are negligible both in terms of the quality of randomness and the measures of bit generation, so that the experimental setup can be used to validate the proposed model.

3.4.2 Effects of a Continuous Clock: The High-Precision Regime

The *high-precision regime* (HPR) is defined as the domain of the parameter space for which the following condition holds:

$$\tau \ll \frac{1}{\lambda} \iff \lambda\tau \ll 1. \tag{3.52}$$

The name of the domain implies that if condition 3.52 is true, then the discrete time measurement yields a very precise result, since the average number of clock periods per photon arrival time will be high. As $\lambda\tau$ increases, the precision decreases, and the setup leaves the HPR. The notion is similar in content to the expression *fast-clock limit* introduced in Ref. [111], but slightly more generous, as $\lambda\tau$ should only be “small”, rather than approaching zero.

In the experiments, $\lambda_{\text{out,max}}$ is limited by the detector's tolerance. The corresponding maximal input photon rate is

$$\lambda_{\text{max}} = \frac{\lambda_{\text{out,max}}}{1 - \lambda_{\text{out,max}}\tau_{\text{d}}} \approx 5.051 \cdot 10^{6} \left[\frac{1}{\text{s}}\right] \tag{3.53}$$

from which the worst-case—highest—values of $\lambda\tau$ and $\lambda\tau_{\text{d}}$ can be calculated to be

$$\max_{\lambda} \lambda\tau = \lambda_{\text{max}}\tau \approx 1.26 \cdot 10^{-3} \text{ and} \tag{3.54}$$

$$\max_{\lambda} \lambda\tau_{\text{d}} = \lambda_{\text{max}}\tau_{\text{d}} \approx 1.01 \cdot 10^{-2}. \tag{3.55}$$

Thus, the products are at least two orders of magnitude smaller than 1 in my experiments, and both the HPR condition (Eq. 3.52) and the argument for using a non-extendable dead time model (Eq. 3.51) hold.

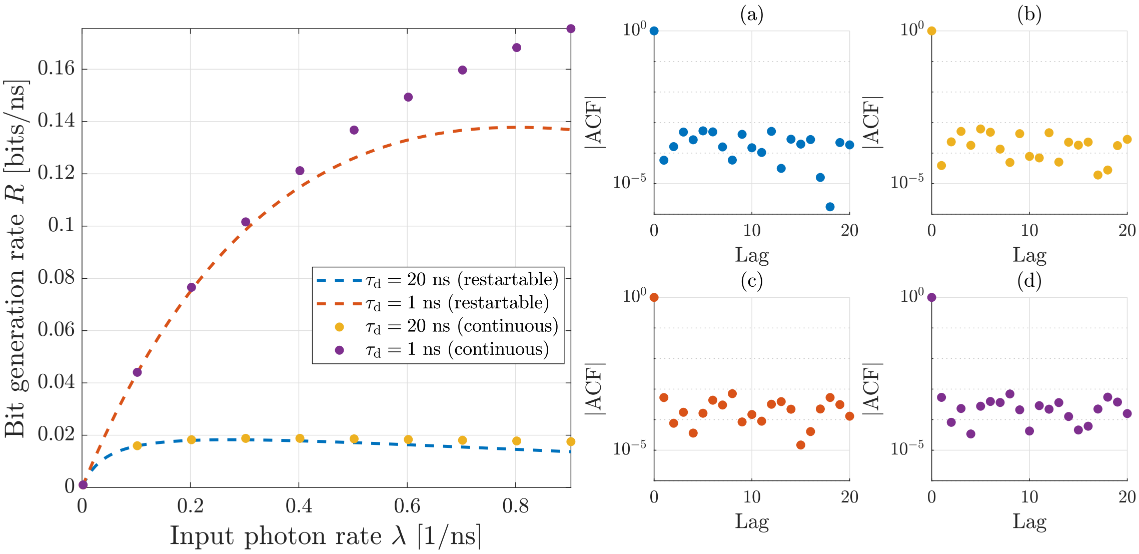

Figure 3.17: Simulations targeting the claim about the high-precision regime. Left: Bit generation rate comparison of restartable clock measurements (dots) of the time sequence for $\tau = 1\,\mathrm{ns}$, and theoretically derived curves for continuous clock measurements (dashed lines) as a function of λ, using two different dead time values. Right: Absolute values of bit sequences' autocorrelation coefficients using $\lambda = 0.002\,\mathrm{ns}$ ($\lambda\tau = 2 \cdot 10^{-3}$). (a) and (c): restartable clock, (b) and (d): continuous clock; top row: $\tau_d = 20\,\mathrm{ns}$, bottom row: $\tau_d = 1\,\mathrm{ns}$. Taken from Ref. [113].

It can be argued that when the QRNG operates in the high-precision regime, the differences between continuous and restartable clock methods are being suppressed, as the drift between the two clocks is just a small fraction of the total time elapsed between detections. In Ref. [111], it was already shown that the coefficient measuring correlation between successive bits, a_1, vanishes in the fast-clock limit. The claim needs to be further supported by simulations, both concerning the autocorrelation functions and the agreement between the respective η_R and R curves.

HPR simulations. The MATLAB simulations generated several 10^7 long arrays of arrival time differences. For each array, a pair from two values of dead time (1 and 20 ns) and ten values of λ between 0 and 1 [1/ns] were chosen. The clock period τ was 1 ns, and bits were generated using both clock types from each dataset. Results are compared in Fig. 3.17. The left subplot shows the bit generation rate as a function of λ for the specified dead time values and clock types. It can be seen that the curves belonging to the same τ_d are in excellent agreement as $\lambda\tau \to 0$, but start deviating around $\lambda\tau = 0.2$.

On the right, autocorrelation functions are shown for the bit sequences generated using the smallest λ. The $\lambda\tau$ product is $2 \cdot 10^{-3} > \lambda_{\max}\tau$, describing a more problematic scenario than what shall be encountered experimentally. The top row represents the two sequences generated with $\tau_d = 20\,\mathrm{ns}$, the bottom row represents those with $\tau_\mathrm{d} = 1\,\mathrm{ns}$. The first column shows the coefficients for a restartable clock, the second those for a continuous clock. There are no significant correlations in either case; most importantly, none between successive bits (Table 3.2) generated with a continuous clock. Moreover, the reported a_1 values are all slightly negative, whereas for a non-restartable clock positive values are expected [111]. This latter expectation was further confirmed by my simulations (e.g. when using $\lambda = 0.902\,[1/\mathrm{ns}]$).

Table 3.2: Autocorrelation coefficient a_1 of simulated bit sequences.

	Clock	
Dead time [ns]	**Restartable**	**Continuous**
1	$-2.48 \cdot 10^{-4}$	$-2.55 \cdot 10^{-4}$
20	$-3.18 \cdot 10^{-4}$	$-3.56 \cdot 10^{-4}$

Altogether, simulations confirmed that the results do not significantly differ between the two clock types in the high-precision regime. Therefore, the experimental setup is able to produce bit sequences and data by which the model can be faithfully validated. This does not necessarily suggest that the experimentally generated bits are going to pass all statistical tests; the cases when $\lambda\tau$ is close to $\lambda_\mathrm{max}\tau$ are going to be especially susceptible to problems. It is, however, not expected that measured bit generation rates are going to depart from theoretical values.

3.4.3 Experimental Validation of the Mathematical Model

In order to validate the predictions of the mathematical model outlined in Section 3.3—that is, whether the theoretical functions derived for η_R and R conform to reality—, bits need to be generated on a wide range of input parameter values. From the three parameters, λ can be tuned without difficulty. τ may be increased from its base value on the TDC, but it will remain non-restartable, whereas the dead time is cumbersome to influence. First, experiments were conducted with the best time measurement precision, τ_d was not altered externally, and the optical power was changed in approximately even steps (around $5 \cdot 10^5$ counts per second) to cover the full available range.

The bit generation efficiency and rate was calculated for each particular step using the available data. Note that at this point, no randomness testing had been conducted yet, focus was only on the macroscopic metrics. Figure 3.18 shows the obtained results for η_R (left) and R (right): markers denote experimental data, whereas dashed lines represent the theoretical curves.

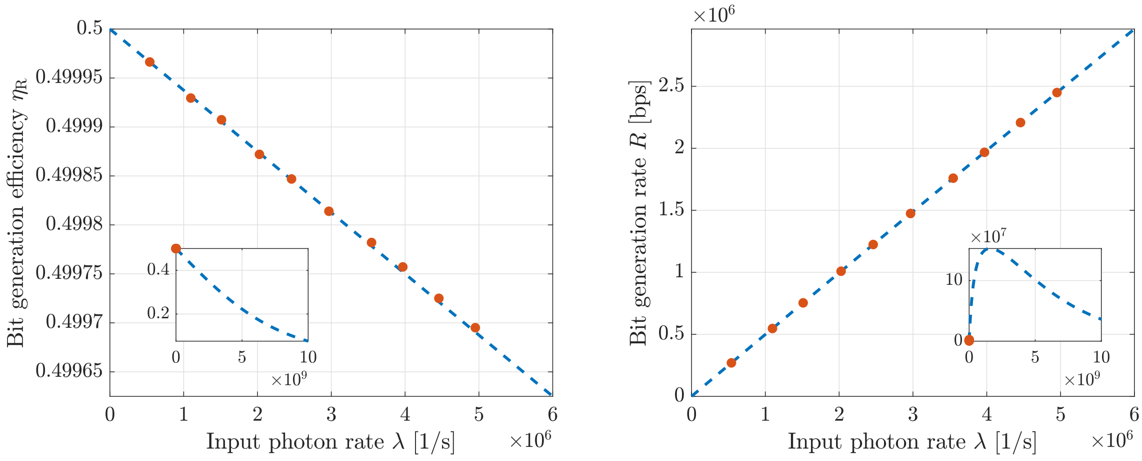

Figure 3.18: Bit generation efficiency η_{R} (left) and bit generation rate R (right) measured for $\tau = 250\,\mathrm{ps}$, $\tau_{\mathrm{d}} = 2\,\mathrm{ns}$ and different values of λ. Lines: theoretical curves, markers: measured values. Insets show functions for a wider range of λ for comprehensive understanding. Taken from Ref. [113].

The agreement between practice and theory is outstanding; the small deviations can be attributed to the finite amount of generated bits, from which the metrics were calculated. Note, however, the small insets in both figures, which show the functions on a larger subset of their domain, revealing that experiments could only inspect their first, quasi-linear sections. For the given time-measurement precision and dead time, the bit generation rate would ideally peak at $\lambda_{\mathrm{opt}} = 1.5952 \cdot 10^9$ [1/s], yielding 152.88 Mbps. The PMT is thus a bottleneck in the current setup, severely limiting the number of generated bits per unit time to approximately 2.5 Mbps. I deemed the evidence gathered this far hopeful, but inadequate for a complete validation of the model. For this, a secondary investigation was started, in which τ was changed as well—but not by changing the settings of the time-to digital converter.

Software-Based Clock Signals. As mentioned above, the TDC is only able to operate with a continuous clock. If τ was increased in hardware, the condition in Eq. 3.52 would become less and less true, leaving the high-precision regime. This solution is therefore unsuitable for the purpose of further model validation. However, the time differences recorded with the base precision could be re-sampled using a much slower software-based clock, which is easy to restart at each detection. This way, the problem of restartability would cease to exist, and the introduced discrepancies would be negligible.

The following reasoning helps support the claim. Suppose that a time interval sequence is measured with a (relatively) high precision, using a clock signal with period τ_{b}. At this point, it does not matter whether it was restartable or not. Now, from the perspective of a software-based clock signal with period τ_{sw}, for which $\tau_{\mathrm{sw}} >> \tau_{\mathrm{b}}$

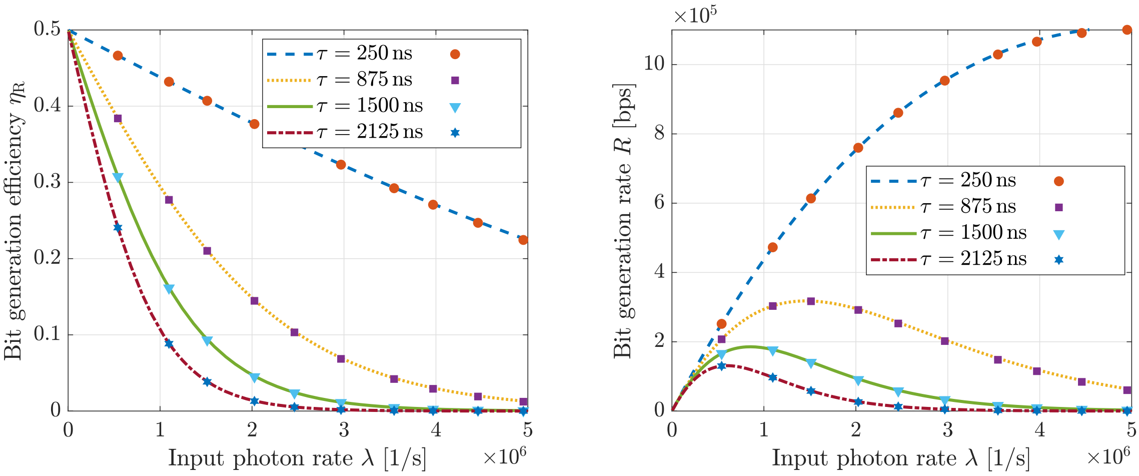

Figure 3.19: Bit generation efficiency η_R (left) and bit generation rate R (right) measured for $\tau_\mathrm{d} = 2$ ns, selected values of software-based resampling clock periods τ and different values of λ. Lines: theoretical curves, markers: measurement values. Taken from Ref. [113].

holds, the previously recorded values are almost analog. The fact that these were already discretized before are only affecting the resampled values at the level of some low-significance decimals. The same effect was present in the simulations of Section 3.3.3, since the generated pseudo-random interval lengths had finite but high precision, compared to which the clock signal was far less accurate. No inspection showed any problems during simulations, neither as deviations from theoretical curves, nor as failures in randomness testing, suggesting that re-sampling is a valid operation.

