Università degli Studi di Padova

Dipartimento di Fisica & Astronomia G. Galilei
Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Fisica

SECURE RANDOM NUMBERS FROM QUANTUM IMAGES

Laureando:
Ugo Zanforlin
Matricola N°: 1057391

Relatore:
Ch.mo Prof. P. Villoresi

Correlatore:
Dott. D. G. Marangon

Anno Accademico 2014 - 2015

Introduzione

Un numero generato senza possedere alcuna caratteristica specifica è chiamato *numero casuale*. Questi numeri sono utilizzati per molteplici scopi. Ricoprono un ruolo fondamentale nel garantire assoluta segretezza di un sistema nel campo della crittografia. L'affidabilità e l'esecuzione nell'ambiente digitale del mondo moderno si poggiano sul grado di casualità che si è in grado di creare. I numeri sono anche usati nelle comunicazioni e nell'informatica, i.e. simulazioni Monte Carlo, predizioni finanziarie, test di biofisica e medicina nucleare, ed in molti altri campi. Sono utili anche per applicazioni prettamente commerciali come la lotteria, slot machines e trasmissioni wireless. [1, 2]

Un **T**rue **R**andom **N**umber **G**enerator (TRNG) è matematicamente definito come un generatore che produce numeri, estratti da un insieme con N elementi, con probabilità $1/N$. Per assicurare che un numero sia effettivamente casuale, esso deve possedere due proprietà fondamentali: uniformità ed imprevedibilità statistica. La prima richiede che i valori casuali debbano essere distribuiti uniformemente all'interno del loro intervallo di esistenza così che ciascuno abbia la medesima probabilità di essere estratto. La seconda, invece, richiede che, senza alcun mezzo possibile, non si sia in grado di predire il numero successivo di una sequenza generata casualmente, anche solo parzialmente.

Questa specifica proprietà, in particolare, è indispensabile in molti protocolli criptografici poiché assicura un alto livello di sicurezza nella distribuzione di chiave (key distribution) e nel processo di autenticazione. Esempi tipici sono gli schemi di condivisione segreta (secret sharing schemes), usati per la memorizzazione sicura di informazioni in supporti informatici tramite la loro criptazione e frammentazione, e nella quantum key distribution, il più alto grado di sicurezza realizzato. [3] Di conseguenza, è cruciale sviluppare un metodo in grado di generare sequenze casuali che rispettino tali proprietà.

I **R**andom **N**umber **G**enerators (RNGs) possono essere suddivisi in due categorie: **P**seudo-**R**andom **N**umber **G**enerators (PRNGs) e physical random number generators (physical RNGs). PRNGs sono basati su algoritmi od anche su combinazioni di tali algoritmi. [4, 5, 6, 7] Sequenze finite di numeri generati da PRNGs posso apparire casuali, così da superare alcuni test statistici specificatamente realizzati per testarne le caratteristiche. Tuttavia, queste sequenze sono create in maniera deterministica da un seme (seed) comune e ciò comporta che i numeri generati, dopo un tempo sufficientemente lungo, si ripeteranno. [8] Questo comportamento può causare seri problemi per applicazioni nel campo della sicurezza o di sistemi computazionali in parallelo, si veda per esempio [9]. Physical RNGs usano casualità da osservabili fisiche come il rumore fotonico, il rumore termico nei resistori, jitter di frequenza di oscillatori e, recentemente, sistemi caotici di LASERs a semiconduttore. [10, 11, 12, 13, 14]

Tali generatori, tuttavia, non possiedono lo stesso rate di generazione dei PRNGs a causa di limitazioni nel rate di acquisizione e nella capacità del meccanismo. Sebbene

sfruttare processi fisici casuali per generare sequenze di numeri sembri ragionevole, sistemi puramente classici possiedono una natura deterministica oltre determinate scale temporali. Processi quantistici, al contrario, permettono di sfruttare l'assoluta presenza della casualità e quindi dell'estrazione casuale. Queste tipologie di generatori formano una categoria a se stante chiamata **Q**uantum **R**andom **N**umber **G**enerators (QRNGs).

La meccanica quantistica elementare mostra che la natura stessa del mondo microscopico è imprevedibile e quindi "casuale". In particolare, sistemi quantistici sono puramente descritti in termini probabilistici. Esistono diversi fenomeni quantistici utili per la generazione di numeri casuali. I più comuni sono quelli relativi al decadimento di nuclei radioattivi [15, 16] dove il processo casuale di una emissione α, β o γ permette di generare bits, 0 or 1, in corrispondenza o meno di una misura da parte del rivelatore. Tuttavia il sistema richiede estrema precauzione nel maneggiare e conservare sostanze radioattive. Al contrario, sistemi ottici sono molto più sicuri e semplici da realizzare. Alcuni di questi sono: beams splitting di singolo fotone, misure di polarizzazione di singolo fotone [17] o periodi light-dark di segnali di fluorescenza relativi a risonanze di singoli ioni intrappolati. [18, 19]

In implementazioni pratiche di QRNGs, il processo fisico non è mai ottenuto perfettamente. Sono sempre presenti influenze (noises) che non sono controllate o controllabili dall'utente. Lo scopo di questo lavoro è di presentare un nuovo approccio per estrarre numeri casuali veramente casuali tramite l'impiego di specifici post-processings. Il modello teorico e matematico è basato sul lavoro di D. Frauchiger, R. Renner and M. Troyer *"True randomness from realistic quantum devices"* [20]. La tesi si focalizza sul metodo base introdotto nel precedente lavoro e lo estende ad uno più generale concentrandosi sull'analisi richiesta.

Cap. (1) presenta l'approccio matematico concentrandosi sulla natura quantistica del processo. Cap. (2) presenta la teoria della misura ed in particolare sulle POVM. Cap. (3) mostra la teoria della statistica di fotorivelazione necessaria per lo studio del processo di rilevazione dei fotoni. Cap. (4) spiega i modelli teorici usati per introdurre gli effetti perturbativi presenti in ogni apparato reale. Cap. (5) mostra il setup usato per l'esperimento. Cap. (6) descrive l'analisi eseguita per caratterizzare la telecamera di rilevazione usata nell'esperimento, in particolar modo la sua efficienza quantica. Cap. (7) riporta i risultati sperimentali dell'analisi sulle influenze elettroniche. Cap. (8) presenta il concetto di entropia necessario per "quantificare" l'informazione estraibile da una stringa di bit. Il capitolo introduce inoltre la generalizzazione del modello quantistico e l'implementazione dei precedenti rumori elettronici nella sua analisi. Cap. (9) mostra gli estrattori classici e quantistici impiegati per generare numeri casuali. Cap. (10) e (11) illustrano i test statistici usati per verificare la casualità ed i relativi risultati sia per i numeri generati per via classica che quantistica.

Introduction

A number generated with no apparent rule is called a *random number*. Random numbers are widely used for various purposes. They play a crucial role in guaranteeing the secrecy of a system in the field of cryptography. The performance and reliability of the digital networked society rely on the degree of randomness that it can generate. They are used in communication and computing such as Monte Carlo simulations, financial predictions, biophysics and nuclear medicine testing and even more. They are also necessary for many commercial applications such as lottery games, slot machines and wireless data transmissions. [1, 2]

A **T**rue **R**andom **N**umber **G**enerator (TRNG) is mathematically defined as a number generator that produces a number out of a large set N with probability $1/N$. To ensure that a number is truly random, there are two fundamental properties that it must possess: statistical uniformity and unpredictability. The first one specify that the values of random numbers must be uniformly distributed over their range so that extracting any of them is equally probable. The latter, instead, requires that with no means it is possible to predict, even partially, the next value in a random sequence.

This specific property, in particular, is indispensable in many cryptographic protocols as this ensures a high-security level of key distribution and authentication. Typical examples are secret sharing schemes, which are used for securely storing data in storage devices by encrypting and dividing it, and quantum key distribution, which is expected to achieve the ultimate security. [3] Therefore, it is crucial to develop a method of generating random sequences with these properties.

Random **N**umber **G**enerators (RNGs) may be classified into two categories: **P**seudo-**R**andom **N**umber **G**enerators (PRNGs) and physical random number generators (physical RNGs). PRNGs are based on algorithms or even a combinations of algorithms. [4, 5, 6, 7] Finite sequences of numbers generated by PRNGs may appear random so to pass some random tests specially made to verify randomness. However, these sequences are created deterministically from a common seed and thus permitting to obtain the same numbers after a sufficiently long time [8]. This behavior may cause serious problems for applications in security or parallel computation systems, see for example [9]. Physical RNGs use randomness from a physical observable such as photon noise, thermal noise in resistors, frequency jitter of oscillators and, more recently, chaotic semiconductor LASERs. [10, 11, 12, 13, 14]

These kind of generators, however, do not possess the same generation rate of the PRNGs due to limitations of the acquisition rate and power of the mechanisms. Although relying on physical random processes to generate sequences of numbers may sound reasonable, purely classical systems have a deterministic nature over relevant time scales. In contrast, quantum processes permit to exploit the absolute power of

randomness and then randomness extraction. These typologies of generators form a category by themselves called **Q**uantum **R**andom **N**umber **G**enerators (QRNGs).

Elementary quantum mechanics explains that the very nature of the microscopic world is unpredictable and thus "random". In particular, quantum systems are described using probabilistic notions. There exist many different quantum phenomenons suitable to generate random numbers. The most common is the decay of radioactive nuclei [15, 16] where the random process of α, β or γ emission permits to generate bits, 0 or 1, if the detector registers an event or not. However, this system requires extra precautions in handling and storing radioactive substances. On the contrary, optical systems are much safer and simpler to realize. Some of these are: single photon beams splitting, single photons polarization measurements [17] or light-dark periods of single trapped ions' resonance fluorescence signal. [18, 19]

In practical implementations of QRNGs, the physical process is never achieved perfectly. There are always influences (noises) that are not fully controlled or controllable by the user. This work's aim is to present a new approach to extract truly random numbers through the usage of specific post-processing. The mathematical and theoretical layout is based on the work of D. Frauchiger, R. Renner and M. Troyer *"True randomness from realistic quantum devices"* [20]. The thesis focuses on the basic method introduced in this work and then extend it to a more general one concentrating on the required analysis.

Chap. (1) presents the mathematical approach concentrating on the quantum nature of the process. Chap. (2) presents the measurement theory focusing on the POVM. Chap. (3) shows the photocount statistic theory necessary to study the process of photon detection. Chap. (4) explains the theoretical models used to account for the noises effects present in every realistic device. Chap. (5) displays the setup used for the experiment. Chap. (6) describes the analysis made to characterize the photon detection device used for the research focusing on its quantum efficiency. Chap. (7) reports the experimental results from the analysis of device's electronic noises. Chap. (8) presents the concept of entropy necessary to "quantify" the information carried by a string bit. It also introduces the generalization of the quantum model and the implementation of the previously introduced electronic noises. Chap. (9) shows both classical and quantum extractors used to generate random strings of bits. Chaps. (10) and (11) explain the statistical tests used to verify randomness and the tests results for both the classical and quantum generated bit strings.

Contents

1 QRNG model

Secure random numbers generation is not just of academic interest, e.g. data analysis, numerical simulations. It is also very relevant for practice, e.g. gambling, lottery extractions, in particular for cryptographic systems where is essential to guarantee their security during key generation and distribution between parties. Both theoretical and practical approaches have been made to quantify randomness. All of them however analyze the number extracted by itself instead of the process that produced it. The current chapter presents the theoretical model designed by *D. Frauchiger, R. Renner and M. Troyer* [20]. The model explains how to extract *truly* random numbers using a post-processing algorithm of the *raw randomness* generated by an imperfect device, i.e. presence of noise or any kind of external influences that cannot be managed.

1.1 True randomness

The usage of spacetime variables notion is essential to introduce a formal and quantitative definition of pure randomness. These are random variables with an associated coordinate that indicates the physical location of the value in relativistic spacetime [21]. The output, X, of a random process as well as all *side information*, i.e. spacetime variables which model any additional information that may be correlated to X. The spacetime coordinates of X should be interpreted as the event where the process starts, generating X. For side information, the coordinates of the corresponding spacetime variables indicate when and where this information is accessible. From these assumptions, it is possible to define a truly random variable.

Definition 1. X is called *ϵ-truly random* if it's *ϵ-close* to uniform and uncorrelated to all other space time variables which are not in the future light cone of X. Denoting this set by Γ_X, the mathematical formulation is defined as follows:

$$\frac{1}{2}\,||P_{X\,\Gamma_X} - P_{\bar{X}}\mathrm{X}P_{\Gamma_X}||_1 \leq \epsilon \tag{1.1}$$

where

$$P_{\bar{X}}(x) = \frac{1}{|\Omega|} \qquad \forall\, x \tag{1.2}$$

and

$$\frac{1}{2}\,||P_X - Q_X||_1 = \frac{1}{2}\sum_x |P_X(x) - Q_X(x)| \tag{1.3}$$

is the *trace distance*.

The trace distance has the following interpretation: if two probability distributions are ϵ-*close* in trace distance, then one may consider the two systems as identical except with probability at most ϵ.

1.2 Leftover Hash Lemma

Considering a random variable X that is partially known to an adversary, i.e. the enemy possesses side information E correlated to X, one may ask whether it is possible to extract from X a part Z that is entirely unknown to adversary, i.e. uniform conditioned on E. If the answer is affirmative, what is the maximum size of Z and how it is computed? The answer to all these question lies in the *Leftover Hash Lemma*. More precisely, the output of a function f selected randomly from a chosen family of functions $\mathcal{F}$, called *two-universal family of hash functions*, provides Z. In most researches, the *universal hashing* method is based on probability theory and *side information* of a purely and entirely classical. By these means the (classical) Lemma states the following:

Lemma 1 (Classical Leftover Hash Lemma)**.** *Let X and E be random variables and let $\mathcal{F}$ be a two-universal family of hash functions with domain Ω and range $\{0,1\}^t$. Then, on average over the choices of f from $\mathcal{F}$, the distribution of the output $Z \equiv f(X)$ is Delta-close from uniform conditioned on E, where:*

$$\Delta = \frac{1}{2}\sqrt{2^{t-H_{min}(X|E)}}$$

The Lemma immediately implies that for a *fixed* joint distribution of X and E, there is a *fixed* function f that extracts almost uniform randomness. More precisely, given any $\Delta > 0$, there exists a function f that produces

$$t = \left\lfloor H_{min}(X|E) - 2\log_2 \frac{1}{2\Delta} \right\rfloor$$

bit that are Δ-*close* to uniform and independent of E.