I tested four different values of τ_sw, each of them at least a thousand times longer than the base precision of 250 ps: 250, 875, 1500 and 2125 ns. The datasets were the ones obtained for the results presented in Fig. 3.18. These clock period selections allow us to observe the peaks in the R-curves. The outcomes are concluded in Fig. 3.19. Once again, the validity of the mathematical model is confirmed with certainty, as the experimental data almost perfectly matches the theoretical predictions.

3.4.4 Randomness Testing of Experimentally Generated Bits

All that is left is to generate suitably long sequences of bits at different count rates (1 to 5 million output counts per second, with a step of ∼1 million) and see whether the generator can pass the statistical tests in the NIST Test Suite. Testing is performed as outlined in Section 1.4.2, using one thousand subsequences, each one million bits long, at each value of λ.

The bit sequences generally performed well on the tests: the ones recorded at the three lowest λ values (1.01, 2.07 and 3.04 million/s) passed each of the 188 total subtests both on the aspect of proportion and the uniformity of p-values. At higher input photon rates, as the high-precision condition weakens, the bit sequences start failing the *Runs* test on both aspects. This can be attributed to the continuous clock signal used for measurement. The Runs test counts the number of blocks of uninterrupted sequences of identical bits—such a block is called a run—and compares it to a theoretical value. The software's detailed output suggest that the number of runs is generally too few: the switches between ones and zeros is not as frequent as it should. This is clearly a consequence of a positive autocorrelation coefficient between successive bits, which stems from the non-restartable nature of the clock. The bit sequence generated from the highest optical power signal failed one more test: a *Non-Overlapping Template Matching* tests, which counts the occurrences of the specific template *000001111*. This template, which rarely switches between different bits, occurs much more frequently than it should in some subsequences, pointing at the presence of a positive autocorrelation between subsequent bits. The test only failed at the proportionality aspect (978/1000 success rate, slightly below the acceptable 980/1000). Table 3.3 sums up the number of successful tests for the different bit sequences, along with the proportion of success and uniformity p-values obtained in case of the Runs test.

Table 3.3: Number of successful tests out of 188 for the time-of-arrival QRNG method for different photon rates with detailed results of the Runs test. Asterisks (*) denote failure of the test on the specific aspect.

Mean λ [10^6/s]	**Successful tests**	**Runs test**	
		Proportion	**Uniformity p-value**
1.01	188	994/1000	0.331408
2.07	188	991/1000	0.382115
3.04	188	985/1000	0.003967
3.72	187	977/1000*	0.000000*
4.80	186	792/1000*	0.000000*

All together, in the HPR, the generator is capable of producing random numbers that cannot be distinguished from the output of an ideal QRNG, without the need for any kind of post-processing. As the HPR condition weakens, the generator becomes prone to slight errors due to the induced autocorrelations. These could be removed by a simple, computationally non-demanding post-processing method, e.g. a bitwise self-delayed XOR operation with a delay longer than the distance over which bits are significantly correlated.

3.5 Conclusion

I have created a mathematical model of a quantum random generation scheme based on the difference of successive photon arrival times. The model focuses mainly on the architecture's macroscopic metrics, the bit generation efficiency and rate. I have also shown that these two functions are not maximized simultaneously; the highest possible rate is achieved by loosening the requirements towards the efficiency.

The model's validity has been supported by simulations, focusing on both the relevant functions and the quality of randomness, while also incorporating the real system's deviations from the model. Finally, experimental results have been demonstrated to be in almost perfect agreement with theory, and the generator was found to work properly under most practical conditions. Slight correlations are only introduced by a continuous clock signal used for time measurement at the configurations yielding the highest bit generation rates. The device is able to produce sequences of random numbers that cannot be distinguished from those coming from a truly uniform distribution. This chapter forms the basis of Thesis II.

Chapter 4

Efficiency Improvement of the Time-of-Arrival QRNG Method

4.1 Introduction

No matter how good are the results reported for the time-of-arrival quantum random number generator described in the previous chapter, it can be seen that it features a significant trade-off. Reliability and stability are the main benefits, but the scheme only extracts a fragment of the available min-entropy, so the bit generation rates are moderate. This chapter introduces a new method of random number generation building on the “old” scheme. This method is able to increase both the bit generation rate and efficiency, while keeping most of the positives of its predecessor; such as an inherently almost-uniform distribution, relatively simple hardware and no post-processing. All this comes at the expense of needing a more stable light source and slightly more complex signal processing.

The method described in Chapter 3 achieves robustness and a uniform distribution of bits through intentionally discarding all cases where two successive time measurements yielded equal values. By doing so, the bit generation efficiency η_{R} remains smaller than 0.5 in any situation—and it has been shown that maximizing the efficiency does not maximize the bit generation rate R simultaneously.

I propose a new bit generation scheme using the same types of time measurement comparisons as described in the previous chapter, which increases both η_{R} and R compared to the old scheme by precisely setting the distribution of comparison signs and only discarding a smaller set of outcomes. All notations are retained from Chapter 3; however, the optimal photon rate, the bit generation rate and the maximal bit generation rate of the old method are now denoted as λ'_{opt}, R' and $R'_{\max}$, respectively, while R refers to the rate of the new method.

4.2 Grouping of Comparisons for Improved Efficiency

Briefly described, the improved method is as follows. The input photon rate λ (once again, including all kinds of losses, along with the quantum efficiency of the single-photon detector) is set to a certain value $\lambda_0(\tau, \Delta\tau)$, such as the probability of equality $P_{\mathrm{eq}}(\lambda_0, \tau, \Delta\tau)$ between successive measurements takes the value 1/3. Thus, the variables describing the sign of the i^{th} comparison, $W_i = \operatorname{sgn}(Y_i)$ will be uniformly distributed on the set $\{-1, 0, 1\}$:

$$P\left[W_i = n\right] = \begin{cases} 1/3, & n \in \{-1, 0, 1\} \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

Now, let us form m-long ($m \in \mathbb{Z}^+$) disjoint groups from successive comparison results to obtain vector-valued variables $V_s = \left(W_{ms}, W_{ms-1}, \ldots W_{ms-(m-1)}\right)$. Since $\{ W_i \}$ are i.i.d random variables—note that the measurement clock is still assumed to restart at every detection—, every V_s is uniformly distributed on $\{ -1, 0, 1 \}^m$. The cardinality of this set is 3^m, so the probability that V_s takes either of its values is $1/3^m$. The min-entropy limiting the maximum number of uniformly distributed random bits that can be generated from such a variable is

$$H_\infty^m \left(V_s\right) = -\log_2 \left(\frac{1}{3^m}\right) = m \cdot \log_2 \left(3\right). \tag{4.2}$$

Only an integer number of bits can be assigned to any random variable; therefore, the best possibility is to assign $\lfloor m \cdot \log_2 (3) \rfloor$ bits to an m-long group. Obviously, this is only enough to cover $2^{\lfloor m \cdot \log_2 (3) \rfloor}$ outcomes, which is strictly less than the total 3^m. Those outcomes that have no associated bit sequence will be discarded; as long as the circumstances are ideal, it does not matter which ones.

It will become obvious that increasing the value of m does not necessarily imply a higher efficiency or generation rate, but the space generated by the outcomes is growing exponentially, making the mapping of bits to outcomes increasingly troublesome.

4.2.1 New Bit Generation Efficiency

The bit generation efficiency is defined the same way as previously: it quantifies the average number of bits generated per random event. To be comparable to the old method, one *random event* should still mean one time difference between successive photon detections. If m was chosen to be one, the available min-entropy is $H_\infty^1 \left(V_s\right) = H_\infty \left(W_i\right) = \log_2 \left(3\right) \approx 1.585$ bits, smaller than two. Ultimately, assigning one bit to two of the three outcomes and discarding one of them is the same as the old method at the specific efficiency of $1/3$.

Generally, however, all other m values yield higher efficiencies than $1/3$, and most, but not all, are above $1/2$—the theoretical limit of the old scheme. The efficiency in terms of m takes the form

$$\eta_{\mathrm{R}} \left(m\right) = \underbrace{\frac{\lfloor m \cdot \log_2 \left(3\right) \rfloor}{2m}}_{\eta_{\mathrm{R,A}}(m)} \cdot \underbrace{\frac{2^{\lfloor m \cdot \log_2 (3) \rfloor}}{3^m}}_{\eta_{\mathrm{R,B}}(m)}, \tag{4.3}$$

depicted in Fig. 4.1. The first term, $\eta_{\mathrm{R,A}} \left(m\right)$, is the number of assigned bits divided by the number of random events required for a bit assignment (m comparisons constitute of $2m$ random events). It is bounded from above

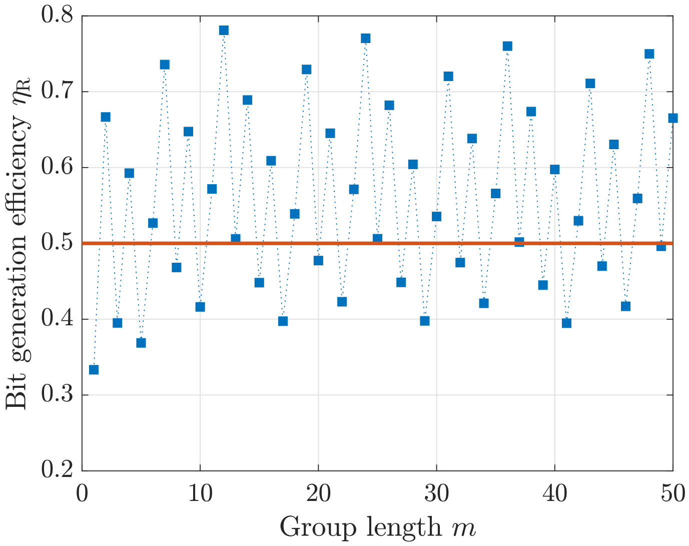

Figure 4.1: Bit generation efficiency $\eta_{\mathrm{R}}(m)$ as a function of group length m. The red line denotes 0.5, the theoretical maximum efficiency of the old method. Taken from Ref. [5].

by $\log_2(3)/2 \approx 0.7925$, which is also a good approximation of the term for large m. The second term, $\eta_{\mathrm{R,B}}(m)$, is the ratio of outcomes that are not discarded. Since $m \cdot \log_2(3) - 1 < \lfloor m \cdot \log_2(3) \rfloor < m \cdot \log_2(3)$, this "oscillates" between $2^{m \cdot \log_2(3)-1}/3^m = 0.5$ and $2^{m \cdot \log_2(3)}/3^m = 1$, and it is largely responsible for the suddenly changing pattern of the efficiency in terms of the group length (Fig. 4.2). Altogether, the efficiency has an upper bound of $\log_2(3)/2 \approx 0.7925$, the product of the individual upper bounds of the two terms. The first three highest efficiency values are the following:

$$\eta_{\mathrm{R}}(m) \approx \begin{cases} 0.6667 & \text{for } m = 2, \\ 0.7358 & \text{for } m = 7, \\ 0.7810 & \text{for } m = 12. \end{cases} \tag{4.4}$$

During theoretical analysis, $m = 2$ is given particular attention, but results are always derived for a general m as well.

4.2.2 Optimal Photon Rate for the New Scheme

As mentioned above, the method only provides perfect uniformity if the generator is operated at a certain combination of measurement clock period, non-extendable dead time and input photon rate—the latter denoted by λ_0. The optimal photon rate is a function of other parameters, τ and $\Delta\tau$, and it can be obtained by solving the equation $P_{\mathrm{eq}} = 1/3$ for λ. Once again, we may separate two cases: zero fractional dead time, so that P_{eq} takes the simple form in Eq. 3.15; and non-zero fractional dead

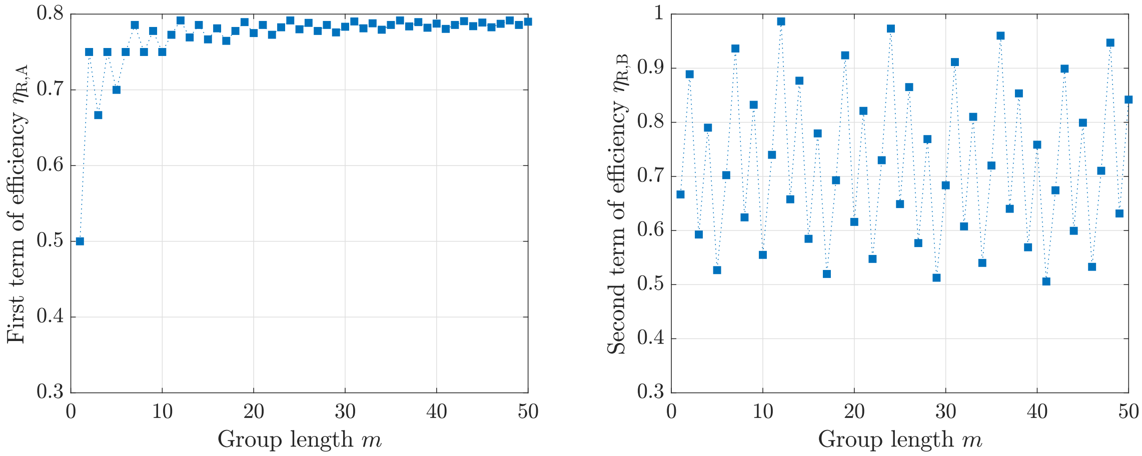

Figure 4.2: First and second terms of the bit generation efficiency $\eta_{\mathrm{R}}(m)$ as a function of group length m.

time, with the probability of equality being a more complicated function (Eq. 3.35). Since P_{eq} is independent of k, the same holds for λ_0. Knowing the optimal photon rate, the bit generation rate can be obtained combining Eqs. 3.1 and 3.28:

$$R_m(\tau, \tau_{\mathrm{d}}) = \frac{\lambda_0}{1 + \lambda_0 \tau_{\mathrm{d}}} \cdot \eta_{\mathrm{R}}(m) \tag{4.5}$$

Keep in mind that λ_0 is the same for all choices of m, since it is the photon rate that makes the prior distribution of W_i uniform.

Zero fractional dead time. In the first situation, the dead time is an integer multiple of the measurement clock period, $\tau_{\mathrm{d}} = k\tau$, $k \in \mathbb{N}$. The equation for the probability of equality,

$$\frac{\mathrm{e}^{\lambda\tau} - 1}{\mathrm{e}^{\lambda\tau} + 1} = \frac{1}{3}, \tag{4.6}$$

can be solved for λ analytically. The optimal photon rate is inversely proportional to τ, just as in the dead time free calculations for the old method (Eq. 3.22):

$$\lambda_0 = \frac{\ln(2)}{\tau}. \tag{4.7}$$

The corresponding bit generation rate is also proportional to the reciprocal of the clock period, but it also depends on the whole length of the dead time—thus, on k as well.

$$R_m(\tau, \tau_{\mathrm{d}} = k\tau) = \frac{\ln(2)}{\left[1 + k \cdot \ln(2)\right]\tau} \cdot \eta_{\mathrm{R}}(m) \tag{4.8}$$

Particularly, substituting $m = 2$ gives the function

$$R_2(\tau, \tau_{\mathrm{d}} = k\tau) = \frac{2\ln(2)}{3\left[1 + k \cdot \ln(2)\right]\tau} \approx \frac{0.4621}{\left[1 + k \cdot \ln(2)\right]\tau}. \tag{4.9}$$

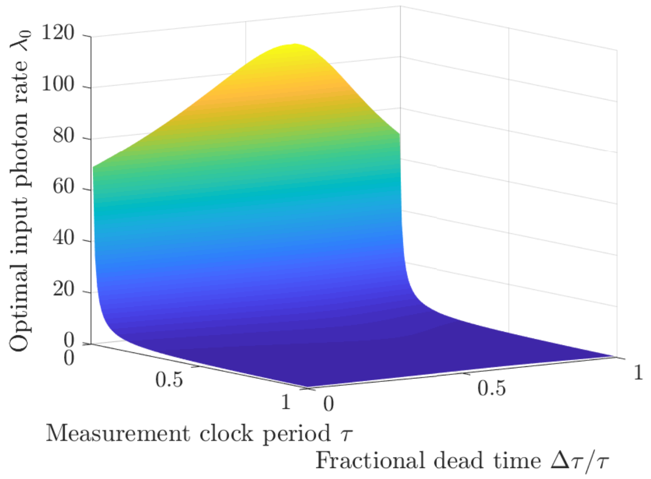

Figure 4.3: Optimal input photon rate λ_0 as a function of τ and $\Delta\tau/\tau$. Units of time and count rate are arbitrary and reciprocal to each other. Taken from Ref. [5].

Non-zero fractional dead time. If the probability of equality takes the more complicated form found in Eq. 3.35, the equation $P_{\text{eq}}(\lambda, \tau, \Delta\tau) = 1/3$ has no analytical solution for λ, and numerical calculations are needed. From λ_0, R_m can be found by simple substitution into Eq. 4.5. The values for λ_0 and R_2 as a function of τ and $\Delta\tau/\tau$ are shown on Fig. 4.3 and Fig. 4.4, respectively (the latter is restricted to $k = 0$).

It can be seen that for constant fractional dead time, λ_0 decreases presumably reciprocally with increasing τ, although the multiplicative coefficient depends on $\Delta\tau/\tau$. For constant τ, however, a fractional dead time between 0.6 and 0.7 maximizes λ_0. Similar observations can be made about the reciprocal decay of R_2 with reducing the time resolution. The highest bit generation rates can be produced for small τ (good measurement resolution) and small fractional dead times.

4.2.3 Bit Generation Gain Over the Old Scheme

I wanted to obtain a proof, along with quantitative results, to show the superiority of the improved scheme over the previous one. It has been done based on the bit generation efficiency by choosing a suitable m with $\eta_{\text{R}}(m) > 0.5$; now, it will be shown that the improvement applies to R as well.

One way to do this is to compare the bit generation rates of the two methods at λ_0—this is not completely fair, though, as λ_0 is not the optimal photon rate for the old method. The two can be analytically compared if $\Delta\tau = 0$; R' then takes the form

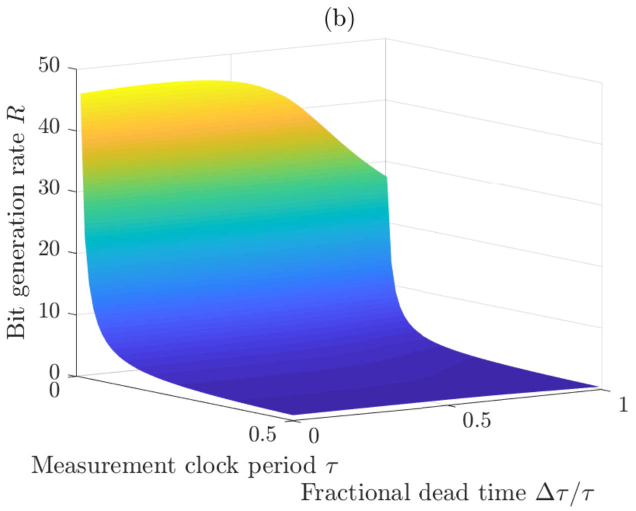

Figure 4.4: Bit generation rate R for $k = 0$ and $m = 2$ as a function of τ and $\Delta\tau/\tau$. Units of time are arbitrary; R is shown in bits per time unit. Taken from [5].

in Eq. 3.31. Substituting Eq. 4.7 yields

$$R'\left(\lambda_0, \tau, \tau_\mathrm{d} = k\tau\right) = \frac{\ln(2)}{3\left[1 + k \cdot \ln(2)\right]\tau} = \frac{1}{2} R_2\left(\tau, \tau_\mathrm{d} = k\tau\right); \tag{4.10}$$

thus, the new method generates bits at twice the rate in λ_0, assuming that $m = 2$. This clearly originates from the fact that η_R is $2/3$ at λ_0 for the new scheme and only $1/3$ for the old; the output count rate's multiplicative factor is the same for both. A better comparison can be obtained by evaluating the relation between the theoretical maximal rates of the two methods. For this, I defined the bit generation rate gain $G_\mathrm{R}m$ as the ratio of the respective peak bit generation rates of the new and old methods for a given τ and τ_d. In case of $m = 2$, the gain can be calculated as

$$G_\mathrm{R2}\left(\tau, \tau_\mathrm{d}\right) = \frac{R_2\left(\tau, \tau_\mathrm{d}\right)}{\max_\lambda R'\left(\lambda, \tau, \tau_\mathrm{d}\right)}. \tag{4.11}$$

Analytical solutions exist only if the dead time is taken to be zero, since the old method does not have a closed form for its peak bit generation rate R'_max for any non-zero τ_d. Using Eqs. 3.22 and 3.24, $\lambda'_\mathrm{opt} = \arg\max_\lambda R'\left(\lambda, \tau, \tau_\mathrm{d}\right)$ and $R'_\mathrm{max} = \max_\lambda R'\left(\lambda, \tau, \tau_\mathrm{d}\right)$ can be expressed in terms of λ_0 and R_2.

$$\lambda'_\mathrm{opt}\left(\tau, \tau_\mathrm{d} = 0\right) = \frac{W_0\left(\mathrm{e}^{-1}\right) + 1}{\tau} = \frac{W_0\left(\mathrm{e}^{-1}\right) + 1}{\ln(2)} \cdot \lambda_0\left(\tau, \tau_\mathrm{d} = 0\right) \tag{4.12}$$

$$\approx 1.8444 \cdot \lambda_0\left(\tau, \tau_\mathrm{d} = 0\right) \tag{4.13}$$

$$R'_\mathrm{max}\left(\tau, \tau_\mathrm{d} = 0\right) = \frac{W_0\left(\mathrm{e}^{-1}\right)}{\tau} = \frac{3 \cdot W_0\left(\mathrm{e}^{-1}\right)}{2\ln(2)} \cdot R_2\left(\tau, \tau_\mathrm{d} = 0\right) \approx 0.6026 \cdot R_2\left(\tau, \tau_\mathrm{d} = 0\right) \tag{4.14}$$

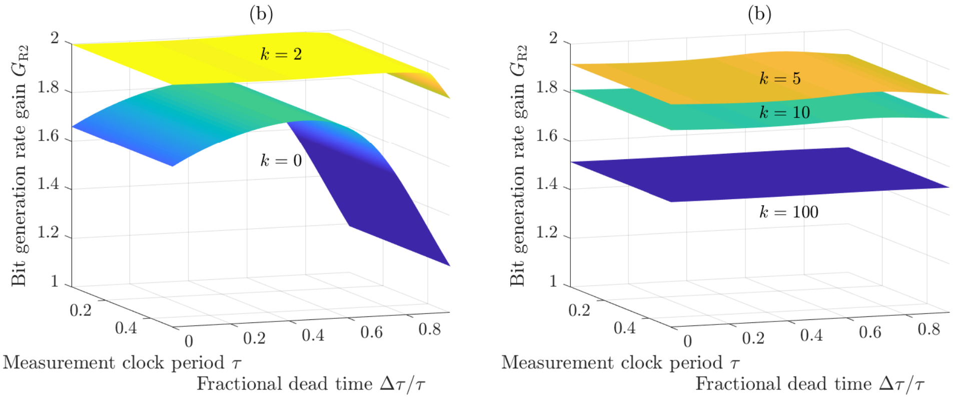

Figure 4.5: Bit generation rate gain G_{R2} as a function of τ and $\Delta\tau/\tau$ for (a) $k = 0$ and 2; (b) $k = 5$, 10 and 100. Both subplots have identical color scaling. Units of time are arbitrary. Taken from [5].

Thus, the gain for $m = 2$ and no dead time is approximately 1.6595. The inclusion of any positive value of dead time is, once again, requiring a numerical solver. Figure 4.5 concludes the results in two subplots. The gain was calculated for $k = 0$, 2, 5, 10 and 100, and $\Delta\tau/\tau$ ranging from 0 to 1. Note that significant numerical errors were found if k is small, τ is close to zero and $\Delta\tau/\tau$ approaches one: the argument of MATLAB's exponential function is too high to produce a meaningful result, and the peak of the old method's rate cannot be found. This section is therefore not shown on the graphs.

The plots reveal several interesting details. First, as the fractional dead time gets to 0, the surface representing $k = 0$ reaches 1.6595, the value obtained previously using Eq. 4.14. Second, the new method is capable of producing higher maximal bit generation rates than the old one ($G_{R2} > 1$) seemingly at all combinations of input parameters if m is set to 2. The gain is practically independent of the measurement clock period, but its k and $\Delta\tau/\tau$ dependencies are noteworthy. For small values of k (0, 1 or 2), G_{R2} decreases rapidly for increasing fractional dead time; if k is higher than that, the surfaces become more and more flat, indicating that the fractional part has a decreasing effect on the gain.

Moreover, a larger k generally implies a smaller G_{R2} the compared values k_1, k_2 are greater than 3. The gain is theoretically maximized and approaches 2 if $k = 1$ or 2 and $\Delta\tau/\tau$ vanishes.