A majority of original works on universal hashing are based entirely on probability theory and side information. These papers are (often implicitly) assumed to be represented by a classical system E (modeled as a random variable). The reasons for this kind of formulation lies in the following reasoning. The process of hashing is purely and entirely *classic*, since values of a random variable X are mapped to those of a random variable Z by a function. It is then reasonable to assume that it is irrelevant to consider the physical nature of the side information and justify a classical treatment. This assumption is, however, not necessarily the case. Some recent works [22, 23], showed that the output of certain extractor functions may be partially known if side information about their input is stored in a *quantum* device of a certain size, while the same output is almost uniform conditioned on any side information stored in a *classical* system of the same size. Knowing this, it is necessary to generalize Lemma 1 so to reflect the

quantum nature of a system E [24, 25, 26]. Considering the quantum nature of E, there is a need to consider the more general case of non-classical side information, i.e. X may be quantum correlated to E. Let ρ_E^x be the state of E when $X = x$ on the product space $\mathcal{H}_X \otimes \mathcal{H}_E$. This situation can be characterized conveniently by a *classical-quantum* state (*CQ-state*) in the following compact form:

$$\rho_{XE} = \sum_{x \in \Omega} P_X(x) \left|x\right\rangle \left\langle x\right| \otimes \rho_E^x \tag{1.4}$$

where the classical value $x \in \Omega$ may be viewed as encoded in mutually orthogonal states $\{\left|x\right\rangle_{x \in \Omega}\}$ on a quantum system X. To quantify the *quality of randomness* it's mainly used the concept of *min-Entropy*, $H_{min}(X)$ (see Chap. (8)), where, to account for the correlation with the system E, may be reformulated as follows:

$$H_{min}(X|E) = \sup\left[\lambda : 2^{-\lambda}\, \mathbb{1}_X \otimes \sigma_E - \rho_{XE} \geq 0\,; \sigma_E \geq 0\right] \tag{1.5}$$

This corresponds to the maximum probability of guessing X given E, and therefore naturally generalizes the classical conditioned *min-Entropy* (see Chap. (8)) [27]. In the same way, the independence condition of Eq. (1.1) naturally becomes:

$$\frac{1}{2}|| \, \rho_{XE} - \rho_{\bar{X}} \otimes \rho_E \, ||_1 \leq \epsilon$$

where $||\cdot||_1 = Tr(|\cdot|)$ denotes the trace norm and $\rho_{\bar{X}} = \frac{1}{|\Omega|}\mathbb{1}_X$ the fully mixed density operator on X. The previous condition characterizes the states for which X is (almost) uniformly distributed and independent of E. An important property of this condition is that the trace norm can only decrease if it is applied a physical mapping, e.g. a measurement, on the system E. Leftover hashing is a special case of randomness extraction where the hash function is chosen from a particular class of functions, called *two-universal* [28, 29, 30, 31, 32, 33, 34]. They are defined as families $\mathcal{F}$ of functions from Ω to $\{0,1\}^t$ such that:

$$Pr\left(f(x) = f(x')\right) \leq \frac{1}{2^t}$$

for $\forall\, x, x' \in \Omega \,|\, x \neq x'$ and f chosen uniformly at random from $\mathcal{F}$.
Now having all the information required, it is possible to state the *Leftover Hash Lemma with Side Information*:

Lemma 2 (Quantum Leftover Hash Lemma)**.** *Let ρ_{XE} be a CQ-state and let $\mathcal{F}$ be a two-universal family of hash functions from Ω to $\{0,1\}^t$. Then*

$$\frac{1}{2}|| \, \rho_{F(X)EF} - \rho_{\bar{Z}} \otimes \rho_{EF} \, ||_1 \leq 2^{-\frac{1}{2}(H_{min}(X|E)-t)} \equiv \epsilon_{hash} \tag{1.6}$$

where

$$\rho_{F(X)EF} = \sum_{f \in \mathcal{F}} \frac{1}{|\mathcal{F}|} \rho_{f(x)E} \otimes \left|f\right\rangle \left\langle f\right| \tag{1.7}$$

and where $\rho_{\bar{Z}}$ is the fully mixed density operator on the space encoding $\{0,1\}^t$

The Lemma tells that, whenever $t < H_{min}(X|E)$, the output $f(x)$ of the hash function is uniform and independent of E, except with probability $\epsilon_{hash} < 1$. The entropy $H_{min}(X|E)$ thus corresponds to the amount of randomness that can be extracted from X requiring uniformity and independence from E. Furthermore, the deviation ϵ_{hash} decreases exponentially as $H_{min}(X|E)$ increases. It is important to note the inclusion of f in the state, $\rho_{F(X)EF}$, as this ensures that $f(x)$ is random even if the function f is known. For a device that generates a continuous sequence of output bits, the hash function is usually applied block-wise and the outcomes are concatenated. In this case, it is sufficient to choose the hash function once, using randomness that is independent of the device. For practice, this means that the hash function may be selected already when manufacturing the device (using independent randomness) and hardcoded on the device. For example, it can be chosen a random matrix $(t \times n)$, k_{ij}, through which define the two-universal hashing function $Y = f(x)$ as the matrix product modulo 2:

$$Y_i = \sum_{j=1}^{n} \mathrm{mod}_2 \left(k_{ij} X_j\right) \tag{1.8}$$

this way the string generated, Y_i with $i = 1, 2, \ldots t$, satisfies the conditions of Lemma 2.

1.3 Side information of an ideal QRNG

On an abstract level, a QRNG may be modeled as a process where a quantum system is prepared in a fixed state and then measured. Under the assumption that (i) the state of the system is pure and that (ii) the measurement on the system is projective, the outcomes are truly random, i.e. independent of anything preexisting [21]. However, for realistic implementations, neither of the two assumptions is usually satisfied. If the preparation is *noisy* then the system is generally, prior to the measurement, in a mixed state. Furthermore, an imperfect implementation of a projective measurement, e.g. with inefficient detectors, is no longer projective, but rather acts as a general Positive-Operator Measure, POVM (see Chap. (2)), on the system [35]. These deviations from assumptions (i) and (ii) mean that there exists side information that may be correlated to the outcomes of the measurement. The aim of the research is to quantify the amount of independent randomness that is still present in the measurements results. More precisely, find a lower bound on the conditional *min-Entropy* of the measurement results given the side information.

Definition 2. A QRNG is defined by a density operator ρ_S on a system S together with a projective measurement $\{\Pi_S^x\}_{x\in\Omega}$ on S. The *raw randomness* is the random variable X obtained by applying this measurement to a system prepared according to ρ_S.

This way, the probability distribution P_X of X is therefore given by the Born rule:

$$P_X(x) = Tr\left(\Pi_S^x \rho_S\right) \tag{1.9}$$

The following framework is based on a QRNG modeled on a PBS (**P**olarising **B**eam **S**plitter). A PBS is a birefringent material which reflects vertically polarized photons and transmits horizontally polarized photons as shown in Fig. (8.4) in Chap. (8). To generate randomness, a diagonally polarized light pulse illuminate the PBS. After passing through the PBS, the light hits one of the two detectors, labeled R_1 and R_2, depending on whether it was reflected (1) or transmitted (2). The output of the device is $X = (X_1, X_2)$ where $X_{1,2}$ are bits indicating whether the corresponding detector $R_{1,2}$ clicked. In the ideal case, where the light pulses contain exactly one photon and where the detectors are maximally efficient, only the outcomes $X = (0, 1)$ and $X = (1, 0)$ are possible. According to quantum theory, the resulting bit, indicating whether $X = (0, 1)$ or $X = (1, 0)$, is uniformly distributed and unpredictable, that is truly random. The situation changes if the ideal detectors are replaced by imperfects ones, which sometimes fail to notice an incoming photon, and if the light source sometimes emits pulses with more than one photon. It is still possible to obtain the previous outcomes but now these can no longer be interpreted as the result of a polarization measurement. Rather, it is determined by the detectors' probabilistic behavior, i.e. whether they are sensitive at the moment when the light pulses arrived. In other words, the device outputs detector noise instead of quantum randomness originating from the PBS. Prior to the realistic model, it is worth dwelling on the theoretical one.

Definition 3 (Ideal PBS-based QRNG)**.** Consider a QRNG based on a PBS, as previously described. The source as well as the PBS may be regarded as part of the state preparation, so that ρ_S corresponds to the joint state of the two light modes traveling to the detectors, $R_{1,2}$ respectively. In the ideal case, where the source emits one single diagonally polarized photon, $\rho_S = |\phi\rangle \langle\phi|_{R_1,R_2}$ is the pure state defined by

$$|\phi\rangle_{R_1,R_2} = \frac{1}{\sqrt{2}} \left(|0\rangle_{R_1} \otimes |1\rangle_{R_2} + |1\rangle_{R_1} \otimes |0\rangle_{R_2} \right)$$

where it has been used the number bases for R_1 and R_2.

Provided the detectors are perfect, their action is defined for $R = R_{1,2}$ by the projectors $\Pi_R^0 = |0\rangle \langle 0|_R$ and $\Pi_R^1 = |1\rangle \langle 1|_R$. Since each of the two modes (R_1 and R_2) is measured separately, the overall measurement is given by:

$$\left\{ \Pi_{R_1}^0 \otimes \Pi_{R_2}^0 \, , \, \Pi_{R_1}^0 \otimes \Pi_{R_2}^1 \, , \, \Pi_{R_1}^1 \otimes \Pi_{R_2}^0 \, , \, \Pi_{R_1}^1 \otimes \Pi_{R_2}^1 \right\}$$

From Eq. (1.9), the *raw randomness* $X = (X_1, X_2)$ is equivalent to a uniformly distributed bit:

$$P_X(0, 1) = P_X(1, 0) = \frac{1}{2}$$

For the realistic model some of the stringent assumptions made on the system are to be discarded.

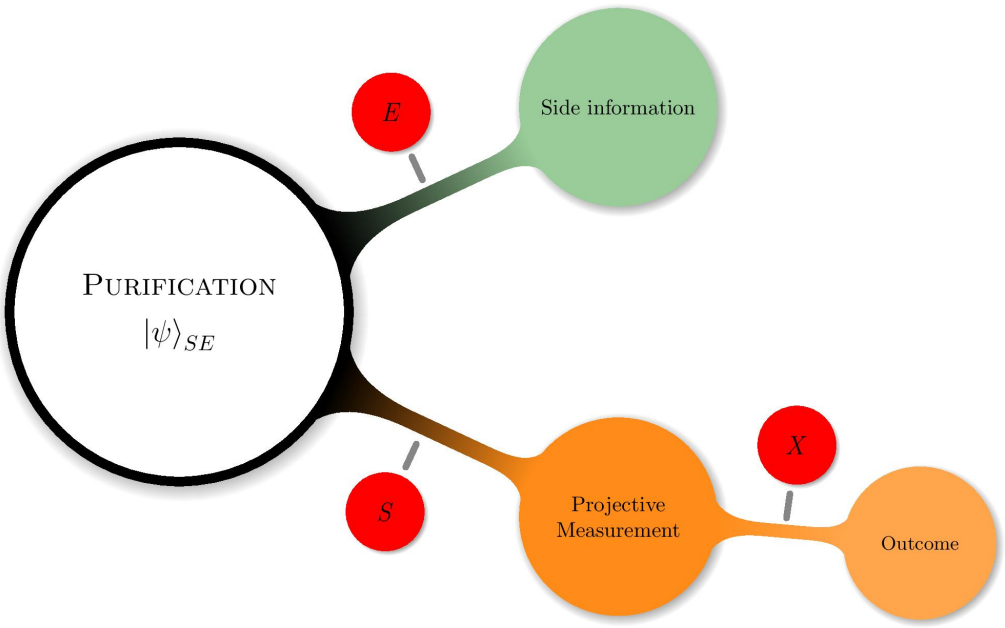


Figure 1.1: Side Information. For a QRNG, defined by a projective measurement on a system S with outcome X, all side information can be obtained from a purifying system E, i.e. an extra system that is chosen such that the joint state on S and E is pure.

Definition 4 (Realistic PBS-based QRNG)**.** A realistic device detects an incoming photon only with bounded probability η. On the subspace of the optical mode, $R = R_1$ or $R = R_2$, spanned by $|0\rangle_R$ (no photon) and $|1\rangle_R$ (1 photon), its action is given by the POVM, $\{V_R^0\,,\, V_R^1\}$ with:

$$V_R^1 = \eta\,|1\rangle\langle 1|_R \qquad \text{and} \qquad V_R^2 = \mathbb{1}_R - V_R^1$$

To describe this as a projective measurement, it is needed to consider an extended space (Neumark's theorem) with an additional subsystem, $R^{'}$, that determines whether the detector is sensitive or not ($|1\rangle_{R'}$ and $|0\rangle_{R'}$ respectively). Specifically, it may be possible to define the extended projective measurement $\left\{\Pi_{RR'}^0\,,\, \Pi_{RR'}^1\right\}$ by:

$$\Pi_{RR'}^1 = |1\rangle\langle 1|_R \otimes |1\rangle\langle 1|_{R'} \qquad \text{and} \qquad \Pi_{RR'}^0 = \mathbb{1}_{RR'} - \Pi_{RR'}^1$$

It is then easily verified that the action of $\{V_R^0\,,\, V_R^1\}$ on any state τ_R is reproduced by the action of $\left\{\Pi_{RR'}^0\,,\, \Pi_{RR'}^1\right\}$ on the product $\tau_R \otimes \tau_{R'}$ where:

$$\tau_{R'} = (1-\eta)\,|0\rangle\langle 0|_{R'} + \eta\,|1\rangle\langle 1|_{R'} \tag{1.10}$$

To asses the quality of a QRNG, it is also needed a description of its side information. To do so, it can be used a purification $|\psi\rangle_{SE}$ of the state ρ_S with purifying system E (see Fig. (1.1)). Any possible side information may now be described as the outcome of a measurement on E.

Example 1 (Side Information for an inefficient detector)**.** Consider an inefficient detector, $R = R_1$ or $R = R_2$ as in Description 4. A classical bit T may determine whether the detector is sensitive to incoming photons or not ($T = 1$ and $T = 0$ respectively). T could then be considered as side information W. This information can indeed be easily obtained from a measurement on a purification of the state $\tau_{R'}$ (see Eq. (1.10)). For example, for the purification

$$|\phi\rangle_{R'E} = \sqrt{1-\eta}\,|0\rangle_{R'} \otimes |0\rangle_E + \sqrt{\eta}\,|1\rangle_{R'} \otimes |1\rangle_E$$

the value T is retrieved as the outcome of the projective measurement $\{|0\rangle\langle 0|_E\,, |1\rangle\langle 1|_E\}$ applied to E.

Based on what has been said, now it is possible to summarize the requirements for the extraction of random numbers combined with the fact that quantum theory is complete:[1]

Definition 5. Consider a QRNG that generats raw randomness X and let E be a purifying system of S. Furthermore, let f be a function chosen uniformly at random and independently of all other values from a two-universal family of hash functions with output length

$$t \leq H_{min}(X|E) - 2\log_2\left(\frac{1}{\epsilon}\right)$$

Then the result $Z = f(X)$ is ϵ-*truly random.*

Proof. According to the definition of ϵ-*true randomness* it is needed to ensure that

$$\frac{1}{2}||\,P_{F(X)WF} - P_{\bar{Z}} \times P_{WF}\,||_1 \leq \epsilon \tag{1.11}$$

where W is any value available outside the future light cone of the event where the measurement X started, F is the random variable indicating the uniform choice of the has function from the two-universal family, and $P_{\bar{Z}}$ is the uniform distribution on $\{0,1\}^t$. It follows from the completeness of quantum theory [21] that such a W can always be obtained by a measurement of all available quantum systems, in this case E. But because the trace distance can only decrease under physical mappings, Eq. (1.11) holds whenever

$$\frac{1}{2}||\,\rho_{F(X)EF} - \rho_{\bar{Z}} \otimes \rho_{EF}\,||_1 \leq \epsilon$$

The claim then follows by the Leftover Hash Lemma (see Sec. (1.2)). □

[1]The specific demonstration is based on what has been previously exposed.

Note that Definition 5 does not require a description of classical side information, i.e. there is no need to model the side information explicitly. This is important for practice, as it could be hard to find an explicit and complete model for all classical side information present in a realistic device.

1.4 Maximum classical noise model

For a realistic QRNG, it is (essentially) never attained an ideal quantum process. The fact is that there are always various type of noise generators, both externally and internally to the system, which can not be fully controlled. In this category fall the side information too. All these influences are generally referred to as "*noise*" and using Definition 5 it is possible to quantify the true randomness of such "noisy" QRNG. However, the criterion involve the conditional *min-Entropy* for quantum systems, which may be hard to evaluate for practical devices. Fortunately it exists a way to find a classical value C which is as good as the side information E in the sense that:

$$H_{min}(X|C) \leq H_{min}(X|E) \tag{1.12}$$

holds.