A more general gain definition can also be found to account for any possible value of m. Since λ_{out} is a common factor of all bit generation rate equations, G_{Rm} can be expressed by multiplying G_{R2} with a ratio of m-dependent efficiencies:

$$G_{\mathrm{R}m}\left(\tau,\tau_{\mathrm{d}}\right)=\frac{R_m\left(\tau,\tau_{\mathrm{d}}\right)}{\max_\lambda\{R'\left(\lambda,\tau,\tau_{\mathrm{d}}\right)\}}=\frac{R_m\left(\tau,\tau_{\mathrm{d}}\right)}{R_2\left(\tau,\tau_{\mathrm{d}}\right)}\cdot G_{\mathrm{R2}}\left(\tau,\tau_{\mathrm{d}}\right)=\frac{\eta_{\mathrm{R}}\left(m\right)}{\eta_{\mathrm{R}}\left(2\right)}\cdot G_{\mathrm{R2}}\left(\tau,\tau_{\mathrm{d}}\right). \tag{4.15}$$

Thus, the surfaces are simply rescaled versions of those in Fig. 4.5. Specifically, $G_{\mathrm{R7}}\approx 1.1037\cdot G_{\mathrm{R2}}$: increasing m from 2 to 7 boosts the yield by 10.37%.
A third way to compare the schemes is looking at their peak performances on a hardware with given capabilities—time measurement resolution, dead time and maximum allowed λ_{out}. This, however, is difficult to generalize; for the particular system I used during my work, the reader may find the comparison results in Sec. 4.4.

4.3 Error Sensitivity

There is a practical difficulty compared to the original method described in Chapter 3: in theory, the Poisson point process governing the bit generation should truly be homogeneous—λ needs to be a constant. Slow changes in the optical power were not of concern in the old scheme, since the self-differentiating method made sure that their effects were eliminated. The new method, on the other hand, requires either very precise power control, or the tuning of the measurement clock period to keep P_{eq} at 1/3.
Even if these are both implemented, the distributions of W_i and V_s will deviate from uniform, since it is impossible to have infinite numerical precision in the software governing the settings. In this section, I devise a realistic one-parameter error model describing the deviations from ideality, obtain bounds which the error parameter should not exceed, and show that even in the case of errors, it is possible to assign bits to outcomes so that the bias remains zero.

4.3.1 Error Model

The error model assumes that the probability of equality $P_{\mathrm{eq}}=P\left[W_i=0\right]$ deviates from the idealistic value 1/3 as described by an error parameter $\epsilon=P_{\mathrm{eq}}-1/3$ Note that $\epsilon\in\left[-1/3;2/3\right]$. Due to the fact that Y_i follows a symmetric probability distribution, $P\left[W_i=-1\right]=P\left[W_i=1\right]$—the same phenomenon that was exploited by the old method. Normalization constraints require that the PMF is now of the

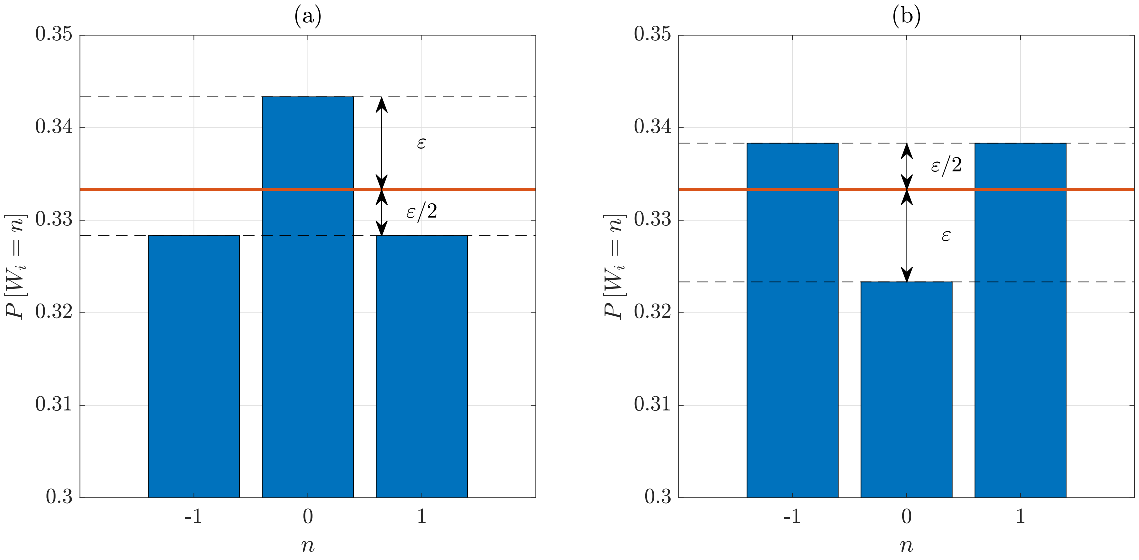

Figure 4.6: Error model for the probability mass function of W_i for (a) positive and (b) negative ϵ.

form (see Fig. 4.6)

$$P\left[W_i = n\right] \begin{cases} \frac{1}{3} + \epsilon, & \text{for } n = 0, \\ \frac{1}{3} - \frac{\epsilon}{2}, & \text{for } n = \pm 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.16}$$

Obviously, the vector-valued variables V_s will neither be uniformly distributed if $\epsilon \neq 0$. The probabilities of V_s (joint probabilities of $W_{ms}, \ldots, W_{ms-(m-1)}$)

$$P\left[V_s = (n_0, n_1, \ldots, n_{m-1})\right] = \prod_{t=0}^{m-1} P\left[W_{ms-t} = n_t\right] \tag{4.17}$$

only depend on the number of equal comparisons (c) involved. Note that the independence of W_i variables was exploited in the calculations once again. As c ranges from 0 to m, the 3^m different joint probabilities may only have one of $m+1$ distinct values, namely

$$p_{c,m}\left(\epsilon\right) = \left(\frac{1}{3} + \epsilon\right)^c \left(\frac{1}{3} - \frac{\epsilon}{2}\right)^{m-c}, \; c = 0, 1, \ldots m, \tag{4.18}$$

which greatly reduces the complexity of the error analysis. In the following discussion, ϵ is assumed to be constant. Although its value fluctuates along with the optical power fluctuations, we can once again use the argument that the time scale of bit generation is significantly shorter as that of thermal effects. Therefore, ϵ can be assumed constant over the generation of e.g. hundreds of thousands of bits.

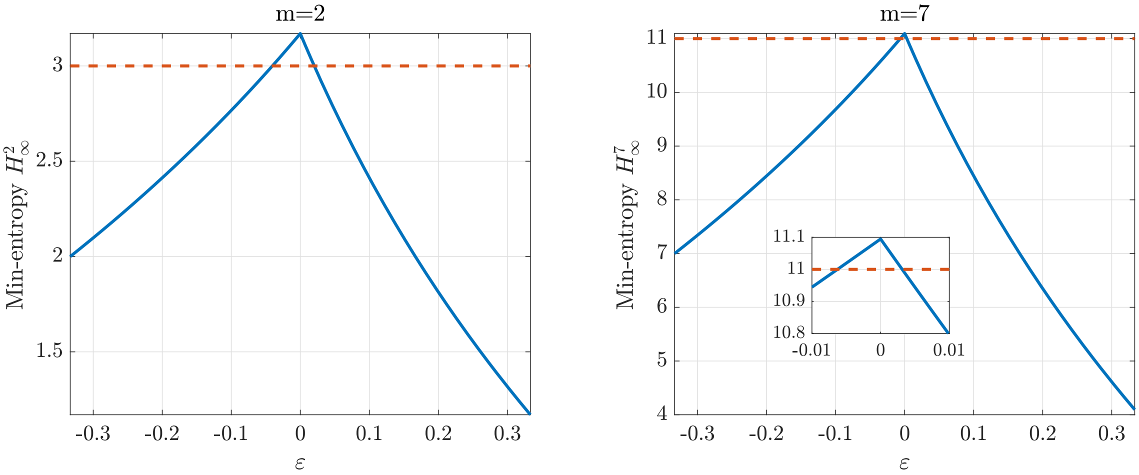

Figure 4.7: Min-entropy of the distorted distribution as a function of ϵ for (a) $m = 2$ and (b) $m = 7$. Red lines show the number of bits assigned by the method for the specified group length. Taken from Ref. [5].

4.3.2 Min-Entropy of the Distorted Distribution

There is an important metric quantifying the deviations from ideality, which can be readily calculated from the obtained probabilities: the min-entropy. As it is known, it only depends on the highest probability value from the PMF. It can be seen that this maximum only depends on the sign of the error parameter for the previously derived distribution.

$$p_{\max,m}(\epsilon) = \max_c p_{c,m}(\epsilon) = \begin{cases} p_{0,m}, & \text{for } \epsilon \leq 0 \\ p_{m,m}, & \text{for } \epsilon \geq 0 \end{cases} \tag{4.19}$$

Correspondingly, the min-entropy can be summarized in terms of ϵ:

$$H_\infty^m(\epsilon) = \begin{cases} m \cdot \log_2(3) - m \cdot \log_2(1+3\epsilon), & \text{for } \epsilon < 0 \\ m \cdot \log_2(3), & \text{for } \epsilon = 0 \\ m \cdot \log_2(3) - m \cdot \log_2(1-1.5\epsilon), & \text{for } \epsilon > 0. \end{cases} \tag{4.20}$$

The min-entropy is smaller than its error-free value for all non-zero ϵ; the rate of decrease is, however, slightly more significant if $\epsilon > 0$ (equalities are favored). If $H_\infty^m(\epsilon)$ becomes smaller than $\lfloor m \cdot \log_2(3) \rfloor$, we cannot assign as many bits as possible to the outcomes. Therefore, ϵ should always stay within a range so that the min-entropy condition $H_\infty^m(\epsilon) \geq \lfloor m \cdot \log_2(3) \rfloor$ is satisfied. This provides us an interval of error parameters, outside of which the method fails theoretically.

Figure 4.7 shows the min-entropy as a function of the error parameter for $m = 2$ and 7. It can be seen that for smaller values of m, the acceptable interval is wider. We cannot, however, conclude that choosing a higher m results in a more sensitive bit generation scheme if the error is within the respective interval—on the contrary, as it will be shown in Section 4.4. One could even argue that the error is dispersed more evenly between a higher number of possible outcomes. The respective intervals for $m = 2$ and 7 are approximately

$$-0.0404 \leq \epsilon_{(2)} \leq 0.0202 \quad \text{and} \tag{4.21}$$

$$-0.00628 \leq \epsilon_{(7)} \leq 0.00314. \tag{4.22}$$

The min-entropy condition in itself does not prove that the method operates free of errors. The resulting bit sequence will still deviate from what we expect from a perfect random number generator. The extent of deviations depend on the magnitude of ϵ, and also on how bits are assigned to different outcomes. Fortunately, the worst problems can be avoided by carefully designing the bit assignment function.

4.3.3 Bias Elimination

Arguably, the greatest flaw of any QRNG is a non-zero bias, when the relative frequencies of zeros and ones differ from each other. Its presence also foreshadows other related statistical problems of the generated bit sequences. As an example, a QRNG that fails the bias related *Frequency* test of the NIST STS will definitely fail the *Runs* test as well—so much so that passing the former is the prerequisite of even running the latter [65].

However, for the error model discussed hereby, we will show that the bias can be systematically eliminated by utilizing the underlying symmetries of the resulting distribution, irrespectively of the value of ϵ. First of all, let us find a natural ordering for the outcomes $V_s = \left(W_{ms}, W_{ms-1}, \ldots W_{ms-(m-1)}\right)$. By increasing each element of the vector-valued variable by one, we get $U_s = \left(W_{ms}+1, W_{ms-1}+1, \ldots W_{ms-(m-1)}+1\right)$. Now, treat U_s as a number in its ternary representation, $W_{ms}+1$ signifying the most significant ternary digit (MST) and $W_{ms-(m-1)}$ being the least significant one (LST). U_s can take on any values between $00\ldots0_2 = 0_\mathrm{d}$ and $22\ldots2_2 = (3^m-1)_\mathrm{d}$, where d denotes the decimal numeral system. The ternary values already imply a "natural" ordering of outcomes. Using this ordering, it is trivial to see that the distribution of U_s (and V_s) will be symmetric around the mid-point represented by the value $(3^m-1)/2$—since the probabilities only depend on the number of ones among the ternary digits, which represent comparisons with equality. See Fig. 4.8 as an example if m is chosen to be 2.

As a shorthand notation, index all outcomes with the decimal representation a of their ternary values, $a \in \mathcal{A}_m = \{\, z \in \mathbb{Z} \mid 0 \leq z < 3^m \,\}$. Denote the probability of

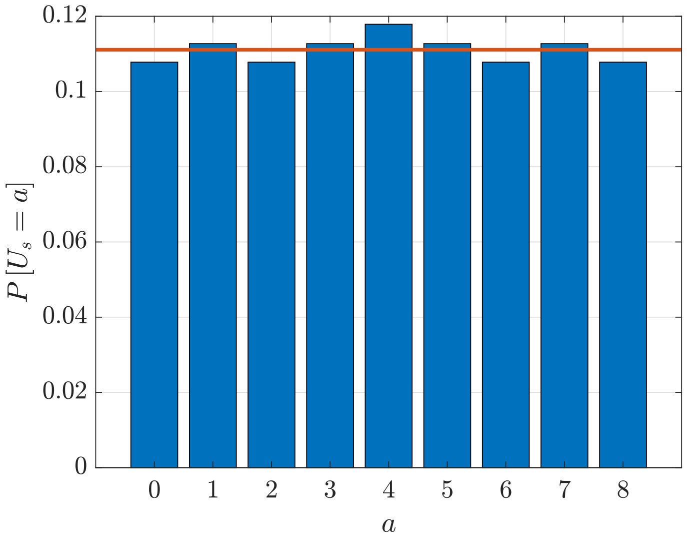

Figure 4.8: Distribution of U_s for $\epsilon = 0.1$ and $m = 2$. The red line denotes $1/9$, the error-free probability of all outcomes.

outcome a with $p(a)$. Define another set, $\mathcal{B}_m \subset \mathcal{A}_m$ that contains $a \in \mathcal{A}$ if and only if a is the index corresponding to a non-discarded outcome. This set has a cardinality of $|\mathcal{B}_m| = 2^{\lfloor m \cdot \log_2(3) \rfloor}$.

To understand how the problem of bit assignment can be approached, we need the notion of the Hamming weight w of bit sequences. w denotes the number of non-zero elements—ones in the binary case—in a sequence. Define w_a as *the Hamming weight of the bit group assigned to outcome a*. For discarded outcomes, w_a shall remain undefined. The amount of bias can be described in terms of the probability-averaged Hamming weight over all non-discarded events, defined as

$$\overline{w}\left(m\right) = \frac{\sum_{a \in \mathcal{B}_m} p\left(a\right) \cdot w_a}{\sum_{a \in \mathcal{B}_m} p\left(a\right)}. \tag{4.23}$$

Here the denominator is a factor of re-normalization, which deletes the effect of discarded outcomes from the average. In bias-free circumstances, on average, one out of two bits is a “1”. Translating this to bit groups assigned to outcomes, the bias-free, ideal probability-averaged Hamming weight of groups should be

$$\overline{w}^{\text{id}}\left(m\right) = \frac{\lfloor m \cdot \log_2\left(3\right) \rfloor}{2}. \tag{4.24}$$

Every particular bijective assignment of bit groups to non-discarded outcomes can be called a *coding function* $C : a \in \mathcal{B}_m \to \{0, 1\}^{\lfloor m \cdot \log_2(3) \rfloor}$. It is possible to systematically eliminate bias in the given random number generation method if there exists such a set $\mathcal{B}_m \subset \mathcal{A}_m$ for which it is possible to construct a coding function C so that $\overline{w}\left(m\right) = \overline{w}^{\text{id}}\left(m\right)$ independently of ϵ. Such a coding function is then called a *bias-free coding on $\mathcal{B}_m$*.

It can be shown that such coding functions do indeed exist for the underlying symmetric distributions. A general, sufficient but not necessary rule of construction is to assign complementary bit groups (those with maximal Hamming distance) to two equiprobable outcomes. An example for $m = 2$ is shown in Table 4.1, along with the relevant parameters mentioned in the discussion. The probability-averaged Hamming weight provided by this coding is

$$\overline{w}(2) = \frac{6 \cdot p_{0,2} + 6 \cdot p_{1,2}}{1 - p_{2,2}} = \frac{6 \cdot (p_{0,2} + p_{1,2})}{4 \cdot (p_{0,2} + p_{1,2})} = \frac{3}{2} = \overline{w}^{\mathrm{id}}(2)\,, \tag{4.25}$$

proving it is free from bias.

Table 4.1: A particular bias-free coding function for $m = 2$. The outcome $a = 4$ is discarded.

$\mathcal{B}_2 = \{0, 1, 2, 3, 5, 6, 7, 8\}$

a	W_{2i}	W_{2i-1}	$p(a) = P[W_{2i}, W_{2i-1}]$	Bits	Weight w_a
0	-1	-1	$p_{0,2}$	000	0
1	-1	0	$p_{1,2}$	001	1
2	-1	1	$p_{0,2}$	010	1
3	0	-1	$p_{1,2}$	011	2
4	0	0	$p_{2,2}$	-	-
5	0	1	$p_{1,2}$	100	1
6	1	-1	$p_{0,2}$	101	2
7	1	0	$p_{1,2}$	110	2
8	1	1	$p_{0,2}$	111	3

The presented coding method can be algorithmically extended to arbitrary values of m. The main idea is the following. The only outcome without an equiprobable "pair" is always the one in the middle of the ordering, described by $U_s = 11\ldots1_3$ or $a = (3^m - 1)/2$. Therefore, this must definitely be discarded. Now, choose pairs of equiprobable outcomes that are placed symmetrically around this mid-point, and discard them. Continue this until you reach the desired amount of non-discarded events. After that, assign the bit groups from $00\ldots0_2$ to $11\ldots1_2$ to the remaining outcomes in increasing order. See Algorithm 1 for details (here the discarded events are directly placed around the mid-point).

Algorithm 1: Bias-free coding generalized for any value of *m*

```
Data: $m$-long array of comparisons W
outcomes:= $3^m$ ;                                  // Total number of outcomes
kept:= $2^{\lfloor m \cdot \log_2(3) \rfloor}$ ;                // Number of not discarded outcomes
discarded:=outcomes−kept ;                        // Number of discarded outcomes
lower_limit := $0.5 \cdot \big[\text{outcomes} - 1 - (\text{outcomes} - \text{discarded} - 1)\big]$;
upper_limit := $0.5 \cdot \big[\text{outcomes} - 1 + (\text{outcomes} - \text{discarded} - 1)\big]$;
for $i = 1$ to $m$ do
    W(i)++ ;                                      // Forming ternary arrays
decimal_value=ternary_to_decimal(W) ;             // Ternary to decimal conversion
if decimal_value<lower_limit then
    bits=decimal_to_binary(decimal_value) ;       // Return the binary value of the ternary array left-padded with zeros
else
    if decimal_value>upper_limit then
        bits=decimal_to_binary(decimal_value)− $\left(3^m - 2^{\lfloor m \cdot \log_2(3) \rfloor}\right)$ ;
        // Return a modified binary value to account for discarded outcomes
    else
        bits=[] ;                                 // Do not return bits
```

A bias-free coding cannot and will not fix all statistical problems that are due to a non-zero ϵ. It actually favors some bit groups over others; for the example given in Table 4.1, negative ϵ values will show preference towards the patterns 000, 010, 101 and 111, whereas positive ϵ values will increase the relative frequencies of 001, 011, 100 and 110. However, since all possible coding functions exhibit some kind of deference under non-uniform distributions, a bias-free choice is always better than others.

4.4 Experimental Results and Applicability

Theoretical results suggest that the newly formed method is capable of increasing the bit generation efficiency and rate of the simple time-of-arrival QRNG scheme, while the requirements toward physical devices change very little. In this section, it will be shown that this can be achieved in practice as well. By monitoring the changes of the photon rate and slightly modifying the parameters to reach a quasi-uniform distribution, the new scheme is capable to generate high-quality random bit sequences, which pass all the NIST tests without any post-processing.

4.4.1 The Experimental Setup

The setup of physical hardware is only slightly modified from that used for the original method (Fig. 3.15). The monitor point is pulled before the first attenuator, and the linear photodetector's output voltage is read every second by a Matlab script realizing a PI controller. The controller tunes the laser's current to maintain a constant reading value specified by a script parameter. Long-term measurements have been conducted on the monitor detector's output voltage, suggesting that once steady-state is reached, its value stays within $\pm 0.185\%$ of the target voltage 99.1% of the time. If this stability could be achieved at the detector's input, the corresponding error parameter would be smaller in absolute value than $\approx 5.7 \cdot 10^{-4}$, within the min-entropy condition's acceptable interval for both $m = 2$ and 7.

The PMT's input photon rate, however, exhibits higher degrees of fluctuation. Although it is quite stable over the duration of several millions of detections, slow drifts can be observed on longer time scales, presumably due to the temperature dependence of the VOAs placed after the monitor point. This effect can only be cancelled indirectly, since it is impossible to efficiently control the power if monitoring is done either between the two attenuators or just before the PMT. The intensity levels at those points are at best similar to—but, especially at the input, significantly lower than—the noise level of linear photodetectors.

Let us revisit some important device parameters from Section 3.4. The PMT has a maximum allowed output count rate of $5 \cdot 10^6$ [1/s], corresponding to a slightly higher input photon rate of $5.05 \cdot 10^6$ [1/s] through the system dead time that is around 2 ns. The time-to-digital converter operates with a best-case resolution of $\tau_\mathrm{b} = 250\,\mathrm{ps}$, its clock signal is continuous, cannot be restarted at every detection. As generally a higher λ requires a smaller τ (higher clock frequency) to set $P_\mathrm{eq} = 1/3$, the fastest possible clock signal we may require using this detection system is that belonging to $\lambda_\mathrm{out,max}$: $138.74\,\mathrm{ns} \approx 555\tau_\mathrm{b}$. Every potential value of τ is therefore at least two and a half orders of magnitude larger than the base time measurement resolution.

4.4.2 Setting and Maintaining the Probability of Equalities

Decreasing the resolution of the TDC card is not an appropriate way of setting τ to its desired value, since all analysis was conducted on a system with a restartable clock. On the other hand, we can once again utilize the idea from Section 3.4.2: we

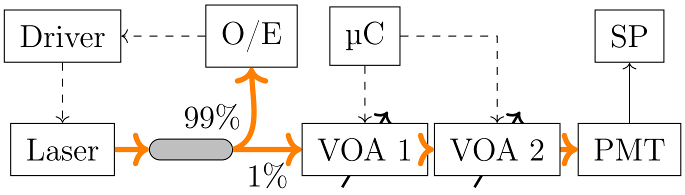

Figure 4.9: Experimental setup of the refined QRNG scheme. Driver: laser driver with a built-in PI controller; O/E: photodiode; VOA: variable optical attenuator; µC: microcontroller; PMT: photomultiplier tube; SP: signal processing. Taken from [5].

can define a restartable clock of period τ in software and apply it to a data set measured with the best possible resolution. If $\tau \gg \tau_\mathrm{b}$ holds, the fact that it is rather a re-discretization and not a discretization of truly analog values, will not become noticeable. This solution offers a second benefit as well: there is no need to control the optical power to perfectly fit a preset measurement resolution—which would be a futile endeavor. Rather, we can periodically estimate λ by counting the time it takes to reach a certain number of detections to account for the intensity drifts, and *tune the software-based clock* to obtain the expected distribution. The estimation period of λ (N_λ detections) should be short enough that the Poisson point process can be assumed homogeneous over its duration, but long enough to keep the estimation precise.

I recorded six different data sets, each at least ten minutes long, at four different input photon rates. The first three used λ values around 1.35, 2.55 and $3.63 \cdot 10^6/\mathrm{s}$, respectively. The last three data sets were measured with the same settings (the highest λ, approximately $4.8 \cdot 10^6/\mathrm{s}$), but due to power fluctuations, they are not of identical quality. Since $\tau_\mathrm{d} \ll \tau$ for each case, the ratio $\Delta\tau/\tau = \tau_\mathrm{d}/\tau$ is close to zero. For this reason, the first idea was to simply update τ using the inversely proportional relationship in Eq. 4.7. This would require less computing power than solving a complex equation for the "true" clock period, but it is not immediately obvious how much it affects the quality of randomness. Altogether, three different clock assignment methods were tried, as listed below.

1. **Simple:** $\tau^\mathrm{s} = \ln(2)/\lambda$.
2. **True:** τ^t obtained from numerically solving $P_\mathrm{eq} = 1/3$.
3. **Heuristic:** $\tau^\mathrm{h} = \left(\tau^\mathrm{s} + \tau^\mathrm{t}\right)/2$.

The heuristic approach was a result of slight initial failures in tests for bit sequences generated using both options 1 and 2, as it will be discussed in the following section.

4.4.3 Results and Statistical Testing

Finding the most suitable estimation period and clock assignment method proved to be a gradual process, in which one of the variables was altered, twelve bit sequences—six for both $m = 2$ and 7—were generated and ultimately tested, until an acceptable result was achieved. Moreover, using the distribution of W_i, a best-fit error parameter ϵ was calculated for each period. As a starting point, τ was chosen to be updated every $N_\lambda = 10^7$ detections, corresponding to approximately 2.04–7.41 seconds for the given range of input photon counts. First, the simple τ assignment was tried for both values of m. This solution already produced satisfactory results unless the photon rate was increased slightly below its allowed maximum and m was chosen to be 2.

Table 4.2: Number of successful tests out of 188 for a photon rate update period of 10^7 detections with different τ-designation methods.

Mean λ [10^6/s]	**Simple**		**Heuristic**		**True**	
	m = 2	**m = 7**	**m = 2**	**m = 7**	**m = 2**	**m = 7**
1.35	188	188	188	188	187	187
2.55	188	188	187	188	187	188
3.63	188	188	188	186	188	188
4.75	177	187	187	185	184	188
4.78	174	188	187	188	188	187
4.91	175	188	184	188	187	187

Table 4.2 concludes the number of successful tests for the different combinations of λ, m and τ assignment. In most cases, the number of failed subtests is between 1 and 4, apart from the last three instances in the first column. On average, bits generated using $m = 7$ performed slightly better compared to those with $m = 2$; this difference is the most pronounced in case of $\tau = \tau^{\mathrm{s}}$.