The previously statement is referred to as the *Maximum classical noise model*. The random variable C may be obtained by a measurement on the system S, but this measurement must not interfere with the measurement carried out by the QRNG (see Fig. (1.2)). Furthermore, Eq. (1.12) can only holds if the measurement of C is maximally informative. This means that the post-measurement state should be pure conditioned on C.

Definition 6. A maximum classical noise model for a QRNG with state ρ_S and projective measurement $\{\Pi_S^x\}_x$ on S is a generalized measurement $\{E_S^c\}_{c\in C}$ on S such that the following requirements are satisfied:

(i) the map

$$\mathcal{P}_{X\leftarrow S} \; : \quad \tau_S \mapsto \sum_x Tr\left(\Pi_S^x \tau_S\right) |x\rangle \langle x|$$

is invariant under composition with the map

$$\mathcal{O}_{S\leftarrow S} \; : \quad \tau_S \mapsto \sum_c E_S^c \tau_S (E_S^c)^\dagger$$

i.e. $\mathcal{P}_{X\leftarrow S} \circ \mathcal{O}_{S\leftarrow S} = \mathcal{P}_{X\leftarrow S}$

(ii) the state

$$\rho_{S|C=c} = \frac{(E_S^c)^\dagger \rho_S E_S^c}{Tr\left[(E_S^c)^\dagger \rho_S E_S^c\right]}$$

obtained by conditioning on the outcome $C = c$ of the measurement $\{E_S^c\}_c$ is pure $\forall\, c \in C$.

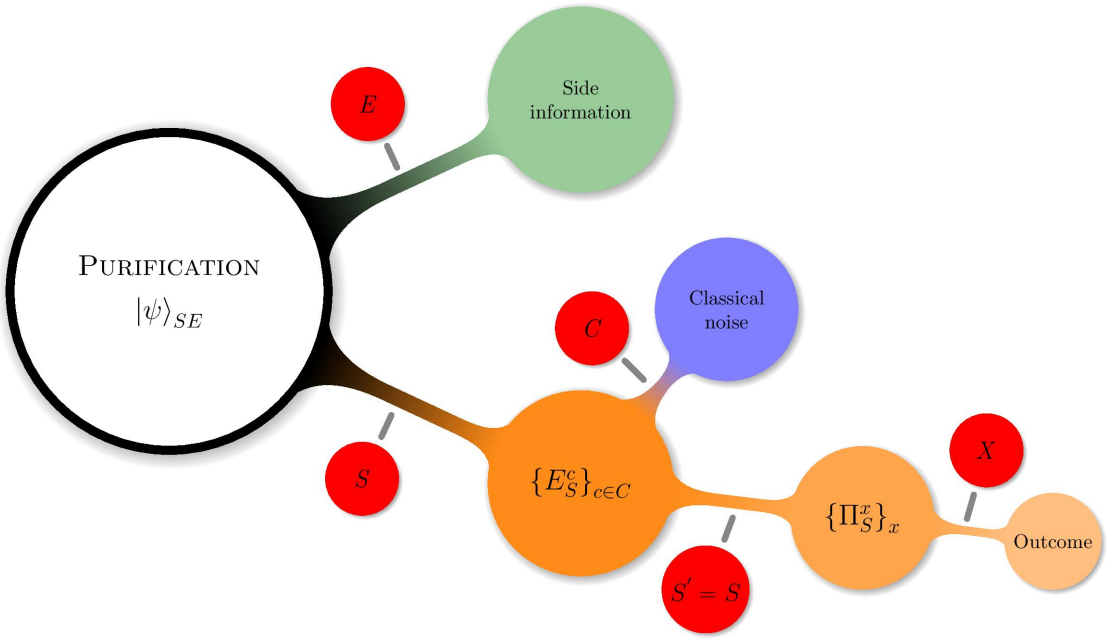


Figure 1.2: Classical noise model. The maximum classical noise C of a QRNG is defined by a measurement on S that does not affect the projective measurement carried out by the QRNG, but gives maximal information about the raw randomness X.

The outcome C of the measurement $\{E_S^c\}_c$ applied to ρ_S is called *Maximum classical noise*.

Example 2 (Maximum classical noise model for an inefficient detector)**.** Consider a system as described in Definition 4 and its description in terms of a projective measurement $\left\{\Pi^0_{RR'}, \Pi^1_{RR'}\right\}$ on an extended system. If the state ρ_R of the optical mode is pure then the measurement $\left\{E^0_{RR'}, E^1_{RR'}\right\}$ defined by

$$E^0_{RR'} = \mathbb{1}_R \otimes |0\rangle\langle 0|_{R'} \qquad \text{and} \qquad E^1_{RR'} = \mathbb{1}_R \otimes |1\rangle\langle 1|_{R'} \tag{1.13}$$

is a maximum classical noise model. To see this, note that the first criterion of Definition 6 is satisfied because this measurement commutes with the measurement $\left\{\Pi^0_{RR'}, \Pi^1_{RR'}\right\}$ of the detector. Furthermore, because $\left\{E^0_{RR'}, E^1_{RR'}\right\}$ restricted to R' is a rank-one measurement, the post-measurement state is pure, so that the second criterion of Definition 6 is also satisfied. Note that the maximum classical noise, C,defined as the outcome of the measurement $\left\{E^0_{RR'}, E^1_{RR'}\right\}$, is a bit that indicates whether the detector is sensitive or not, as in Example 1. For the PBS-based QRNG with two detectors, R_1 and R_2, the classical noise would be $C = (T_1, T_2)$, where T_1 and T_2 are the corresponding indicator bits for each detector.

The statement of Eq. (1.12) is then verified using the following Lemma.

Lemma 3. *Consider a QRNG that generates raw randomness X and let E be a purifying system. Then, for any maximum classical noise C*

$$H_{min}(X|C) \leq H_{min}(X|E) \tag{1.14}$$

Proof. The first requirement od Definition 6 guarantees that the random variables C and X are defined simultaneously. Because, by the second requirement of Definition 6, the state S conditioned on C is pure, it is necessarily indipendent of E. Since X is obtained by a measurement on S, it is also independent of E, conditioned on C. Hence it is obtained the Markov chain

$$X \leftrightarrow C \leftrightarrow E$$

which implies

$$H_{min}(X|C) = H_{min}(X|CE)$$

The assertion then follows from the data processing inequality for the *min-Entropy*[2]

$$H_{min}(X|CE) \leq H_{min}(X|E) \qquad \square$$

From the joint probability distribution determined by the Born rule

$$P_{XC}(x,c) = Tr\left(\Pi_S^x E_S^c \rho_S \left(E_S^c\right)^\dagger\right) \tag{1.15}$$

the conditional *min-Entropy*, $H_{min}(X|C)$, can be calculated using Eq. (8.3).

1.5 Quantum randomness

When analyzing realistic QRNGs, purely classical random variables conveniently describe them. As shown, a maximum classical noise model and, hence, a random variable C (see Definition 6) captures all side information. Similarly, it can be introduced a random variable, Q, that accounts for the "quantum randomness", i.e. the part of the randomness that is intrinsically unpredictable. The idea is to define this as the randomness that "remains" after accounting for the maximum classical noise C.

Definition 7. Consider a QRNG that generates raw randomness X and let C be maximum classical noise, jointly distributed according to P_{XC}. Let P_Q be a probability distribution and let $\Omega : (q,c) \mapsto x$ be a function such that

$$P_{XC} = P_{\Omega(Q,C)C}$$

[2]The *min-Entropy* satisfies the data processing inequality: discarding side information of a system can only increase the entropy, i.e. for any two systems E and $E^{'}$, $H_{min}(X|EE^{'}) \leq H_{min}(X|E)$.

where

$$P_{\Omega(Q,C)C}(x,c) = \sum_{q\,:\,\Omega(q,c)=x} P_Q(q)P_C(c)$$

The corresponding random variable Q is called *quantum randomness*.

Example 3 (Quantum randomness of a PBS-based QRNG)**.** The quantum randomness of the PBS-based QRNG of Definition 3 may be defined as the path that the photon takes after the PBS (i.e. whether it travels to R_1 or R_2). For a single diagonally polarized photon, Q would therefore be a uniformly distributed bit. Then, for inefficient detectors with maximum classical noise T_1 and T_2 defined as in Example 2, the function $\chi : (q, t_1, t_2) \mapsto x = (x_1, x_2)$ is given by

$$\chi(q,t_1,t_2) = \begin{cases} (t_1,0) & \textit{if } \; q=1 \\ (0,t_2) & \textit{if } \; q=2 \end{cases}$$

2 POVM

A measurement of a system is one of the most important and fundamental key to quantum theory. There exist many different interpretations of a quantum measurement, most of them rely on the von Neumann's projective postulate based on the "collapsed wave function" model. This chapter presents the general definition of quantum measurement dwelling on the POVM (**P**ositive-**O**perator **V**alued **M**easure).

2.1 POVM measurement

The state of a quantum system, $|\psi\rangle$, is a vector in a complex vector space (*Hilbert space*) $\mathcal{H}$. If the set of vectors $|n\rangle$ with $n = 0, 1, 2, \ldots, N-1$ is an orthonormal basis for this space, then $|\psi\rangle$ may be expressed as:

$$|\psi\rangle = \sum_{n=0}^{N-1} c_n |n\rangle \tag{2.1}$$

for some complex coefficients c_n where $\sum_n |c_n|^2 = 1$. The quantum measurement of a system evolves from the "*von Neumann's projective postulate*":[1] [36]

Postulate. *If an ideal measurement (that minimally perturbs the system) of II kind (that if the measurement is performed, the same value is retrieved immediately after)of an observable $\mathbb{O}$ on the state $|\psi\rangle$ gives an outcome in a set Δ, immediately after the state of the system is projected on the subset of the spectrum of $\mathbb{O}$ contained in Δ.*

The elements $|\psi\rangle$ are called "*pure state*" and provide a maximal knowledge of a system. Sometimes it is not known with certainty which of the pure state the systems is, but only that it is in one of them, $(|\phi_1\rangle, |\phi_2\rangle, \cdots, |\phi_m\rangle)$ with probability $(p_1, p_2, \ldots, p_m)$. Naturally these probabilities satisfy the common relation, $\sum_m p_m = 1$. In these particular cases, it is used the more compact form, ρ, called density matrix:

$$\rho \equiv \sum_i p_i |\phi_i\rangle \langle\phi_i| \tag{2.2}$$

The previous expression is obtained from the definition of the mean of a quantum observable.

[1] It is necessary to specify that the measurements that fall in this definition belong to a more wide and generic field called PVM (**P**rojection **V**alued **M**easurements)

Let $\mathbb{O}$ be an operator of a quantum observable, the mean of $\mathbb{O}$ on the pure state $|\psi\rangle$ is:

$$\langle \mathbb{O} \rangle_{\psi} = \langle \psi | \mathbb{O} | \psi \rangle \tag{2.3}$$

For a system Σ that describes a density matrix ρ, Eq. (2.3) becomes:

$$\langle \mathbb{O} \rangle_{\Sigma} = \sum_{i=1}^{m} p_i \langle \phi_i | \mathbb{O} | \phi_i \rangle \tag{2.4}$$

If $|\phi\rangle \in \mathcal{H}$:

$$\langle \phi | \mathbb{O} | \phi \rangle = Tr\left(|\phi\rangle \langle \phi| \, \mathbb{O} \right) \tag{2.5}$$

To prove Eq. (2.5) let $\{ |\chi_j\rangle : j = 1, \ldots, \infty \}$ be a base for $\mathcal{H}$ with $|\chi_1\rangle = |\phi\rangle$. The r.h. side of Eq. (2.5) becomes:

$$\begin{aligned} Tr(|\phi\rangle \langle \phi| \, \mathbb{O}) &\equiv \sum_{j=1}^{\infty} \langle \chi_j | \left(|\phi\rangle \langle \phi| \, \mathbb{O} \right) |\chi_j\rangle \\ &= \sum_{j=1}^{\infty} \underbrace{\langle \chi_j | \chi_1 \rangle}_{\delta_{j1}} \langle \chi_1 | \mathbb{O} | \chi_j \rangle \\ &= \underbrace{\langle \chi_1 | \chi_1 \rangle}_{1} \langle \chi_1 | \mathbb{O} | \chi_1 \rangle \\ &= \langle \phi | \mathbb{O} | \phi \rangle \end{aligned} \tag{2.6}$$

From the relation expressed in Eq. (2.6), Eq. (2.4) becomes:

$$\begin{aligned} \langle \mathbb{O} \rangle_{\Sigma} &= \sum_{i=1}^{m} p_i Tr\left(|\phi_i\rangle \langle \phi_i| \, \mathbb{O} \right) \\ &= Tr\left[\sum_{i=1}^{m} p_i |\phi_i\rangle \langle \phi_i| \, \mathbb{O} \right] \\ &= Tr\left(\rho \, \mathbb{O} \right) \end{aligned} \tag{2.7}$$

From the previous statement it can be shown that the density matrix satisfies the following properties:

(i) $\rho = \rho^{\dagger}$

(ii) $\rho \geq 0$

(iii) $\mathrm{Tr}\,(\rho)$=1

The mathematical action of Hermitian matrices $\mathbb{A}$ on the orthonormal vectors $|i\rangle$ represents the physical action of quantum measurement:

$$\mathbb{A} = \sum_{k=1}^{n} a_k \mathbb{P}_k \qquad \text{with} \qquad \mathbb{P}_k = |a_k\rangle \langle a_k| \tag{2.8}$$

where $\mathbb{P}_k$ is called projective operator, or more generally projector, that can be regarded as the answer to question "*What is the probability of finding the state* $|\psi\rangle$ *in the state* $|k\rangle$". [36]

These operators satisfy the following properties:

(i) Hermitiannes: $\mathbb{P}_i = \mathbb{P}_i^\dagger$

(ii) Positiveness: $\langle n|\mathbb{P}_i|n\rangle \geq 0 \quad \text{per} \quad \forall\ |n\rangle$

(iii) Completeness: $\sum_{i=1}^{n} \mathbb{P}_i = \mathbb{1}_n$

(iv) Orthogonality and projectivity: $\mathbb{P}_i \mathbb{P}_j = \delta_{ij} \mathbb{P}_i$

This way, after making the measurement, the system will be found in the ρ_m state with probability p_m:[2]

$$p_m = \langle m|\rho|m\rangle = Tr(P_m\, \rho\, P_m) = Tr(P_m^2\, \rho) = Tr(P_m\, \rho) \tag{2.9}$$

$$\rho_m = |m\rangle \langle m| = \frac{P_m\, \rho\, P_m}{Tr(P_m\, \rho\, P_m)} = \frac{P_m\, \rho\, P_m}{p_m} \tag{2.10}$$

All these conditions hold only for an ideal system, where the observed object couples with the measurement device and *collapse* onto the experimental observed state. For a more realistic system, it is necessary to express the state through the matrix density because a non-ideal measurement provide statistical mixture used to account the non-maximal knowledge of the system itself. [37] This fact results in the definition of new projective operators $\widetilde{\mathbb{A}}$ (although they no longer meet the condition of projectivity) defined as following:

$$\widetilde{\mathbb{A}} = \sum_{k=1}^{n} \widetilde{a}_k \widetilde{\mathbb{P}}_k \qquad \text{with} \qquad \widetilde{\mathbb{P}}_k = \sum_{k} p_{j|k} \mathbb{P}_j \tag{2.11}$$

where $p_{j|k}$ stands for the conditioned probability of outcome j given k.

As the set $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_n\}$ provides a complete description of an ideal measurement, the set $\{\widetilde{\mathbb{P}}_1, \widetilde{\mathbb{P}}_2, \ldots, \widetilde{\mathbb{P}}_n\}$ provides a description of a realistic measurement.

[2] For the probability computation it has been used the cyclic property of trace: $Tr[ABC] = Tr[BCA] = Tr[CAB]$.

The new operators, however, satisfy only some of the properties of the $\mathbb{P}$ stated previously:

(i) Hermitiannes: $\widetilde{\mathbb{P}}_i = \widetilde{\mathbb{P}}_i^\dagger$

(ii) Positiveness: $\langle n|\,\widetilde{\mathbb{P}}\,|n\rangle \geq 0 \quad \text{for} \quad \forall\,|n\rangle$

(iii) Completeness: $\sum_i \widetilde{\mathbb{P}}_i = \mathbb{1}$

(iv) Generally no orthogonality and no projectivity: $\widetilde{\mathbb{P}}_i\widetilde{\mathbb{P}}_j \neq \delta_{ij}\widetilde{\mathbb{P}}_i$

As it can be seen from (iv), the new matrices are no longer projective and orthogonal ensuring that there are an infinite number of possible choices of operators $\widetilde{\mathbb{P}}$ through which define $\widetilde{\mathbb{A}}$. The matrices $\{\mathbb{P}_1, \mathbb{P}_2, \ldots, \mathbb{P}_n\}$ are so called POVM (**P**ositive-**O**perator **V**alued **M**easures) and the generalized measurement "POVM measurement". Performing a measurement on the system through a POVM it is obtained the state $\widetilde{\rho}_m$ with probability $\widetilde{p}_m$:

$$\widetilde{\rho}_m = \frac{1}{\widetilde{p}_m}\,\widetilde{\mathbb{P}}_k^\dagger\,\rho\,\widetilde{\mathbb{P}}_k \qquad \text{with} \qquad \widetilde{p}_m = Tr\left[\widetilde{\mathbb{P}}_k^\dagger\,\rho\,\widetilde{\mathbb{P}}_k\right] \tag{2.12}$$

These definitions however do not account that the matrices $\widetilde{\mathbb{P}}$ are no longer projective and that the probabilities $\widetilde{p}_m$ do not sum to one:

$$\sum_k Tr\left[\widetilde{\mathbb{P}}_k^\dagger\,\rho\,\widetilde{\mathbb{P}}_k\right] = \sum_k Tr\left[\widetilde{\mathbb{P}}_k^2\,\rho\right] \neq Tr\left[\sum_k \widetilde{\mathbb{P}}_k\,\rho\right] = Tr(\rho) = 1$$

The simplest method to change that is to define some "*measurement operators*" $\mathbb{M}_k$ that linearly act on the system's space through which redefine $\widetilde{\mathbb{P}}$:

$$\widetilde{\mathbb{P}}_k = \mathbb{M}_k\mathbb{M}_k^\dagger \tag{2.13}$$

This way Eq. (2.12) becomes:

$$\widetilde{\rho}_m = \frac{1}{p_m}\,\mathbb{M}_k^\dagger\,\rho\,\mathbb{M}_k \qquad \text{with} \qquad p_m = Tr\left[\mathbb{M}_k^\dagger\,\rho\,\mathbb{M}_k\right] \tag{2.14}$$

where:

$$\sum_k Tr\left[\mathbb{M}_k^\dagger\,\rho\,\mathbb{M}_k\right] = \sum_k Tr\left[\mathbb{M}_k\,\mathbb{M}_k^\dagger\,\rho\right] = Tr\left[\sum_k \widetilde{\mathbb{P}}_k\,\rho\right] = Tr(\rho) = 1$$

It is remarkable to notice that a POVM measurement repeated on the system does not reproduce the same result, unlike a PVM measurement:

$$Tr\left(\mathbb{M}_j^\dagger\,\rho_k\,\mathbb{M}_j\right) = \frac{Tr\left(\mathbb{M}_j^\dagger\mathbb{M}_k^\dagger\,\rho\,\mathbb{M}_k\mathbb{M}_j\right)}{Tr\left(\mathbb{M}_k^\dagger\,\rho\,\mathbb{M}_k\right)} \neq \delta_{kj} \tag{2.15}$$

because generally $\mathbb{M}_k\mathbb{M}_j \neq \delta_{k,j}\mathbb{M}_k$.

The relation between PVM and POVM thus becomes evident: it is possible to trace back to the theory of PVM as a special case of the POVM theory. Of course there are many contexts in which the PVM measurements generate spontaneously POVM measurements in spaces of reduced size. Also thanks to the Neumrak's theorem it can be shown that POVM measurements can be realized as PVM measurements in a space of greater dimension. [37]

3 Photocount statistic

The task of quantum optics is to study the particle nature of a beam of light through the usage of the "*photons*" concept. The basic definition of *photon* is that of energy quantization of a classical electromagnetic wave with which define a beam of light. This chapter, based on "*Quantum Optics - An introduction*" by Mark Fox [38], presents a semi-classical, and quantum, approach to study the statistical properties of a photons' flux. The quantum section will also show the rudimentary concepts of the quantization of the electromagnetic field and the notations used to analyze photonic systems employing "*The Quantum Theory of Light*" by Rodney Loudon [39].

3.1 Semi-classical theory of photodetection

Let's consider a PCD (**P**hoton-**C**ounting **D**etector) such as a photomultiplier illuminated by a light beam. The light interacts with the photocathode's atoms liberating individual electrons by the photoelectric effect. These single photoelectrons then trigger a sequence of secondary avalanches generating a current pulse sufficiently intense to be detected with an electronic counter. The statistical nature of the timing between the output pulses can then be explained by making the following three assumptions about the photodetection process:

1. The probability of the emission of a photoelectron in a short time interval Δt is proportional to the intensity I, the area A illuminated and the time interval Δt
2. If Δt is sufficiently small, the probability of emitting two photoelectrons is negligibly small
3. Photoemission events registered in different time intervals are statistically independent of each other

The probability of observing one photoemission event in the time interval $[t, t + \Delta t]$, from assumption (1), is:

$$P(1, t, t + \Delta t) = \xi I(t) \Delta t \tag{3.1}$$

where ξ is proportional to the area illuminated and is equal to emission probability per unit time and intensity while $I(t)$ is the incident radiation intensity.

Assuming that the probability of emitting two photoelectrons in the same time interval is negligibly small (assumption (2)), is the same as stating the following expression:

$$P(0,t,t+\Delta t) = 1 - P(1,t,t+\Delta t) = 1 - \xi I(t)\Delta t \tag{3.2}$$

Asking that the events are to be statistically independent of each other (assumption (3)) means that the probability of detecting n events in the time interval $[0, t+\Delta t]$ is the same as saying:

$$P(n,0,t+\Delta t) = P(n,0,t)P(0,t,t+\Delta t) + P(n-1,0,t)P(1,t,t+\Delta t) \tag{3.3}$$

where $P(n,0,t)$ and $P(n-1,0,t)$ correspond to the probabilities of having n and *n-1* events in the time interval $[0,t]$ respectively. Changing the previous notation as, $P(n,0,t) = P_n(t)$, Eq. (3.3) may be rewritten in the more compact form:

$$\frac{P_n(t+\Delta t - P_n(t))}{\Delta t} = \xi I(t)\left[P_{n-1}(t) - P_n(t)\right] \tag{3.4}$$

where in the limit $\Delta t \to 0$ becomes:

$$\frac{dP_n(t)}{dt} = \xi I(t)\left[P_{n-1}(t) - P_n(t)\right] \tag{3.5}$$

Supposing that the light intensity is constant, ($\xi I(t) = const. = C$), that is, assume the radiation perfectly coherent, Eq. (3.5) becomes:

$$\frac{dP_n(t)}{dt} + CP_n(t) = CP_{n-1}(t) \tag{3.6}$$

with the following condition $P_0(0) = 1$.
The equation's solution con be subdivide into two terms: $n = 0$ and $n \geqslant 1$. When $n = 0$ necessarily $P_{n-1}(t) = 0$ because registering a negative number of events has no meaning. The solution then is:

$$P_0(t) = \exp\left(-Ct\right) \tag{3.7}$$

with the same boundary condition $P_0(0)$=1.
For the $n \geqslant 1$ case, Eq. (3.6) is then multiply by the integrating factor $\exp(Ct)$ to obtain:

$$\frac{d}{dt}\left(e^{Ct}P_n(t)\right) = Ce^{Ct}P_{n-1}(t) \tag{3.8}$$

which, on integrating, gives:

$$P_n(t) = e^{-Ct}\int_0^t Ce^{Ct'}P_{n-1}(t')dt' \tag{3.9}$$

To solve Eq. (3.9), it is used a recursively method employing $P_0(t)$ given by Eq. (3.7):

$$\begin{aligned}
P_1(t) &= e^{-Ct}\int_0^t Ce^{Ct'}P_0(t')dt' = (Ct)\,e^{-Ct} \\
P_2(t) &= e^{-Ct}\int_0^t Ce^{Ct'}P_1(t')dt' = \frac{(Ct)^2}{2}e^{-Ct} \\
P_3(t) &= e^{-Ct}\int_0^t Ce^{Ct'}P_2(t')dt' = \frac{(Ct)^3}{3!}e^{-Ct} \\
&\vdots \\
P_n(t) &= e^{-Ct}\int_0^t Ce^{Ct'}P_{n-1}(t')dt' = \frac{(Ct)^n}{n!}e^{-Ct}
\end{aligned} \tag{3.10}$$

When the intensity is not constant, the term Ct has to be replaced by $\int_0^t \xi I(t')dt'$. It is possible to cast Eq. (3.10) into a simpler form noticing that the event probability per unit time is equal to $\xi I(t)$, thus, if the intensity is constant, the mean count rate $\overline{n}$ in the time interval $[0,t]$ is just given by:

$$\overline{n} = \xi I t = Ct \tag{3.11}$$

Hence Eq. (3.10) assumes the following form:

$$P_n(t) = \frac{\overline{n}^{\,n}}{n!}e^{-\overline{n}} \tag{3.12}$$

which shows that, for a constant $I(t)$, it is obtained a Poissonian distribution. Eq. (3.12) demonstrates that it is possible to explain the photocount statistic through the Poissonian distribution without invoking concepts like photons or quantum field radiation. [38] The fundamental demand is that the emission of photoelectrons has to be a probabilistic process triggered by the absorption of a quantum of energy from the light beam. At the same time, it is clear that this approach does not show a photon statistic of any kind. Only with a quantum approach it is possible to relate the two concepts.

3.2 Quantum theory of photodetection

3.2.1 Quantization of the electromagnetic field

The electromagnetic field is associated to a quantum system of a three-dimensional harmonic oscillators. It is possible to simplify the notation narrowing the notation to a one-dimension system. This way, the quantum-mechanical Hamiltonian for a one-dimensional harmonic oscillator, with frequency ω, is:

$$\hat{H} = \frac{\hat{p}^{\,2}}{2m} + \frac{1}{2}m\omega^2\hat{q}^{\,2} \tag{3.13}$$

where the position and momentum operators, $\hat{q}$ and $\hat{p}$ respectively, satisfy the commutation relation:

$$[\hat{q}, \hat{p}] = i\hbar$$

Replacing the operators $\hat{q}$ and $\hat{p}$ with two dimensionless operators, $\hat{a}$ and $\hat{a}^\dagger$, Eq. (3.13) becomes:

$$\hat{H} = \hbar\omega \left(\hat{a}^\dagger \hat{a} + \frac{1}{2}\right) \tag{3.14}$$

where $\hat{a}$ and $\hat{a}^\dagger$ corresponds to the destruction and creation operators respectively, for the harmonic oscillator. They satisfy the following commutation relation:

$$\left[\hat{a}, \hat{a}^\dagger\right] = 1 \tag{3.15}$$

Let $|n\rangle$ be an energy eigenstate with eigenvalue E_n of the Hamiltonian operator. The eigenvalue equation is:

$$\hat{H}\,|n\rangle = \hbar\omega \left(\hat{a}^\dagger \hat{a} + \frac{1}{2}\right)|n\rangle = E_n\,|n\rangle \tag{3.16}$$
$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \qquad\qquad n = 0, 1, 2\,...$$

where n is the eigenvalue of the "*particle number*" operator $\hat{n} = \hat{a}^\dagger \hat{a}$. This way, $|n\rangle$ is also called "*number state*":

$$\hat{n}\,|n\rangle = n\,|n\rangle$$

Given the previous notation, suppose that the radiation is confined inside a cubic cavity of edge L. The vector-potential, $\hat{A}_{\mathbf{k}\lambda}$, may be expressed as a linear combination of permitted oscillating modes that satisfy the periodic boundary condition (bold font indicate a three-dimensional variable):

$$\hat{A}_{\mathbf{k}\lambda} = \sum_{\mathbf{k}} \sum_{\lambda=1,2} e_{\mathbf{k}\lambda} \hat{A}_{\mathbf{k}\lambda}(\mathbf{r}, t) \tag{3.17}$$
$$k_i = 2\pi \frac{n_{i,l}}{L} \qquad i = x, y, z \text{ ed } n_l = 0, \pm 1, \pm 2, \pm 3 \cdots$$

where $e_{\mathbf{k}\lambda}$ is the spatial unit vector, $\mathbf{k}$ is the wavevector representing the radiation while λ is a polarization index that specify if the radiation is vertically (1) or horizontally (2) polarized. It is fundamental to stress out that Eq. (3.17) does not refer to classical function but to a quantum field operator. It is common to express $\hat{A}_{\mathbf{k}\lambda}$ with the destruction and creation operators as to simplify the notation.

Naturally, both operators now have to follow a more general commutation relation:

$$\left[\hat{a}_{\mathbf{k}\lambda}, \hat{a}^{\dagger}_{\mathbf{k}'\lambda'}\right] = \delta_{\mathbf{k},\mathbf{k}'}\,\delta_{\lambda,\lambda'}$$

Eq. (3.17) thus becomes:

$$\hat{A}_{\mathbf{k}\lambda} = \sqrt{\frac{\hbar}{2\epsilon_0 V \omega_k}}\left[\hat{a}_{\mathbf{k}\lambda}\exp\left(-i\omega_k t + i\mathbf{k}\cdot\mathbf{r}\right) + \hat{a}^{\dagger}_{\mathbf{k}\lambda}\exp\left(i\omega_k t - i\mathbf{k}\cdot\mathbf{r}\right)\right] \tag{3.18}$$

Using Eq. (3.14) and (3.18), the Hamiltonian operator $\hat{H}$ takes the following expression:

$$\hat{H}_R = \sum_{\mathbf{k}}\sum_{\lambda}\hbar\omega\left(\hat{a}^{\dagger}_{\mathbf{k}\lambda}\hat{a}_{\mathbf{k}\lambda} + \frac{1}{2}\right) \tag{3.19}$$

where the R stands for "Radiation". [39]

3.2.2 Photodetection

According to what said in Subsec. (3.2.1), there are two fundamentals physical quantities that quantum optics uses to relate photodetection properties and a source of light radiation: the mean and variance of incident photons flux. [39] Before going into details it is necessary to introduce the "*photons number*" concept detected by a device. For a detector that integrates the incident photons flux over a time interval T, the photons number may be computed with the following operator:

$$\hat{M}(t,T) = \int_{t}^{t+T} dt'\,\hat{f}(t') = \int_{t}^{t+T} dt'\,\hat{a}^{\dagger}(t')\hat{a}(t') \tag{3.20}$$

Normal system are of course not ideal. There exist many phenomenons that prevent a total conversion of incident photons in "*photocounts*". That is why it is introduced a parameter that takes into account all the "imperfections" of the device: quantum efficiency. A simple model that shows the meaning of quantum efficiency is the one showed in Fig. (3.1). An ideal detector is preceded by a BS (**B**eam **S**plitter) with transmission and reflective coefficients:

$$\mathscr{R} = i(1-\eta)^{1/2} \qquad \text{and} \qquad \mathscr{T} = \eta^{1/2} \tag{3.21}$$

so to satisfy the following conditions:

$$|\,\mathscr{R}\,|^2 + |\,\mathscr{T}\,|^2 = 1$$
$$\mathscr{R}\mathscr{T}^* + \mathscr{R}^*\mathscr{T} = 0$$

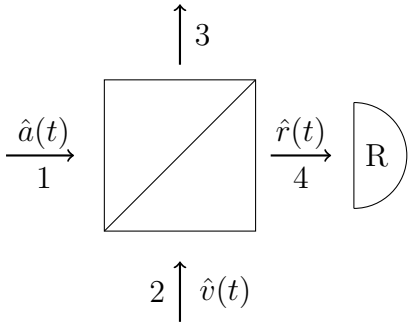


Figure 3.1: Simple scheme of an inefficient photodetector showing the input field $\hat{a}(t)$ entering on the left, the beam splitter with quantum efficiency η and the ideal detector R.