The reported ϵ values were always small enough to stay within the limits provided by both Eq. 4.21 and 4.22, keeping the min-entropy condition satisfied for $m = 2$ and 7 as well. Moreover, the goodness of fit was always close to one, indicating that my error model correctly describes the deforming distribution. Generally, the following tendencies can be seen (see Fig. 4.10). First, for a given clock assignment method, the mean of ϵ decreases (tends towards negative values), while its variance increases with increasing λ. Second, as the simple and heuristic methods assign a τ value that slightly decrease the probability of equality, these solutions cause ϵ to almost always

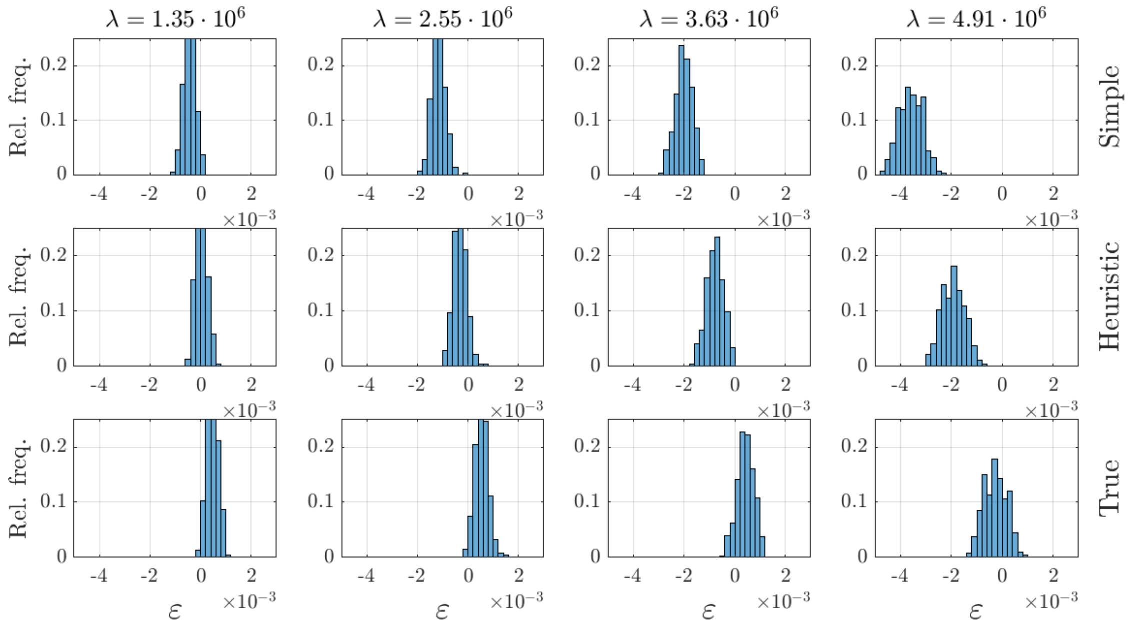

Figure 4.10: Histograms of best-fit ϵ values for different clock assignment methods (rows) and photon rates (columns) for an update period $N_\lambda = 10^7$.

remain negative. Both effects are the most pronounced in case of a simple clock assignment.

For smaller values of λ, the true assignment favors almost exclusively positive ϵ. Even in this theoretically best scenario, about half of the generated bit sequences failed a small number of tests. It was seen in Section 4.3 that the method is more sensitive if $\epsilon > 0$ with regard to its min-entropy. Since it can also be true with regard to other issues, the heuristic method was meant to provide an intermediate solution offering slightly negative ϵ. Ultimately, this idea did not solve the problem, with the reported success ratio decreasing somewhat.

Rather than twisting the way of calculating the clock period any further, I decided to try different values of N_λ. A slightly longer ($2 \cdot 10^7$) update period produced similar results, a significantly shorter one (10^6) degraded the randomness quality noticeably, but choosing $N_\lambda = 5 \cdot 10^6$ was really promising. The simple τ assignment was already abandoned in this scenario, leaving only the heuristic and true methods. The test results (Table 4.3) show a general improvement over previous values of N_λ even in the heuristic case. Moreover, with a true τ and $m = 2$, three of the six bit sequences failed only one of the Non-Overlapping Template tests (a separate one in each case) on proportionality. The respective passing ratios of 979, 979 and 1000 fell just outside the acceptable interval 980–999. The three remaining sequences did, however, pass all tests. The best results were obtained for a choice of true τ and $m = 7$: none of the six sequences showed any noticeable deviations from a uniform distribution.

Table 4.3: Number of successful tests out of 188 for a photon rate update period of $5 \cdot 10^6$ detections with different τ-designation methods.

Mean λ [10^6/s]	**Heuristic**		**True**	
	m = 2	**m = 7**	**m = 2**	**m = 7**
1.35	188	187	187	188
2.55	188	188	188	188
3.63	187	188	188	188
4.75	188	188	187	188
4.78	186	186	188	188
4.91	188	187	187	188

By changing the update length, the ϵ values remained at the same order of magnitude, and also exhibited the same tendencies as before (Fig. 4.11). However, the variance increased, shown by the fact that histograms span across wider intervals.

A final performance comparison needs to be made between the two schemes on the same physical hardware. Assuming that $m = 7$, the maximal bit generation rate on the current setup is

$$\lambda_{\text{out,max}} \cdot \eta_{\text{R}}\,(7) \approx 3.679\,\text{Mbps}. \tag{4.26}$$

This is 47.25% greater than the 2.5 Mbps achievable with the old method (Section 3.4.3). Note also that the old method failed some tests, most notably the Runs test when operating near the maximal bit generation rate, due to the continuous clock inducing a positive correlation between neighbouring bits. However, the new design passed everything in the vicinity of said maximum.

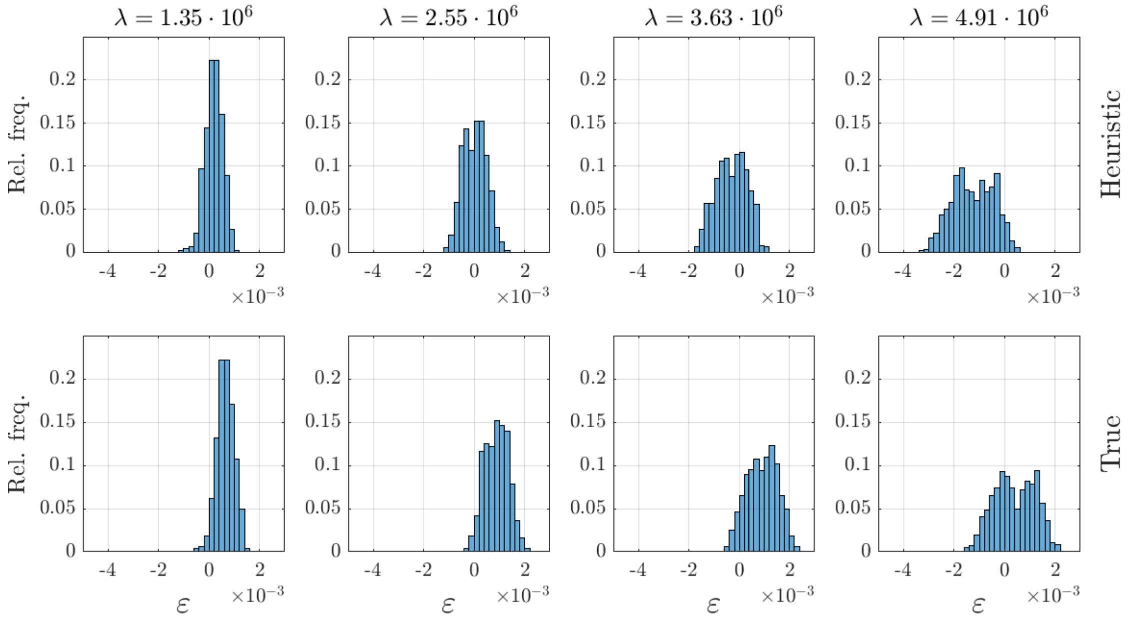

Figure 4.11: Histograms of best-fit ϵ values for different clock assignment methods (rows) and photon rates (columns) for an update period $N_\lambda = 5 \cdot 10^6$.

4.5 Conclusion

The new bit generation method proposed in this chapter proved to be better than the one discussed earlier. By increasing the efficiency through maintaining an almost-uniform distribution and careful grouping of its outcomes, less events need to be discarded, leading to an overall gain in terms of the bit generation rate. Although the uniformity demands precise power control and updating a software-based clock signal's period to counteract drifts in optical intensity, this is a minor inconvenience in light of the advantages. The effects of deviations from uniformity (a non-zero ϵ) have been calculated in terms of the min-entropy. This chapter forms the basis of Thesis III.

Further research could provide more insight into how this error influences other aspects of randomness for the generated bit sequences, and how the choice of group length m relates to sensitivity towards errors. Moreover, the basic idea behind the new scheme can be generalized as well. Forming groups of outcomes from uniform distributions, for which the cardinality of the sample space is not a power of 2, may increase the bit generation efficiency for good choices of group length. How this applies for cardinalities other than 3, the case which was discussed beforehand, is also worthy of inspection.

Chapter 5

Novel Scientific Results - Summary of Theses

Thesis I.

I proposed a specific solution to use vertical cavity surface-emitting lasers in the BB84 DV-QKD protocol, which are responsible for sending two of the four states. This could lead to the potential cost and size reduction of the transmitter circuitry. I also analyzed other degrees of freedom of quantum states, which provide opportunities to eavesdrop without getting noticed. I suggested countermeasures against these loopholes based on realistic arguments, which take into account the practical realization constraints.

I.a) I introduced a new transmitter design for the BB84 protocol, which only uses two VCSELs instead of the four light sources found in the trivial design. To obtain all four states of BB84, the VCSELs' polarization switching mechanism is exploited. This switching happens between two orthogonal eigenstates; on-demand switching is called polarization modulation. General benefits of VCSELs in low-power applications also make this solution preferable over edge-emitting lasers.

I.b) I proposed two different current-induced polarization modulation scenarios. The first biases the VCSEL near threshold, and applies different amplitude pulses to obtain different polarizations; the second sets the bias near the polarization switching point and applies small-signal modulation around the bias. I showed that two problems can be solved by inserting an electro-absorption modulator into the light's path. First, attenuation can be changed quickly to account for the inherently different power levels between the obtained eigenstates. Second, the modulator is also suitable for pulse-shaping, which cancels incorrectly polarized parts of the pulse and creates an arbitrarily precise temporal overlap between differently polarized signals.

I.c) I highlighted the fact that the new transmitter design is susceptible to a spectral attack, where the eavesdropper measures the frequency of photons to distinguish between differently polarized quantum states. In the trivial design, this could be overcome by using four lasers with overlapping spectra; however, VCSEL polarization eigenmodes are always separated in frequency. I proposed a method that offers protection even if Eve can measure frequency without destroying photons. First, choose two VCSELs, for which the spectra of lower and higher frequency eigenstates are pairwise largely overlapping. Then assign bits to frequencies in a complementary way. This renders bits and frequencies, and also bases and frequencies, uncorrelated. The eavesdropper is thus unable to gain information about the polarization by the spectral degree of freedom—not more then she could if a trivial transmitter were in use.

Related own publications: **C6**, **C7**, **B1**

Thesis II.

I derived the mathematical model of a quantum random number generation method. It is based on comparing the measured lengths of two successive time intervals between photon detections, and assigning bits based on the sign of the difference, discarding equal cases. I focused mainly on how two figures of merit—the bit generation efficiency and the bit generation rate—change as functions of the relevant parameters. I simulated the method, also incorporating deviations from the model found in the physical setup, before evaluating the model's validity through experiments. The theoretically derived and measured results show an excellent agreement.

II.a) I determined the decisive parameters, with which the method can be modelled faithfully, and discarded those which are negligible in my physical implementation of the generator. I derived formulae for the bit generation rate and efficiency analytically, based on probability theoretical arguments, as a function of the parameters: the input photon rate λ, the time measurement precision τ and the detection system's dead time τ_d. Fixing the latter two, and changing only λ, I showed that it is not possible to maximize both figures simultaneously, and the maximal bit generation rate corresponds to a lower-than-ideal efficiency.

II.b) I showed that the efficiency only depends on the fractional part of the ratio τ_d/τ, and is invariant under changing the ratio's integer part, greatly simplifying the analysis, providing solutions for all possible parameter combinations. The bit generation rate, however, is a function of the whole length of the dead time, but it arises readily from the efficiency formula, through a multiplication by the output photon rate of the detector.

II.c) I introduced the concept of the high-precision regime, when $\lambda\tau \ll 1$, and argued that in the HPR, the negative effects of a continuous clock are almost negligible compared to a restartable clock. I supported this claim by further simulations, showing that the correlations between bits are kept low in the HPR, while the bit generation efficiency and rate agree with those calculated in the model.

II.d) I conducted experiments and verified the validity of my mathematical model. I showed that the measured bit generation efficiency and rate, as a function of the input photon rate for fixed values of τ and τ_d, are in a remarkable agreement with the theoretical predictions. The effects of device limitations—continuous time measurement clock, non-constant dead time—are small enough so that the figures of merit do not deviate from the derived values. Furthermore, I tested the quality of randomness of the generated sequences, showing that in the HPR, a continuous clock does not compromise the uniformity of generated bits.

Related own publications: **J1**

Thesis III.

I created a refined version of the random number generation method known from Thesis II., which increases both the bit generation efficiency and the bit generation rate by forming groups of m comparison signs and assigning multiple bits to each group. I compared the refined method with the old one quantitatively, and obtained formulae for the bit generation rate gain. I also took into account the effects of non-uniformity due to small but non-vanishing light intensity fluctuations and analyzed how those affect the quality of randomness. Finally, I verified experimentally that the proposed method is indeed capable of creating random numbers, which pass all NIST statistical tests without post-processing.