Referring to Fig. 3.1, the radiation, parameterized by the operator $\hat{a}(t)$, enters in arm 1. The operator $\hat{v}(t)$,representing the continuum vacuum state, enters in arm 2. The detector R registers only the radiation coming out in arm 4, $\hat{r}(t)$, while the remaining is lost in arm 3. This way, $\hat{r}(t)$ may be rewritten using Eq. (3.21):

$$\hat{r}(t) = \eta^{1/2}\,\hat{a}(t) + i(1-\eta)^{1/2}\,\hat{v}(t) \tag{3.22}$$

Similarly to Eq. (3.20), the new photocount operator takes the same form except for the fact that this time the incoming field operator is $\hat{r}(t)$:

$$\hat{M}_R(t,T) = \int_t^{t+T} dt^{'}\hat{r}^{\dagger}(t^{'})\hat{r}(t^{'}) \tag{3.23}$$

The "*photocounts*" mean then becomes:

$$\langle m\rangle = \left\langle \hat{M}_R(t,T)\right\rangle = \eta\left\langle \hat{M}(t,T)\right\rangle \tag{3.24}$$

where m identifies the counts number and the angle brackets the quantum expectation value. The second factorial momentum than is:

$$\begin{aligned}\langle m(m-1)\rangle &= \left\langle \hat{M}_D^2(t,T)\right\rangle \\ &= \int_t^{t+T} dt^{'}\int_t^{t+T} dt^{''}\left\langle \hat{r}^{\dagger}(t^{'})\hat{r}^{\dagger}(t^{''})\hat{r}(t^{''})\hat{r}(t^{'})\right\rangle\end{aligned} \tag{3.25}$$

As a consequence of the model, the relations between the mean value and second momentum of the "*photocounts*" and the photons number of the incident light are the same as that of a standard BS.

This way, the expression for the photocount distribution variance is:

$$(\Delta m)^2 = \eta^2 \left\langle \Delta \hat{M}^2(t,T) \right\rangle + \eta(1-\eta) \left\langle \hat{M}(t,T) \right\rangle \tag{3.26}$$

where the first contribution on the r.h.s is the variance of the photons number of the radiation hitting the detector, represented by the operator in Eq. (3.20) scaled by the square quantum efficiency. The second contribution instead is a "*partition noise*" resulting from the random selection of a fraction η of the incident photons by the defective detector.

4 System's noise models

For a more realistic model, it has to be taken into account all the perturbation effects that modify the analysis of a physical system. In common light radiation detection using SPADs, three are the principal phenomenons that characterize the electric noise of a detector: *optical crosstalk*, *dark count* and *afterpulsing*. The chapter introduces the concepts of *crosstalk*, *dark count* and *afterpulsing* dwelling on the methods used to defined them.

4.1 Optical Crosstalk

Single photon avalanche detectors, SPADs, fall under the category of silicon-based photomultipliers (SiPMs). Every component or pixel, of the SPAD array that makes the detector (see Sec. (5.2)), produces an impulse of constant amplitude. The pulse is then registered, analogically or digitally, through the same output channel giving a signal proportional to the sum of the single pixels. Unfortunately, the count resolution is strongly limited by the crosstalk effect which modifies the linear response of the device. [40] This perturbation then produces an increase of the electronic noise. When a pixel creates a signal, thanks to a detected photon or a thermal generation, the carriers induce a secondary emission of IR (**I**nfra-**R**ed) photons that may subsequently creates other hot carriers in a near region. Fig. (4.1) shows an example of this kind of process.

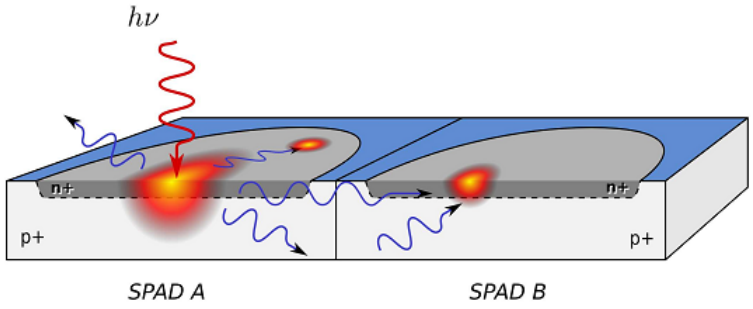
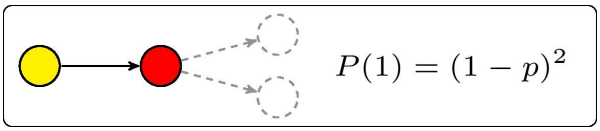
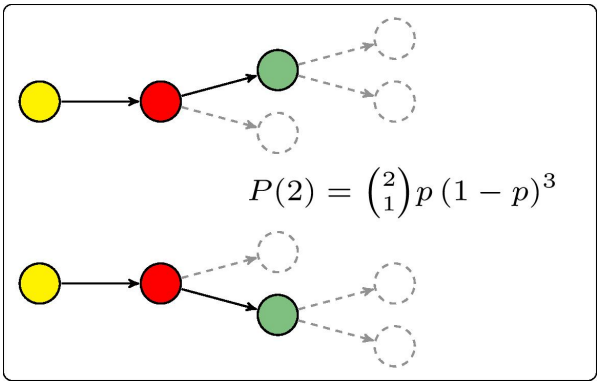


Figure 4.1: Example of optical crosstalk between two detectors, A and B. When a primary photon starts an avalanche on the A SPAD, IR photons may be emitted and propagate through the detector. If these photons interact with the B SPAD they may generate a secondary avalanche.

This stochastic process is called optical crosstalk. The phenomenon occurs almost instantaneously with probability proportional to the SiPM gain. The probability, ε, is usually defined as the ratio of the dark counts that have generated crosstalk and the total dark counts rate. To calculate ε it has been adopted the geometric model developed by Spanish researchers (L. Gallego, J. Rosado, F. Blanco, F. Arqueros - *"Modeling crosstalk in silicon photomultipliers"* [41], which includes sequential avalanche generations effects.



(a) Probability for 1 successful crosstalk event (red filled circle).



(b) Probability for 2 successful crosstalk event (red and green filled circle).

Figure 4.2: Crosstalk histories for a system of 2 neighbors. Given a crosstalk event (yellow filled circle), the primary pixel (red filled circle) generates a sequence of crosstalk events (green filled circle). If a crosstalk events does not generates another one (dashed grey circle), the cascading sequence stops and the corresponding probability is calculated in terms of multinomial combinations.

The model defines a binomial distribution to describe s successful events of crosstalk for a set of n neighbors of the primary pixel:[1]

$$P(s) = \binom{n}{s} p^s (1-p)^{n-s} \tag{4.1}$$

where p indicates the probability that a primary avalanche may induces a secondary one in an adjacent pixel. This parameter is related to ε as following:

$$P(1) = (1-p)^n = 1 - \varepsilon \tag{4.2}$$

[1]The distribution reduces to a Bernoulli distribution for $n = 1$ and tends to a Poissonian with mean $\lambda = -\log(1-\varepsilon)$ for $n \to \infty$ keeping ε constant.

The probability of successful crosstalk events for n neighbors pixel is obtained applying repeatedly Eq. (4.1) for each pixel of the detector assuming that a primary (or secondary, tertiary...) pixel may generate a successful crosstalk event. Through these assumptions, the probability of obtaining j crosstalk events is proportional to the number of "histories" with $j-1$ crosstalk events. All the records are obtained considering all the possible combination of "on and off" neighbors. Fig. (4.2) schematically represents the process up to a maximum of 2 crosstalk events for a model of $n=2$ neighbors. For an arbitrary number of n neighbors and j crosstalk events, the probability is:

$$P(j) = h_{n,j-1} p^{j-1} (1-p)^{jn-j+1} \tag{4.3}$$

where $h_{n,j-1}$ defines the numbers of crosstalk "histories" summed up to the $(j-1)$-th for n neighbors. Although calculating $h_{n,j-1}$ is computationally hard for a large n and j, it is possible to generalized the formula with the following recursive method:

$$h_{n,e} = \sum_{l=1}^{e} \binom{n}{l} \sum_{i_1=0}^{e-l} h_{n,i_1} \sum_{i_2=0}^{e-l-i_1} h_{n,i_2} \quad \cdots \quad \sum_{i_{l-1}=0}^{e-l-i_1-\cdots-i_{l-2}} h_{n,i_{l-1}} h_{n,e-l-i_1-\cdots-i_{l-1}} \tag{4.4}$$

assuming $h_{n,0} = 1$.

It is remarkable to notice that the inclusion, or not, of avalanches effects does not influence the total probability of crosstalk, ε. As a matter of fact, the probability relies only on the original binomial distribution that a primary pixel will not induce any crosstalk event. [41]

4.2 Dark Count

SPADs are detectors that operate biased at a voltage V_A well above breakdown voltage V_B for an amount of time sufficient to guarantee the creation of an avalanche. At this bias, the electric field is so high that a single charge carrier injected into the junction can trigger a self-sustaining avalanche [42]. The current rises rapidly to a steady level in the milliampere range ($\sim mA$)[2]. Ideally, the system should not produce any signal in "dark regimen"[3]. In real devices, hot carriers may induce a hole-electron pair which may lead to a signal generation because of the Geiger working mode of SPADs. When this situation occurs, the *fake* signal is misinterpreted as a good one offsetting the whole analysis. The dark count rate, μ, representing the counts number of dark events per unit time, quantifies the phenomenon. Assuming the rate to be portrayed by a Poisson statistic, the rate is estimated as following:

$$\mu = -\frac{1}{\tau} \ln\left(1 - \frac{N_C}{N_G}\right) \tag{4.5}$$

[2]The amplitude of the signal reaches this range because the detectors are operated in Geiger mode with a gain factor, M_g, of order $10^6 - 10^8$.

[3]No incoming radiation.

where N_C is the mean number of dark count, N_G the total frames number for a single acquisition and τ the time of integration. It is important to stress out that the dark count rate is heavily influenced by the working temperature of the device. Working at low temperatures reduce the hot carriers generation (hole-electron pairs) responsible for the *fake* produced signal. Fig. (4.3) depicts a dark count rate-temperature relation for a SPAD.

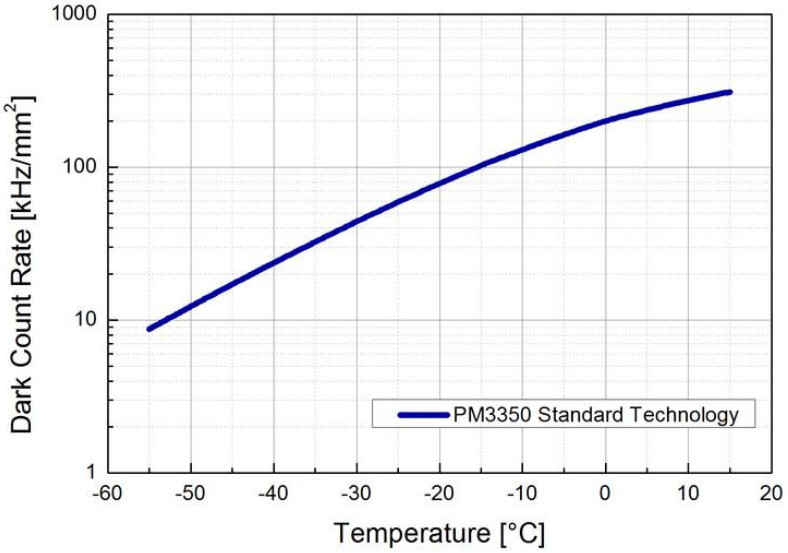


Figure 4.3: Example of the relationship between dark count rate and working temperature. As the temperature lowers, the rate decreases becoming zeros at $T = -273.15\ ^{\circ}C$. This graphic shows the equivalent form of the dark count rate expressing it as counts number per unit time per squared millimeter, $[^{kHz}/_{mm^2}]$. (*Courtesy of KETEK GmbH, München*)

4.3 Afterpulsing

The Afterpulsing phenomenon falls under the "dark count second generation" definition. During the avalanche process some carriers (hole or electrons) may sink to deep levels present in the depletion region of the *p-n* junction subsequently released at random. The time interval between releases is a stochastic process around a common mean value that depends on the characteristics of the deep levels. These carriers may then produce secondary avalanches correlated to the primary one thus causing an afterpulsing effect. [43] The number of captured carriers, during an impulse, increase as the total number of carriers that cross the junction increases. As a result, the afterpulsing amplifies as the delay time of the avalanche quenching and the applied voltage increase.

Usually, these features are correlated to the quantum efficiency or the time resolution needed for a particular device. [42]

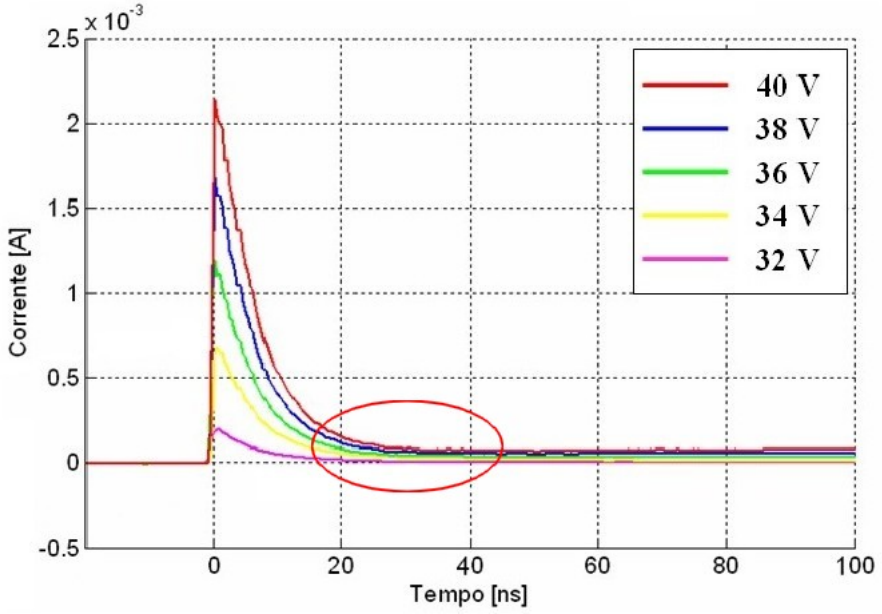


Figure 4.4: Example of the relationship between the amplitude of a signal generated by an avalanche and the reading time. The red circled zone indicates where the time correlation of signal may occur. As the applied voltage increases, the system can not rapidly quench the signal resulting in the creation of a secondary avalanche correlated to the first one.

Fig. (4.4) shows a typical trend of the relationship between the output current and the reading time of a photomultipliers' signal. When the applied voltage[4] raises, the system's quenching can no longer stop the avalanche thus inducing a possible secondary avalanche.

[4]The voltage has to be greater than the breakdown limit otherwise a photoelectron can not trigger an avalanche process.

5 Setup

This chapter presents the experimental set-up used for the photon analysis of the system. It will also be introduced the main characteristics of the different light sources employed and those of the photodetector. The last part illustrates the arrangements taken for the data acquisition.

5.1 Light sources: LASER and LED

The LASER sources employed is an LD (**L**aser **D**iode) semiconductor with a near infrared spectrum emission ($\lambda = 808\,nm$). The system connects to a **S**ingle **M**ode (SM) optic fiber through which the radiation propagates. The standard output of the fiber allows it to be attached to many anchoring devices on the optical bench. The LASER is operated by an LDC (**L**aser **D**iode **C**ontroller) that provides extreme precision at the current level of operation and absolute control over working temperature. Fig. (5.1) shows the LASER source while Fig. (5.2) shows the technical specifications of construction.

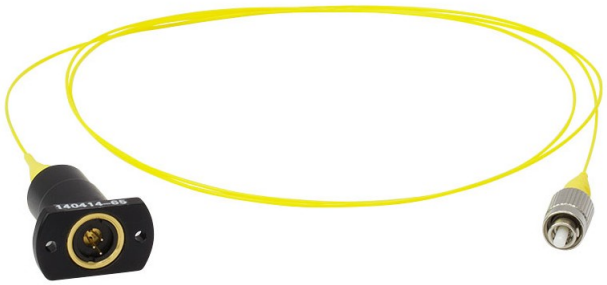


Figure 5.1: The LD semiconductor of operative wavelength $\lambda = 808\,nm$.

The LED (**L**ight **E**mitting **D**iode) source used is made of a simple p-n junction (see Fig. (5.3)) using gallium nitride (GaN) as semiconductor which is commonly employed for blue light spectrum emission. The reason for the usage of this kind of light source lies in the fact that the quantum efficiency of the detector reaches a maximum value in the wavelength interval (420 - 490 nm) (see Fig. (5.7)). The wavelength of the LED source is $\lambda = 470\,nm$. Fig. (5.4) shows the functioning LED source.

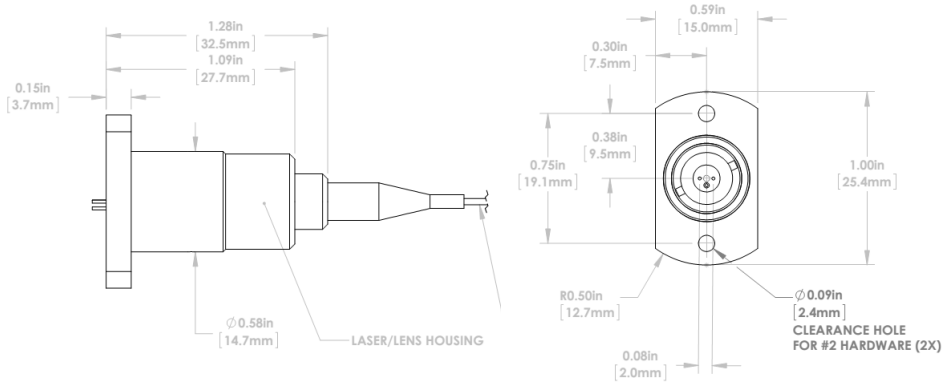


Figure 5.2: Technical specifications of the LD semiconductor.

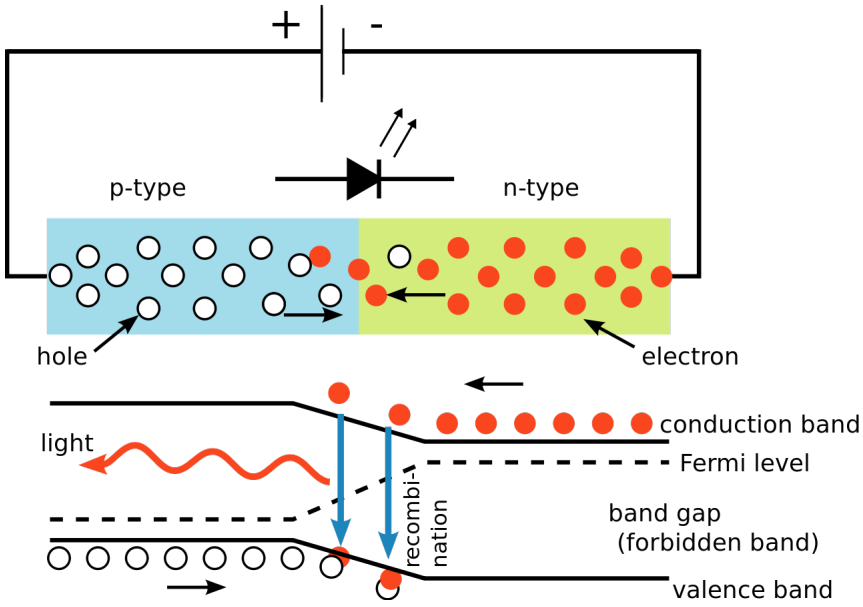


Figure 5.3: Schematic representation of a p-n junction used for a LED source.

5.2 Detector

For the analysis and detection of photons emitted by the light sources it has been used a single photon detector camera, SPC^2 gently provided by the Micro Photon Devices S.r.l. of Bolzano (see Fig. (5.5)). [44] The camera is based on a 2-D imaging array of 32x32 smart pixels (see Fig. (5.6)). Each pixel includes an SPAD (**S**ingle **P**hoton **A**valanche **D**iode) detector, an AFE (**A**nalog **F**ront-**E**nd) and a digital processing electronics. This configuration allows a "*single photon*" sensitivity, high electronic noise immunity and a fast readout speed of the signal.

Figure 5.4: The LED source of operative wavelength $lambda = 470\,nm$.

The system operates at a maximum frame rate of 49000 fps with negligible interframe dead time. This camera varies from the conventional CCD (**C**harge-**C**oupled **D**evice) or CMOS (**C**omplementary **M**etal **O**xide **S**emiconductor) sensors for it performs an entirely digital acquisition of the light signal. Each pixel effectively counts the number of photons that are detected by the sensor during the acquisition time. It features an high photon detection efficiency in the visible spectral region and a low dark count rate even at room temperature (see Fig. (5.7)).

Figure 5.5: SPC^2 detector.

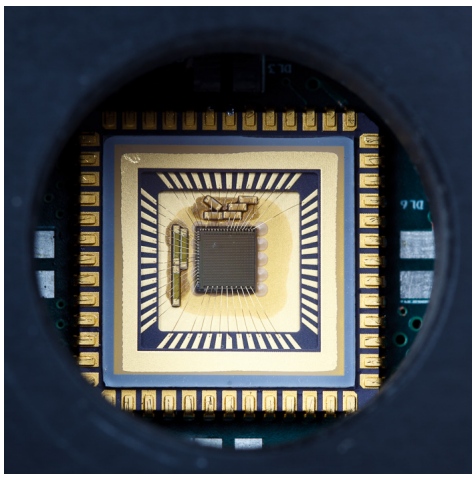


Figure 5.6: Detection region of the SPC^2 camera. The zone is an array of 32x32 smart pixels each including an SPAD.

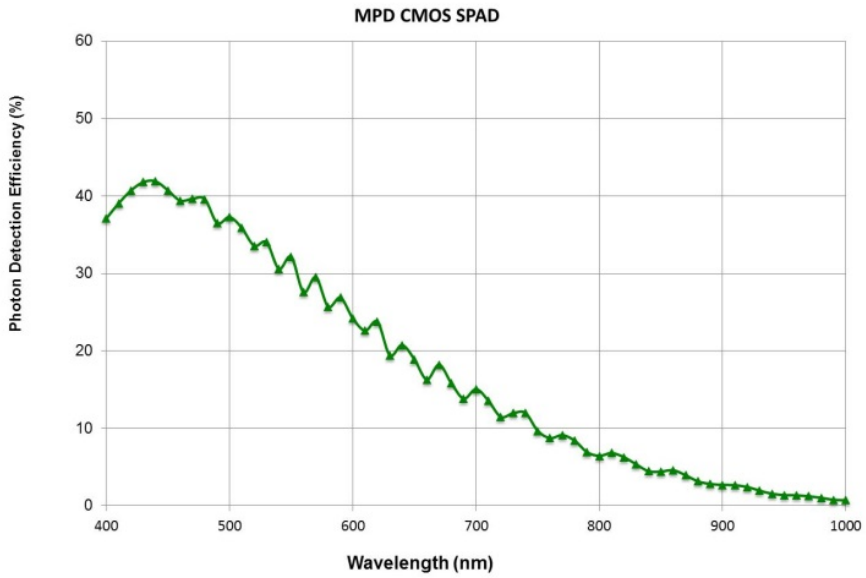


Figure 5.7: Relation between the SPC^2 quantum efficiency and the wavelength of the incident light beam.

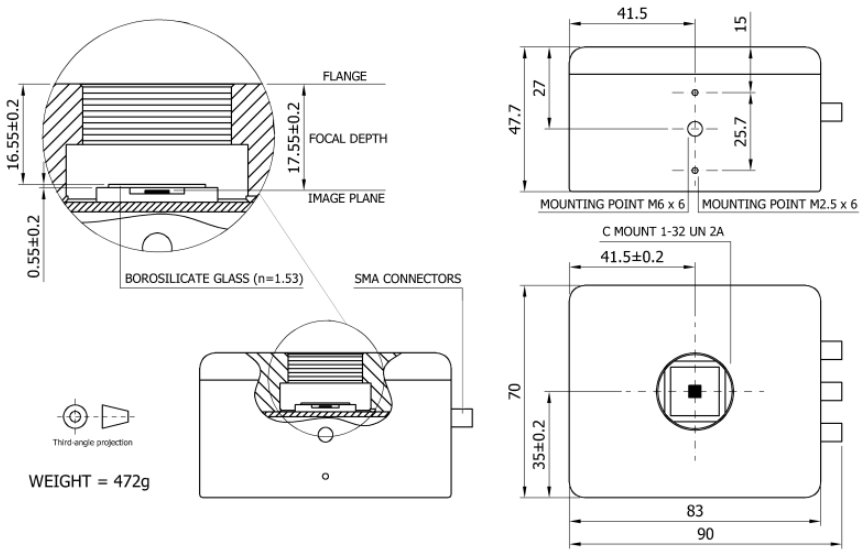


Figure 5.8: Technical draws of the SPC^2 camera.

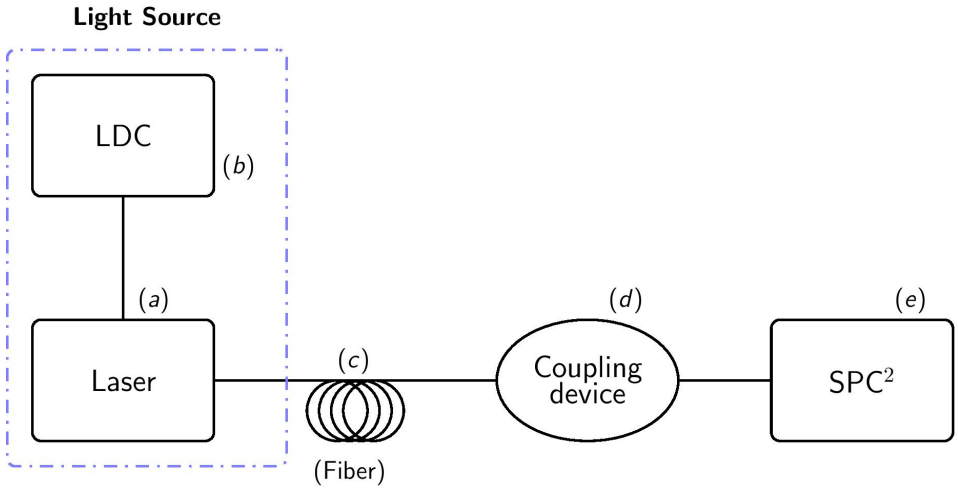


Figure 5.9: Schematic representation of the acquisition data configuration for the LASER source.

5.3 Acquisition data configuration

Each data acquisition depends mainly on two simple parameters: the working power of the light source (LASER or LED) and the relative distance between the detector and the source itself. The working power allows to retrieve information about the photon distribution of the system (Gaussian beam for the LASER, uniform distribution for the LED). The relative distance gives information about the light beam behavior and dynamic after the free space propagation (the space dependence is express by the z parameter for the LASER (see Eq. (6.1)) and it is already accounted for in the A_r parameter of Eq. (6.8) for the LED source). Fig. (5.9) reports a schematic representation of the acquisition data configuration adopted for the LASER source. Referring to the figure: the LASER source (a), connected to the LDC (b), generates the radiation that is collected into the optical fiber (c). The fiber is then secured to a coupling device (d) directly attached to the optical bench on the same axis as of the camera's (e). The coupling device is a translator with micrometer drives that permit to translate the fiber transversely and parallel to the optic bench manually adjusting the relative distance with the camera. Fig. (5.10), instead, shows the acquisition data configuration for the LED source. For this light source, the LED (a) is powered by a voltage generator (b) and directly secured on the coupling device (d) always in axis with the camera (e).

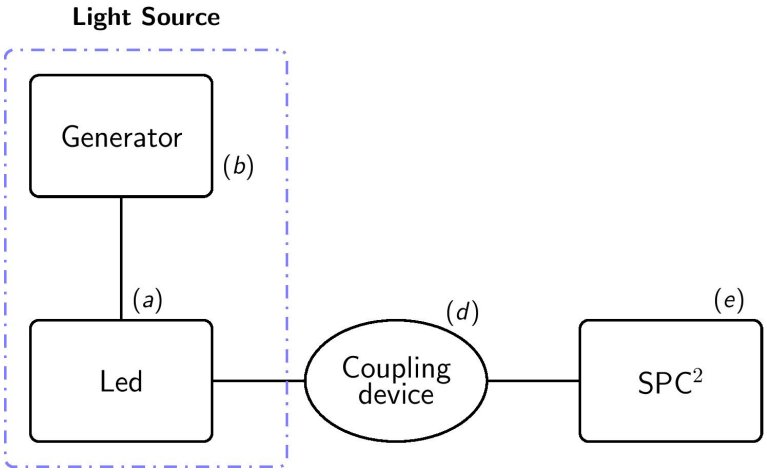


Figure 5.10: Schematic representation of the acquisition data configuration for the LED source.

6 Quantum efficiency

To develop a QRNG model, it is necessary to possess a maximal knowledge about the photodetection device used for the photons analysis. Some of its features, in primis quantum efficiency, play a fundamental role for a proper study of a system. The first part of this chapter shows the analysis made to parameterize the LASER and LED sources. The middle and last part covers a brief return to quantum photodetection theory and its relationship with the quantum efficiency of the detector.

6.1 Detector analysis

The SPC^2 camera includes a software application that allows the user to modify the acquisition settings. Some of these configurations are: variable integration time, system's dead time, (thermal and electric) background subtraction algorithm. The parameters were chosen so that each pixel would not detect more than one photon per frame. This way the system precisely represented the theoretical model expressed in Chap. (1). It is important to point out that the settings allow a "single photon detection" for each working power or distance configuration between the camera and the light sources. Working in "sampling mode", the camera acquired a total of 131043 frames (see Fig. (6.1)).