III.a) I proposed to change the previous bit generation method such that the signs of comparisons are uniformly distributed. I derived that it is only possible for certain combinations of input photon rate λ, time measurement precision τ and dead time τ_d. I showed that this alone is not suitable for extracting more bits than one per comparison, since the floor function applied to the min-entropy is still one. This limitation is always present if the cardinality of the sample space is not a power of two. However, one can form m-long groups of successive comparisons, for which the difference between the min-entropy of the new vector variables and its integer part is smaller. Therefore, higher bit generation efficiencies are feasible. I showed that values of $m = 2$ and 7 are corresponding to higher efficiencies, while the resulting space of outcomes is still kept relatively small.

III.b) I quantified the improvements over the old method in several different ways: comparing the bit generation rates in the settings tailored for the refined method, and introducing the bit generation rate gain $G_{\mathrm{R}m}$ as the ratio of the maximum/optimal rates of the new and old methods given a fixed dead time and time measurement precision. Depending on the how long τ_{d} is compared to τ, the gain varies significantly between 1 and 2, showing that the new method is always capable of generating bits faster than the old one.

III.c) I created an error model for the random number generation method, where the distribution of comparison signs is not uniform, but symmetric. I derived the maximum tolerable error limits within which the min-entropy exceeds the number of bits assigned to each group. I showed that bias can be systematically eliminated under all conditions by utilizing the symmetries by choosing carefully which outcomes to discard, and by assigning bit groups of maximal Hamming distance to equiprobable outcomes. I also provided a general algorithm—valid for any value of m—that realizes such a bias-free coding function.

III.d) I conducted experiments, and confirmed that the proposed method is capable of generating high-quality random numbers even with practical limitations: slight photon rate fluctuations and a continuous measurement clock signal. I overcame the problem of slow power drifts by tuning τ in software accordingly. I showed that the error model gives a good description of real-life deviations from uniformity, and the error magnitude can be kept within the allowed range. I demonstrated that even with non-idealities, it is possible to fine-tune the settings so the random bits generated by the new method do not need post-processing for passing all randomness tests. The generation rate on the given hardware showed a 47.25% increase compared to the old method.

Related own publications: **J5**

Bibliography

[1] Ágoston Schranz and Eszter Udvary. “Mathematical analysis of a quantum random number generator based on the time difference between photon detections”. In: *Optical Engineering* 59.4 (2020), p. 044104. DOI: 10.1117/1.OE.59.4.044104.

[2] Ádám Marosits, Ágoston Schranz, and Eszter Udvary. “Amplified spontaneous emission based quantum random number generator”. In: *Infocommunications Journal* 12.2 (2020), pp. 12–17. DOI: 10.36244/icj.2020.2.2.

[3] Ágoston Schranz and Eszter Udvary. “Error probability in polarization sensitive communication systems in terms of moments of the channel’s rotation angle”. In: *Optical and Quantum Electronics* 53.1 (Jan. 2021), p. 62. ISSN: 0306-8919. DOI: 10.1007/s11082-020-02690-1.

[4] Balázs Matolcsy, Eszter Udvary, and Ágoston Schranz. “Common-mode noise filtering with space-divided differential 2x2 VLC for V2V applications”. In: *Optical and Quantum Electronics* 53.4 (2021), p. 182. DOI: 10.1007/s11082-021-02808-z.

[5] Ágoston Schranz, Eszter Udvary, and Balázs Matolcsy. “Efficiency Improvement of a Time-of-Arrival Quantum Random Number Generator”. In: *Optical Engineering* 60.3 (2021), p. 034112. DOI: 10.1117/1.OE.60.3.034112.

[6] Ágoston Schranz, Eszter Udvary, and Zsolt Kis. “Photon statistics determination for single photon based quantum key distribution”. In: *18th International Conference on Transparent Optical Networks (ICTON)*. IEEE. 2016, pp. 1–4. DOI: 10.1109/ICTON.2016.7550483.

[7] Eszter Udvary, Ágoston Schranz, and Balázs Matolcsy. “Dispersion and off-set filtering in RSOA based networks”. In: *18th International Conference on Transparent Optical Networks (ICTON)*. 2016, pp. 1–4. DOI: 10.1109/ICTON.2016.7550690.

[8] Ágoston Schranz. "Experimental Investigation of VCSEL for Quantum Communications". In: *Mesterpróba 2016.* 2016, pp. 8–11.

[9] Ágoston Schranz. "Investigation of VCSEL Polarization for Quantum Key Distribution". In: *International Interdisciplinary PhD Workshop 2016.* 2016, pp. 117–120.

[10] Gábor Szabó, Ágoston Schranz, and Eszter Udvary. "Nonlinear Modulation Characteristics of LEDs in Radio on Visible Light Systems". In: *International Interdisciplinary PhD Workshop 2016.* 2016, pp. 6–10.

[11] Ágoston Schranz and Eszter Udvary. "Transmitter Design Proposal for the BB84 Quantum Key Distribution Protocol using Polarization Modulated Vertical Cavity Surface-emitting Lasers". In: *Proceedings of the 6^{th} International Conference on Photonics, Optics and Laser Technology.* INSTICC. SciTePress, 2018, pp. 252–258. ISBN: 978-9-897-58286-8. DOI: 10.5220/0006638002520258.

[12] Ágoston Schranz and Eszter Udvary. "Quantum Bit Error Rate Analysis of the Polarization based BB84 Protocol in the Presence of Channel Errors". In: *Proceedings of the 7th International Conference on Photonics, Optics and Laser Technology - Volume 1: PHOTOPTICS.* INSTICC. SciTePress, 2019, pp. 181–189. ISBN: 978-9-897-58364-3. DOI: 10.5220/0007384101810189.

[13] Ágoston Schranz, Ádám Marosits, and Eszter Udvary. "Effects of Sampling Rate on Amplified Spontaneous Emission Based Single-Bit Quantum Random Number Generation". In: *21st International Conference on Transparent Optical Networks (ICTON).* IEEE. 2019, pp. 1–4. DOI: 10.1109/ICTON.2019.8840188.

[14] Ágoston Schranz and Eszter Udvary. "Polarization Modulated Vertical-Cavity Surface-Emitting Lasers in Quantum Key Distribution". In: *Optics, Photonics and Laser Technology 2018.* Vol. 223. Springer Series in Optical Sciences. Springer, Cham., 2019, pp. 75–92. DOI: 10.1007/978-3-030-30113-2_4.

[15] Ronald L Rivest, Adi Shamir, and Leonard Adleman. "A method for obtaining digital signatures and public-key cryptosystems". In: *Communications of the ACM* 21.2 (1978), pp. 120–126.

[16] Peter W Shor. "Algorithms for quantum computation: discrete logarithms and factoring". In: *Proceedings 35th annual symposium on foundations of computer science.* IEEE. 1994, pp. 124–134.

[17] Enrique Martin-Lopez et al. "Experimental realization of Shor's quantum factoring algorithm using qubit recycling". In: *Nature photonics* 6.11 (2012), pp. 773–776.

[18] Frank Miller. *Telegraphic code to insure privacy and secrecy in the transmission of telegrams*. CM Cornwell, 1882.

[19] GS Vernam. "Secret signaling system". 1310719. July 1919. URL: `https://patents.google.com/patent/US1310719A/en`.

[20] Claude E Shannon. "Communication theory of secrecy systems". In: *The Bell system technical journal* 28.4 (1949), pp. 656–715.

[21] Claude E Shannon. "A mathematical theory of communication". In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[22] Stefano Pirandola et al. "Advances in quantum cryptography". In: *Advances in Optics and Photonics* 12.4 (2020), pp. 1012–1236.

[23] Sándor Imre and László Gyöngyösi. "Introduction to Quantum Information Theory". In: *Advanced Quantum Communications*. John Wiley & Sons, Ltd, 2012. Chap. 2, pp. 11–64. ISBN: 978-1-118-33746-2. DOI: `10.1002/9781118337462.ch2`.

[24] William K Wootters and Wojciech H Zurek. "A single quantum cannot be cloned". In: *Nature* 299.5886 (1982), pp. 802–803.

[25] Charles H Bennett. "Quantum cryptography using any two nonorthogonal states". In: *Physical review letters* 68.21 (1992), p. 3121.

[26] Roy J Glauber. "Coherent and incoherent states of the radiation field". In: *Physical Review* 131.6 (1963), p. 2766.

[27] Bruno Huttner et al. "Quantum cryptography with coherent states". In: *Physical Review A* 51.3 (1995), p. 1863.

[28] Norbert Lütkenhaus and Mika Jahma. "Quantum key distribution with realistic states: photon-number statistics in the photon-number splitting attack". In: *New Journal of Physics* 4.1 (2002), p. 44.

[29] Charles H Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". In: *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*. Vol. 1. IEEE. 1984, pp. 175–179.

[30] A Ruiz-Alba et al. "Practical Quantum Key Distribution based on the BB84 protocol". In: *Waves*. Vol. 3. 2011, pp. 4–14.

[31] Christopher A Fuchs et al. "Optimal eavesdropping in quantum cryptography. I. Information bound and optimal strategy". In: *Physical Review A* 56.2 (1997), p. 1163.

[32] Dagmar Bruß et al. "Phase-covariant quantum cloning". In: *Physical Review A* 62.1 (2000), p. 012302.

[33] Nicolas J Cerf et al. "Security of quantum key distribution using d-level systems". In: *Physical Review Letters* 88.12 (2002), p. 127902.

[34] Valerio Scarani et al. "Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulse implementations". In: *Physical review letters* 92.5 (2004), p. 057901.

[35] Won-Young Hwang. "Quantum key distribution with high loss: toward global secure communication". In: *Physical Review Letters* 91.5 (2003), p. 057901.

[36] Artur K Ekert. "Quantum cryptography based on Bell's theorem". In: *Physical review letters* 67.6 (1991), p. 661.

[37] Makoto Matsumoto and Takuji Nishimura. "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator". In: *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 8.1 (1998), pp. 3–30.

[38] Miguel Herrero-Collantes and Juan Carlos García-Escartín. "Quantum random number generators". In: *Reviews of Modern Physics* 89.1 (2017), p. 015004.

[39] Helmut Schmidt. "Quantum-mechanical random-number generator". In: *Journal of Applied Physics* 41.2 (1970), pp. 462–468.

[40] CH Vincent. "The generation of truly random binary numbers". In: *Journal of Physics E: Scientific Instruments* 3.8 (1970), p. 594.

[41] John Walker. *HotBits: Genuine random numbers, generated by radioactive decay.* 1996. URL: http://www.fourmilab.ch/hotbits/.

[42] Thomas Jennewein et al. "A fast and compact quantum random number generator". In: *Review of Scientific Instruments* 71.4 (2000), pp. 1675–1680.

[43] André Stefanov et al. "Optical quantum random number generator". In: *Journal of Modern Optics* 47.4 (2000), pp. 595–598.

[44] G. Ribordy and O. Guinnard. *Method and apparatus for generating true random numbers by way of a quantum optics process.* US Patent 7,519,641. Apr. 2009. URL: http://www.google.com/patents/US7519641.

[45] Patrick Bronner et al. "Demonstrating quantum random with single photons". In: *European journal of physics* 30.5 (2009), p. 1189.

[46] Qing Luo et al. "Quantum random number generator based on single-photon emitter in gallium nitride". In: *Optics Letters* 45.15 (2020), pp. 4224–4227.

[47] Markus Gräfe et al. "On-chip generation of high-order single-photon W-states". In: *Nature Photonics* 8.10 (2014), pp. 791–795.

[48] Harald Fürst et al. "High speed optical quantum random number generation". In: *Optics express* 18.12 (2010), pp. 13029–13037.

[49] Emanoela de Jesus Lopes Soares, Fabio Alencar Mendonca, and Rubens Viana Ramos. "Quantum Random Number Generator Using Only One Single-Photon Detector". In: *IEEE Photonics Technology Letters* 26.9 (2014), pp. 851–853.

[50] Simone Tisa et al. "High-speed quantum random number generation using CMOS photon counting detectors". In: *IEEE Journal of Selected Topics in Quantum Electronics* 21.3 (2015), pp. 23–29.

[51] Jian-min Wang et al. "A bias-free quantum random number generation using photon arrival time selectively". In: *IEEE Photonics Journal* 7.2 (2015), pp. 1–8.

[52] Fang-Xiang Wang et al. "Robust quantum random number generator based on avalanche photodiodes". In: *Journal of Lightwave Technology* 33.15 (2015), pp. 3319–3326.

[53] Min Ren et al. "Quantum random-number generator based on a photon-number-resolving detector". In: *Physical Review A* 83.2 (2011), p. 023820.

[54] Yi Jian et al. "Two-bit quantum random number generator based on photon-number-resolving detection". In: *Review of Scientific Instruments* 82.7 (2011), p. 073109.

[55] Bruno Sanguinetti et al. "Quantum random number generation on a mobile phone". In: *Physical Review X* 4.3 (2014), p. 031056.

[56] Caitlin RS Williams et al. "Fast physical random number generator using amplified spontaneous emission". In: *Optics express* 18.23 (2010), pp. 23584–23597.

[57] Taiki Yamazaki and Atsushi Uchida. “Performance of random number generators using noise-based superluminescent diode and chaos-based semiconductor lasers”. In: *IEEE Journal of Selected Topics in Quantum Electronics* 19.4 (2013), pp. 0600309–0600309.