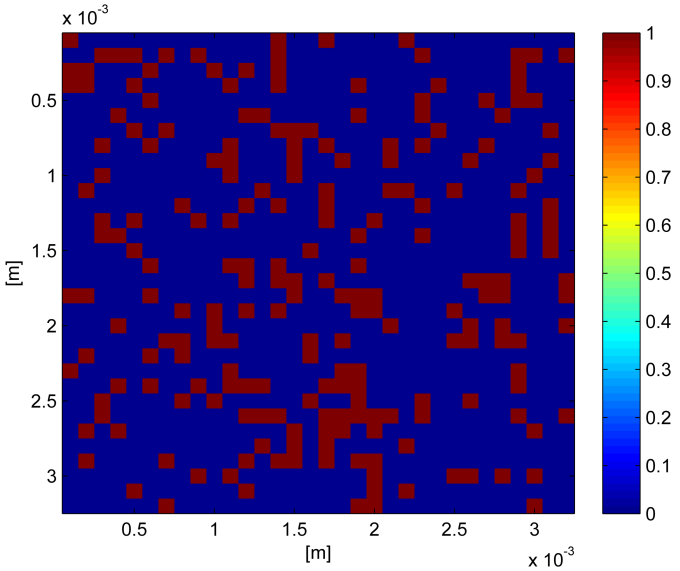
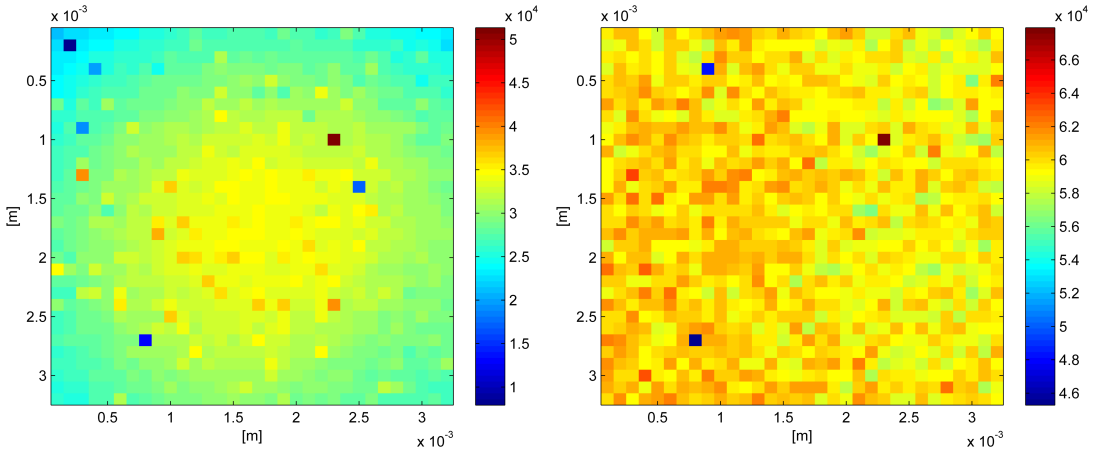


Figure 6.1: A single frame acquisition. The red-filled entries corresponds to a single detection while the blue-filled to none.

The integration time for each frame was of 200 ns for a total acquisition time of 2,67 s. All the frames were then saved as TIFF file subsequently converted in matrices of 32x32 entries to faithfully reproduce the camera.[1] As introduced in Sec. (5.3) the measurements were made for different LASER and LED working powers and distance configurations together with a background acquisition. Summing all the single frames reproduces the light beam intensity profile. For the LASER source the image resembles a TEM^{00} because of the Gaussian shape of the LASER output while an almost constant distribution for the LED. The images in Fig. (6.2) shows saturation effects due to the presence of some pixels that generate noise (thermal, electronic, crosstalk, afterpulsing, dark count, etc. . .).

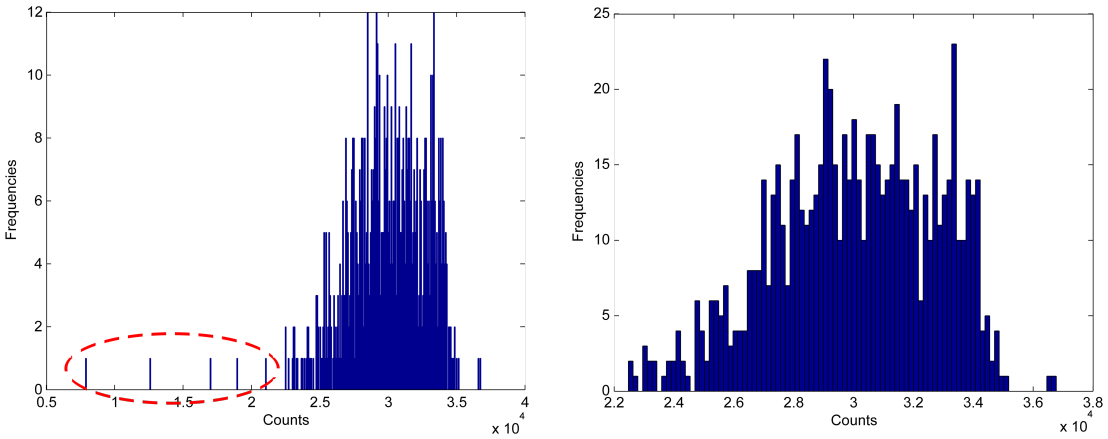


(a) LASER intensity profile (b) LED intensity profile

Figure 6.2: Images of all frames summed up. For the LASER source it is evident the Gaussian shape of the intensity profile while for the LED an almost constant distribution.

To account for these pixels it has been made an histogram using the background samples (see Fig. (6.3)). The x axis shows the single pixel counts while the y axis shows the frequencies with which they occur. The red circled zone in Fig. (6.3a) highlights the pixels that deviate from the mean value more than 3σ. Tab. (6.1) reports the selected pixels. All of them were then compared with the one extracted from all the other acquisitions so to mark the ones that were present in every sample.

[1] See Sec. (5.2) for technical specifications.



(a) Distribution before discarding pixels

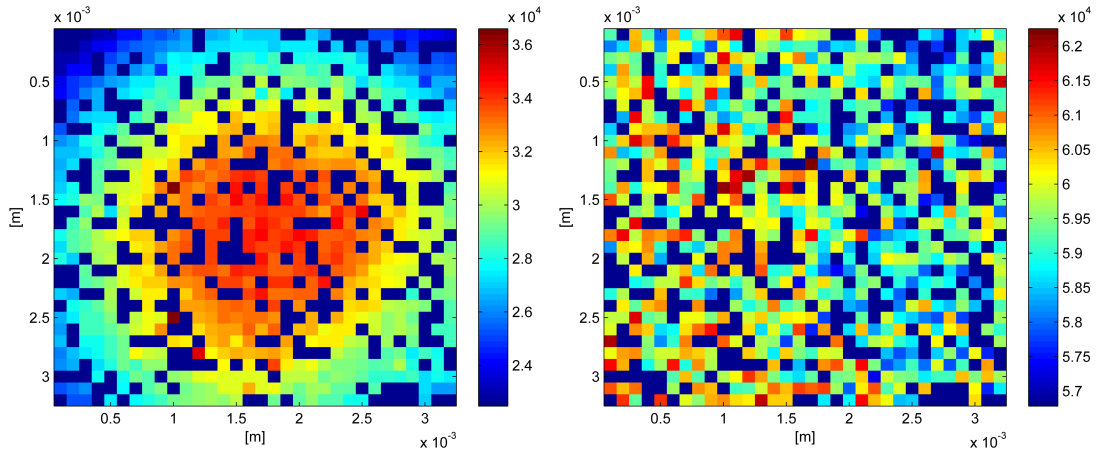
(b) Distribution after discarding pixels

Figure 6.3: Histograms relative to a background acquisition.

Through the usage of a software, expressly made for the purpose, the selected pixels were removed from the matrices. Fig. (6.3b) shows the result of the software implementation. In Fig. (6.4) are reproduced the same acquisitions of Fig. (6.2) after this analysis.

Pixel		Acquisition									
Row	Column	1	2	3	4	5	6	7	8	9	10
1	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	1		✓								✓
8	21	✓			✓	✓	✓		✓		✓
9	3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
14	25	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
26	9	✓			✓	✓	✓			✓	
27	8	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 6.1: Table of selected pixels. A check mark indicates if the pixel was found on that particular acquisition.



(a) LASER intensity profile after discarding pixels

(b) LED intensity profile after discarding pixels

Figure 6.4: Light sources profiles after discarding pixels. The two different profiles are now sharper than those of Fig. (6.2).

Taking in exam the LASER source, four variables are needed to fully parameterize it: the waist W_0, the operative wavelength ($\lambda = 808\,nm$), the distance of sampling relative to the optical fiber output z and the working power P.[2] All these parameters define the intensity profile according to the following formula:

$$I(x, y, z) = \frac{2P}{\pi W^2(z)} \exp\left[-\frac{2(x - x_0)^2 + 2(y - y_0)^2}{W^2(z)}\right] \tag{6.1}$$

$$W(z) = W_0\sqrt{1 + \left(\frac{z}{z_0}\right)^2}$$

$$z_0 = \frac{\pi W_0^2}{\lambda}$$

To validate the assumption of "Gaussian profile" it has been extracted a section of the image in correspondence to the intensity peak. Performing a fit of the extracted data using Eq. (6.1) demonstrated the reason behind the previous assumption. Fig. (6.5) shows the section relative to an acquisition of working power $P = 1.154\,\mu W$ and sampling distance of $z = 2.5\,cm$.[3]

[2]To determine the value of the working power P it has been used a Power Meter placed close to the camera.

[3]The fit of the data has been made normalizing the experimental and theoretical distributions.

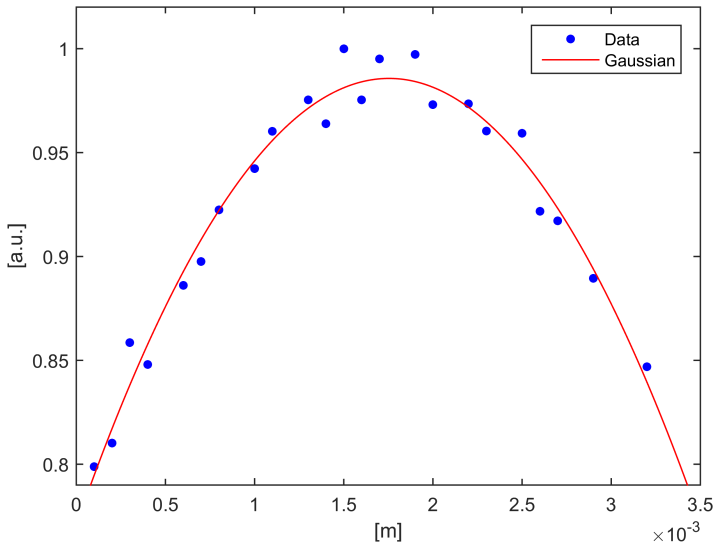


Figure 6.5: Fit of the LASER intensity profile section in correspondence of the intensity peak. The data and fitting equation were normalized.

From the good agreement between the data and Eq. (6.1) a 3-D fit was made (see Fig. (6.6)) extracting the three parameters: the centroids (x_0, y_0) and waist W_0. Tab. (6.2) shows the results.

x_0	y_0	Waist
$1.74 \pm 0.02 \times 10^{-3}\,\mathrm{m}$	$1.86 \pm 0.02 \times 10^{-3}\,\mathrm{m}$	$2.52 \pm 0.08 \times 10^{-6}\,\mathrm{m}$

Table 6.2: Parameters extracted from the 3-D fit of the LASER intensity profile.

The waist parameter reported in Tab. (6.2) refers to a single acquisition fit. In fact, most of the acquisition fits, reported different values of W_0, in most cases even not compatible with each others. This fact could be attributed to a misalignment of the light source or the camera especially for lower working powers ($\sim nW$) . It was then decided to discard some of the values to reduce the fluctuations to a maximum 10 % and to extract a reasonable waist value. This way the mean value is:

$$W_0 = 2.83 \pm 0.08 \times 10^{-6}\,\mathrm{m} \tag{6.2}$$

Having all the information about the LASER source, it is now possible to focus on the quantum efficiency of the camera. Exploiting the working mechanics of the SPC^2 it may be desirable to shift the analysis to the photodetection statistic theory. Using the information expressed in Sec. (3.2), it is possible to relate the statistical light source photons emission with the statistical photodetection of a device.

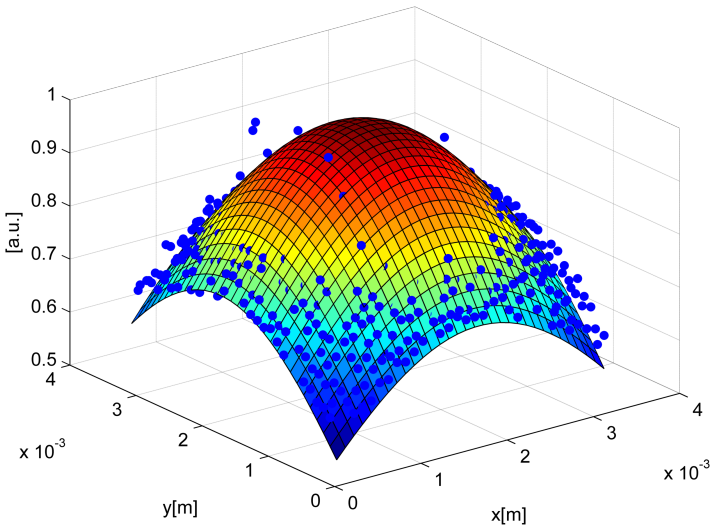


Figure 6.6: 3-D Gaussian fit of the LASER intensity profile.

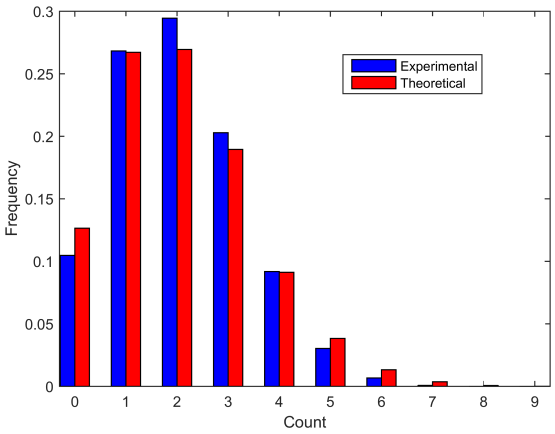
Let $(\Delta N)^2$ be the photodetection variance and $(\Delta n)^2$ the one relative to the light source photons emission. The two variances are then bound to satisfy the following relation:[4]

$$(\Delta N)^2 = \eta^2 (\Delta n)^2 + \eta (1 - \eta) \overline{n} \tag{6.3}$$

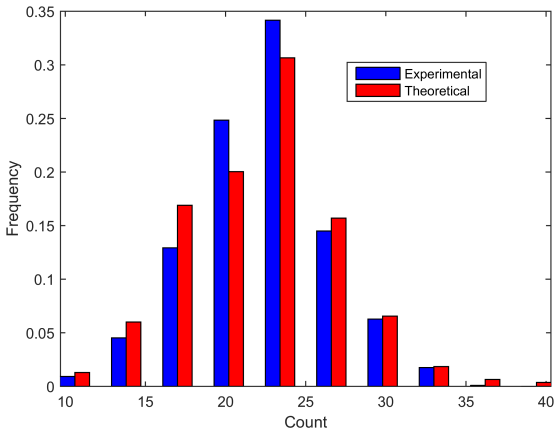
where $\overline{n}$ is the incident photons number mean, η is the detector's quantum efficiency defined as $\eta = \frac{\overline{N}}{\overline{n}}$ and $\overline{N}$ is the photodetection number mean.