[58] Xiaowen Li et al. “Scalable parallel physical random number generator based on a superluminescent LED”. In: *Optics letters* 36.6 (2011), pp. 1020–1022.

[59] Apostolos Argyris et al. “Sub-Tb/s physical random bit generators based on direct detection of amplified spontaneous emission signals”. In: *Journal of Lightwave Technology* 30.9 (2012), pp. 1329–1334.

[60] Y Liu et al. “Implementation of 1.6 Tb s^{-1} truly random number generation based on a super-luminescent emitting diode”. In: *Laser Physics Letters* 10.4 (2013), p. 045001.

[61] Lei Li et al. “Random bit generator using delayed self-difference of filtered amplified spontaneous emission”. In: *IEEE Photonics Journal* 6.1 (2014), pp. 1–9.

[62] Anthony Martin et al. “Quantum random number generation for 1.25-GHz quantum key distribution systems”. In: *Journal of Lightwave Technology* 33.13 (2015), pp. 2855–2859.

[63] Jie Yang et al. “Randomness quantification for quantum random number generation based on detection of amplified spontaneous emission noise”. In: *Quantum Science and Technology* 6.1 (2020), p. 015002.

[64] Emil Simion. “The relevance of statistical tests in cryptography”. In: *IEEE Security & Privacy* 13.1 (2015), pp. 66–70.

[65] Andrew L. Rukhin et al. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. Tech. rep. Spec. Pub. 800-22, Rev. 1a. Gaithersburg, MD, United States: National Institute of Standards & Technology, 2010.

[66] Pierre L’Ecuyer and Richard Simard. “TestU01: A C library for empirical testing of random number generators”. In: *ACM Transactions on Mathematical Software (TOMS)* 33.4 (2007), pp. 1–40.

[67] George Marsaglia. *DIEHARD: a battery of tests of randomness*. 1996. URL: http://stat.fsu.edu/pub/diehard/.

[68] Robert G Brown. *Dieharder: A random number test suite*. 2016. URL: https://www.phy.duke.edu/~rgb/General/dieharder.php.

[69] Rainer Michalzik. "VCSEL fundamentals". In: *VCSELs: Fundamentals, Technology and Applications of Vertical-Cavity Surface-Emitting Lasers*. Ed. by Rainer Michalzik. Berlin, Heidelberg: Springer, 2013. Chap. 2, pp. 19–75. ISBN: 978-3-642-24986-0. DOI: 10.1007/978-3-642-24986-0_2.

[70] Rainer Michalzik. "VCSELs: A Research Review". In: *VCSELs: Fundamentals, Technology and Applications of Vertical-Cavity Surface-Emitting Lasers*. Ed. by Rainer Michalzik. Berlin, Heidelberg: Springer, 2013. Chap. 2, pp. 3–18. ISBN: 978-3-642-24986-0. DOI: 10.1007/978-3-642-24986-0_1.

[71] Gwenaelle Vest et al. "Design and evaluation of a handheld quantum key distribution sender module". In: *IEEE Journal of Selected Topics in Quantum Electronics* 21.3 (2015), pp. 131–137.

[72] Rainer Michalzik and Johannes Michael Ostermann. "Polarization Control of VCSELs". In: *VCSELs: Fundamentals, Technology and Applications of Vertical-Cavity Surface-Emitting Lasers*. Ed. by Rainer Michalzik. Berlin, Heidelberg: Springer, 2013. Chap. 5, pp. 147–179. ISBN: 978-3-642-24986-0. DOI: 10.1007/978-3-642-24986-0_5.

[73] Josep Martin-Regalado et al. "Polarization properties of vertical-cavity surface-emitting lasers". In: *IEEE Journal of Quantum Electronics* 33.5 (1997), pp. 765–783.

[74] Kent D Choquette et al. "Gain-dependent polarization properties of vertical-cavity lasers". In: *IEEE Journal of Selected Topics in Quantum Electronics* 1.2 (1995), pp. 661–666.

[75] M San Miguel, Q Feng, and Jerome V Moloney. "Light-polarization dynamics in surface-emitting semiconductor lasers". In: *Physical Review A* 52.2 (1995), p. 1728.

[76] Krassimir Panajotov and Franco Prati. "Polarization Dynamics of VCSELs". In: *VCSELs: Fundamentals, Technology and Applications of Vertical-Cavity Surface-Emitting Lasers*. Ed. by Rainer Michalzik. Berlin, Heidelberg: Springer, 2013. Chap. 6, pp. 181–231. ISBN: 978-3-642-24986-0. DOI: 10.1007/978-3-642-24986-0_6.

[77] Adam B Kaplan. "Investigating the Polarization Properties of Vertical-Cavity Surface-Emitting Lasers". B.A. honors thesis. Amherst College, 2007.

[78] Salam Nazhan and Zabih Ghassemlooy. "Polarization Switching Dependence of VCSEL on Variable Polarization Optical Feedback". In: *IEEE Journal of Quantum Electronics* 53.4 (2017), pp. 1–7.

[79] S Bandyopadhyay et al. “VCSEL polarization control by optical injection”. In: *Journal of lightwave technology* 21.10 (2003), pp. 2395–2404.

[80] WH Loh and Chung L Tang. “Numerical investigation of ultrahigh frequency polarization self-modulation in semiconductor lasers”. In: *IEEE journal of quantum electronics* 27.3 (1991), pp. 389–395.

[81] Shijun Jiang et al. “High-frequency polarization self-modulation in vertical-cavity surface-emitting lasers”. In: *Applied physics letters* 63.26 (1993), pp. 3545–3547.

[82] Shijun Jiang, George Pan, and Mario Dagenais. “Fast polarization self-modulation in a vertical-cavity surface-emitting laser”. In: *Proceedings of 1994 Nonlinear Optics: Materials, Fundamentals and Applications*. IEEE. 1994, pp. 388–390.

[83] J Martin-Regalado et al. “Polarization switching in vertical-cavity surface emitting lasers observed at constant active region temperature”. In: *Applied physics letters* 70.25 (1997), pp. 3350–3352.

[84] Kent D Choquette et al. “Polarization modulation of cruciform vertical-cavity laser diodes”. In: *Applied physics letters* 64.21 (1994), pp. 2767–2769.

[85] Guy Verschaffelt et al. “Frequency response of current-driven polarization modulation in vertical-cavity surface-emitting lasers”. In: *Applied physics letters* 80.13 (2002), pp. 2248–2250.

[86] Guy Verschaffelt et al. “Frequency response of polarization switching in vertical-cavity surface-emitting lasers”. In: *IEEE journal of quantum electronics* 39.10 (2003), pp. 1177–1186.

[87] Krassimir Panajotov et al. “Polarization switching in VCSEL’s due to thermal lensing”. In: *IEEE Photonics Technology Letters* 10.1 (1998), pp. 6–8.

[88] GM Isoe, AWR Leitch, and TB Gibbon. “VCSEL polarization modulation for pulse-per-second clock signal transfer in optical frequency distribution systems”. In: *Optoelectronics Letters* 14.5 (2018), pp. 376–379.

[89] Ajit V Barve et al. “Fast polarization modulation in vertical cavity lasers with electrical RF injection”. In: *ISLC 2012 International Semiconductor Laser Conference*. IEEE. 2012, pp. 1–2.

[90] Ajit V Barve et al. “Ultrafast polarization modulation in vertical cavity surface emitting lasers with frequency dependent current injection”. In: *Applied Physics Letters* 101.25 (2012), p. 251104.

[91] Ajit V Barve et al. “High speed polarization modulation of oxide confined asymmetric VCSELs in multimode regime”. In: *2013 IEEE Photonics Conference*. IEEE. 2013, pp. 246–247.

[92] Ajit V Barve et al. “Fast, electrically controlled polarization modulation of multimode vertical-cavity surface-emitting lasers by RF frequency modulation”. In: *Optics express* 21.25 (2013), pp. 31092–31097.

[93] Ajit V Barve et al. “Ultrafast electrical polarization modulation in VCSEL with asymmetric current injection”. In: *Optical Interconnects Conference, 2014 IEEE*. IEEE. 2014, pp. 91–92.

[94] Larry A Coldren and Ajit V Barve. “On-chip VCSEL interconnects enabled by 3-D interposer-based integration and polarization modulation”. In: *2015 IEEE Summer Topicals Meeting Series (SUM)*. IEEE. 2015, pp. 150–151.

[95] Kwok K Ng. “Electroabsorption Modulator”. In: *Complete Guide to Semiconductor Devices*. John Wiley & Sons, Ltd, 2009. Chap. 69, pp. 521–527. ISBN: 978-1-118-01476-9. DOI: https://doi.org/10.1002/9781118014769.ch69.

[96] Shashank Gupta et al. “50GHz Ge waveguide electro-absorption modulator integrated in a 220nm SOI photonics platform”. In: *Optical Fiber Communication Conference*. Optical Society of America. 2015, Tu2A–4.

[97] M Trajkovic et al. “55GHz EAM bandwidth and beyond in InP active-passive photonic integration platform”. In: *CLEO: Science and Innovations*. Optical Society of America. 2018, JTh5A–8.

[98] Zhi Liu et al. “56 Gbps high-speed Ge electro-absorption modulator”. In: *Photonics Research* 8.10 (2020), pp. 1648–1652.

[99] Sebastian Nauerth et al. “Information leakage via side channels in freespace BB84 quantum cryptography”. In: *New Journal of Physics* 11.6 (2009), p. 065001. DOI: 10.1088/1367-2630/11/6/065001.

[100] Hai-Qiang Ma, Yuejian Xie, and Ling-An Wu. “Random number generation based on the time of arrival of single photons”. In: *Applied optics* 44.36 (2005), pp. 7760–7763.

[101] LM Yu et al. “Note: A sampling method for quantum random bit generation”. In: *Review of Scientific Instruments* 81.4 (2010), p. 046107.

[102] James F Dynes et al. “A high speed, postprocessing free, quantum random number generator”. In: *Applied physics letters* 93.3 (2008), p. 031109.

[103] Michael A Wayne et al. “Photon arrival time quantum random number generation”. In: *Journal of Modern Optics* 56.4 (2009), pp. 516–522.

[104] Michael Wahl et al. “An ultrafast quantum random number generator with provably bounded output bias based on photon arrival time measurements”. In: *Applied Physics Letters* 98.17 (2011), p. 171105.

[105] Michael Wahl et al. “Addendum: An ultrafast quantum random number generator with provably bounded output bias based on photon arrival time measurements [Appl. Phys. Lett. 98, 171105 (2011)]”. In: *Applied Physics Letters* 101.15 (2012), p. 171105.

[106] KS Kravtsov et al. “Minimalist design of a robust real-time quantum random number generator”. In: *JOSA B* 32.8 (2015), pp. 1743–1747.

[107] Raghad Saeed Hasan et al. “A true random number generator based on the photon arrival time registered in a coincidence window between two single-photon counting modules”. In: *Chinese journal of physics* 56.1 (2018), pp. 385–391.

[108] Alessandro Tomasi et al. “Model, validation, and characterization of a robust Quantum Random Number Generator based on photon arrival time comparison”. In: *Journal of Lightwave Technology* 36.18 (2018), pp. 3843–3854.

[109] Hesong Xu et al. “A SPAD-based random number generator pixel based on the arrival time of photons”. In: *Integration* 64 (2019), pp. 22–28.

[110] Nicola Massari et al. “A Compact TDC-based Quantum Random Number Generator”. In: *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. IEEE. 2019, pp. 815–818. DOI: 10.1109/ICECS46596.2019.8964941.

[111] Mario Stipčević and B Medved Rogina. “Quantum random number generator based on photonic emission in semiconductors”. In: *Review of scientific instruments* 78.4 (2007), p. 045104. DOI: 10.1063/1.2720728.

[112] Abbas Khanmohammadi et al. “A monolithic silicon quantum random number generator based on measurement of photon detection time”. In: *IEEE Photonics Journal* 7.5 (2015), pp. 1–13.

[113] Ágoston Schranz, Eszter Udvary, and Balázs Matolcsy. “Evaluation of a Time-of-Arrival Quantum Random Number Generator with Device Limitations”. Unpublished manuscript.

[114] M.O. Scully and M.S. Zubairy. *Quantum Optics*. Cambridge: Cambridge University Press, 1997. ISBN: 9780521435956. URL: `https://books.google.hu/books?id=20ISsQCKKmQC`.

[115] William R. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach*. 2nd. Springer, 1994, pp. 122–126. ISBN: 978-3-540-57280-0.

[116] Jörg W Müller. "Generalized dead times". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 301.3 (1991), pp. 543–551.

[117] Catherine Forbes et al. *Statistical distributions*. 4th edition. John Wiley & Sons, 2011.

[118] Malvin Carl Teich and Bahaa EA Saleh. "Effects of random deletion and additive noise on bunched and antibunched photon-counting statistics". In: *Optics letters* 7.8 (1982), pp. 365–367.

[119] Alfréd Rényi et al. "On measures of entropy and information". In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California. 1961.

[120] Robert Konig, Renato Renner, and Christian Schaffner. "The Operational Meaning of Min- and Max-Entropy". In: *IEEE Transactions on Information Theory* 55.9 (Sept. 2009), pp. 4337–4347. DOI: `10.1109/tit.2009.2025545`.

List of Figures

List of Tables