A remarkable consequence follows from Eq. (6.3): if the incident light is characterized by a Poissonian distribution with $(\Delta n)^2 = \overline{n}$, then $(\Delta N)^2 = \eta \overline{n} = \overline{N}$ for every value of η. In other words, the photodetection statistics always coincides with a Poissonian distribution. Assuming that the LASER beam may be regarded as a light source of Poissonian photons distribution is widely supported by the coherent states study of a quantum system.[5] To better test this hypothesis it was decided to run a statistic analysis on all the acquired frames. The frames were summed up to different sequence lengths making an histogram of the results. Fig. (6.7a) shows the relation between the experimental distribution for a 11 frames sum (blue colored) and the theoretical Poissonian distribution with mean $\lambda = 2.03$ (red colored). Fig. (6.7b) instead shows an experimental distribution for a 121 frames sum and the theoretical one with mean $\lambda = 22.3$. Looking at Fig. (6.7), it is clear that the assumption of Poissonian distribution is well justified.

[4] To a more detailed derivation see Sec. (3.2).
[5] A more detailed relation is reported in App. (A).



(a) Experimental distribution of a 11 frames sum and theoretical distribution with mean $\lambda = 2.03$



(b) Experimental distribution of a 121 frames sum and theoretical distribution with mean $\lambda = 22.3$

Figure 6.7: Relationship between the experimental and theoretical distributions of photodetection. The blue distributions refer to the experimental ones while the red distributions to the theoretical.

According to the results obtained so far and the camera settings (one photodetection per frame and pixel) the photodetection probability P_{riv} is:

$$P_{riv} = 1 - P(0) \tag{6.4}$$

where $P(0)$ corresponds to the probability of no photodetection. Writing explicitly the Poisson distribution

$$P(n) = e^{\lambda} \frac{\lambda^n}{n!} \tag{6.5}$$

as a function of the photons number mean per unit frame and pixel $\overline{N}$ and of the quantum efficiency η, Eq. (6.4) becomes:

$$\eta = -\frac{1}{\overline{n}} \log\left(1 - \overline{N}\right) \tag{6.6}$$

where $\overline{n}$ is the incident photons number per unit frame and pixel. $\overline{n}$ was then retrieved using the following standard relations between energy, optical intensity and power:

$$P \cdot A_r = \frac{E}{T}$$
$$E = \frac{\overline{n} h c}{\lambda}$$

where T is the integration time of a single frame ($200\,ns$), h and c are the Planck constant and the speed of light respectively, λ is the wavelength of the optical radiation,

P is the working power of the LASER source rescaled by the A_r factor as the ratio of a single SPAD area and the total optical area of the camera.[6] This way the intensity profile of Eq. (6.1) takes the following form:

$$\overline{n}(x, y, z) = \frac{2P\lambda T A_r}{hc\pi W^2(z)} \exp\left[-\frac{2(x - x_0)^2 + 2(y - y_0)^2}{W^2(z)}\right] \tag{6.7}$$

Integrating Eq. (6.7) relatively to each pixel area and inserting the result in Eq. (6.6), a quantum efficiency "per pixel" was retrieved. Fig. (6.8) shows the values for a single acquisition. The overall quantum efficiency is the same for the most part of pixels confirming the validity of the previous theories and assumptions. Exploiting an average on all the pixels, a quantum efficiency relative to the LASER wavelength ($\lambda = 808\,nm$) is retrieved:

$$\eta = 5.47 \pm 0.38\,\%$$

The result shows a good compatibility with the theoretical value provided by the camera producer (see Fig. (5.7)).

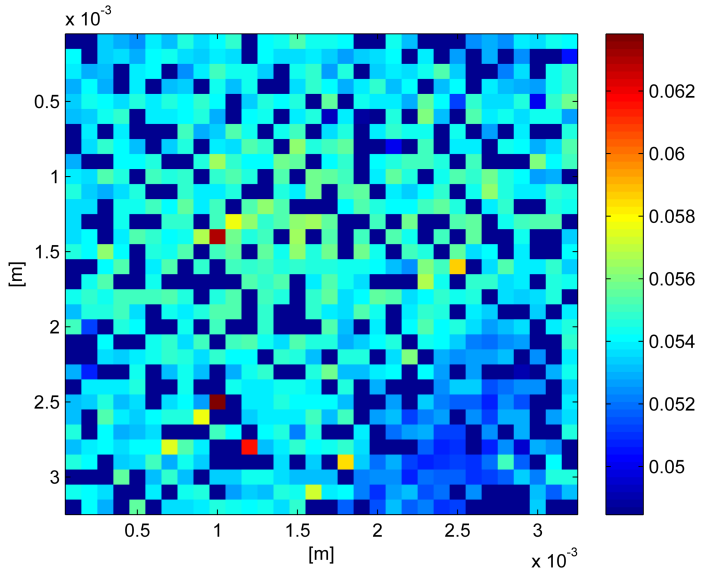


Figure 6.8: Quantum efficiency per pixel relative to a single acquisition for the LASER source.

[6] The rescaling factor is introduced because not the entirety of the camera allows a photodetection. To a more detailed explanations see Sec. (5.2).

Taking now in exam the LED source, the same analysis procedures were made on the system revising the new nature of the optical radiation. The LED radiation does not possess a well-defined intensity profile but shows a chaotic nature of emission instead while maintaining a high monochromaticity. These characteristics make the LED an excellent uniformly emitting light source. Accounting for these considerations, Eq. (6.7) becomes:

$$\overline{n} = \frac{PA_rT\lambda}{hc} \tag{6.8}$$

maintaining the same variables relations with the exception of the wavelength, now $\lambda = 470\,nm$, and of A_r that now corresponds to the ratio of a single SPAD area and the active area of the Power Meter. The reason behind the last change is because of the no more spatial confinement of the light radiation. Placing the Power Meter close to the camera allows to register the exact portion of light that falls in the camera objective accounting for the spatial diffusion of the radiation. As it was done for the LASER acquisitions, it was retrieved a section of a "LED image" observing the intensity profile of the radiation measured. Fig. (6.9) shows the section of a single acquisition relative to a working power of $P = 2.816\,\mu W$ and at a sampling distance of $z = 15\,cm$.[7] The intensity profile results constant validating the previous hypotheses.

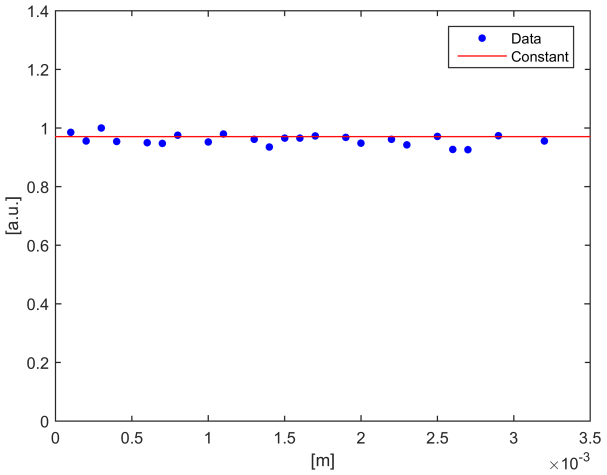


Figure 6.9: Fit of the LED intensity profile section. The data and fitting equation were normalized.

[7] As it was done for the LASER source, the fit was made after normalizing the data.

Contrarily to the LASER source, the photoemission statistic is no more described by a Poissonian distribution because of the different light emission. Despite this fact, the system still follows Eq. (6.3). That is because, in a semiclassical approximation, the photodetection statistic is expressed by a Poissonian distribution. In fact, the photodetection process meets the following conditions:

- The probability of the emission of a photoelectron in a short time interval Δt is proportional to the intensity I, the area A illuminated and the time interval Δt
- If Δt is sufficiently small, the probability of emitting two photoelectrons is negligibly small
- Photoemission events registered in different time intervals are statistically independent of each other

Applying the same analysis, it was retrieved a quantum efficiency "per pixel" shown in Fig. (6.10).

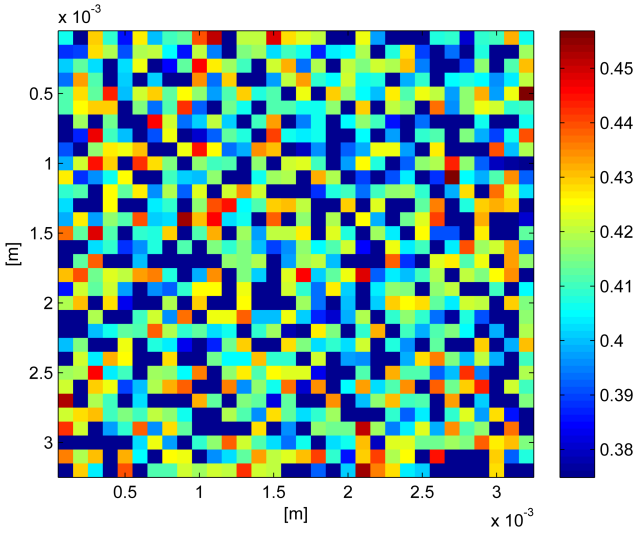


Figure 6.10: Quantum efficiency per pixel relative to a single acquisition for the LED source.

The average on all the pixels gives the following quantum efficiency for a wavelength of ($\lambda = 470\,nm$):

$$\eta = 41.4 \pm 0.21\,\%$$

Even in this case there is good accordance with the theoretical value provided by the camera producer (see Fig. (5.7)).

7 Experimental system's noises

This chapter shows the analysis made to account for the *optical crosstalk*, *dark count* and *afterpulse* effects present in every realistic devices.

7.1 Crosstalk

The crosstalk model, presented in Sec. (4.1), allows to compute the successful event probability relying on the intrinsic binomial distribution for a pixel activation and on the "*histories*" of cascading events generated by a primary pixel. The experimental setup and the camera's acquisition settings allow to directly analyze the "background" obtained for each power-distance configuration (see Sec. (4.1)). Expanding Eq. (4.3) till $j \leq 5$, for a generic number n of neighbors, it is possible to establish a maximum limit of cascading events to determine the p parameter. Following are shown the calculated expansions:

$$
\begin{aligned}
P(1) &= (1-p)^n = 1-\varepsilon \\
P(2) &= np(1-p)^{2n-1} \\
P(3) &= \frac{1}{2}n(3n-1)p^2(1-p)^{3n-2} \\
P(4) &= \frac{1}{3}n(8n^2-6n+1)p^3(1-p)^{4n-3} \\
P(5) &= \frac{1}{4}n\left(\frac{125}{6}n^2-25n^2+\frac{55}{6}n-1\right)p^4(1-p)^{5n-4}
\end{aligned}
$$

To analyze the 32x32 array it is necessary to apply a recognition algorithm that prevents to double count a specific history for two different triggered pixel. For example, assuming a as primary pixel, the secondary b pixel will be activated if there is a crosstalk event, but it is not true for the reverse process. In fact, the b pixel is not the product of the a pixel ($a \rightarrow b \neq a \leftarrow b$). To investigate the system's geometry using an "8 neighbors scheme" (see Fig. (7.1)), it has been crated a mask that could account for all the system's criterion. Fig. (7.2) shows the mask applied to the array of pixels color-filled to underline the difference between the algorithm used. The *blue* pixels use the algorithm shown in Fig. (7.3a), the *azure* that of Fig. (7.3b), the *green* that of Fig. (7.3c), the *orange* that of Fig. (7.3d) while the *red* pixel does not use any algorithm.

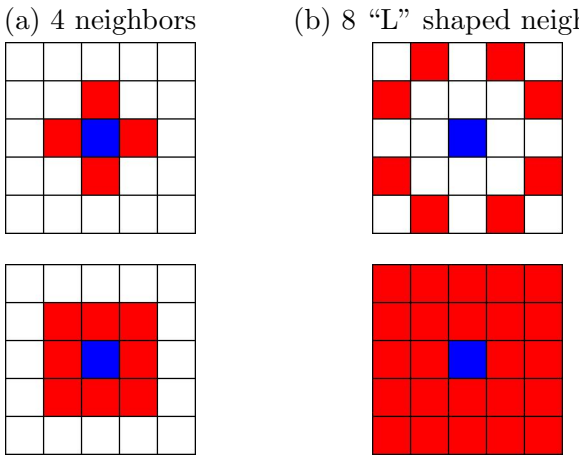


Figure 7.1: Possible configuration for the geometric analysis of the crosstalk effect. The blue-filled pixel is the primary pixel, the red-filled pixels are the selected pixels for the crosstalk analysis. The adopted configuration for the system is that of 8 neighbors (c).

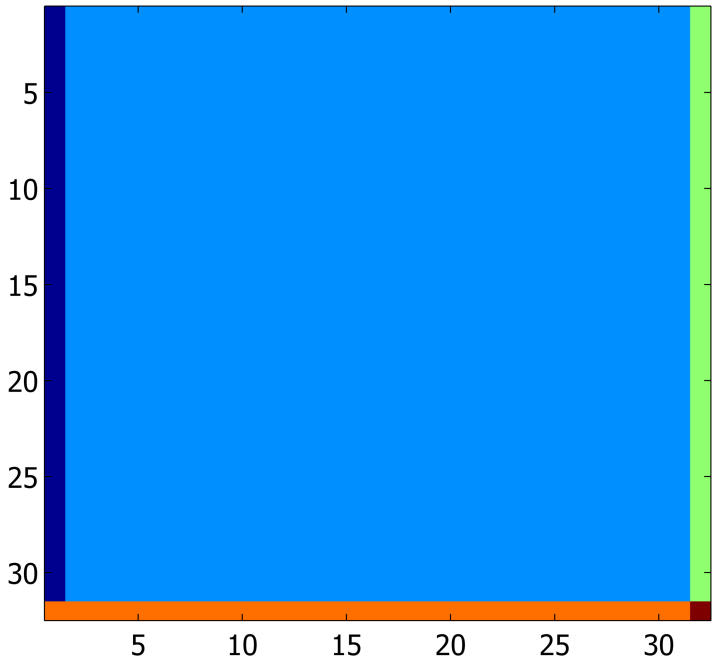


Figure 7.2: Crosstalk analysis mask. For every color-filled pixel, corresponds a different algorithm of crosstalk analysis.

After applying the different algorithms for each pixel to all the 131043 acquisition frames, the successful crosstalk events were recorded and highlighted. Fig. (7.4) reports

the results of such analysis. Different zones of the image possess greater contrast of others suggesting that the crosstalk effect is more frequent. Such system, however, does not provide which of the most intense pixel is the cascading generator. In fact, it is not possible to distinguish between the two "time reversal histories" $[(i \to j)\,\text{or}\,(i \leftarrow j)]$.

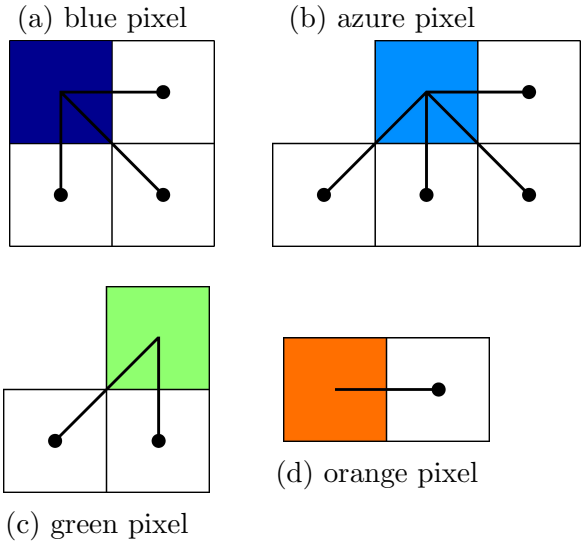


Figure 7.3: Rappresentation of the different algorithms used to the crosstalk analysis. Each different color-filled pixel in Fig. (7.2) corresponds to one of the algorithms.

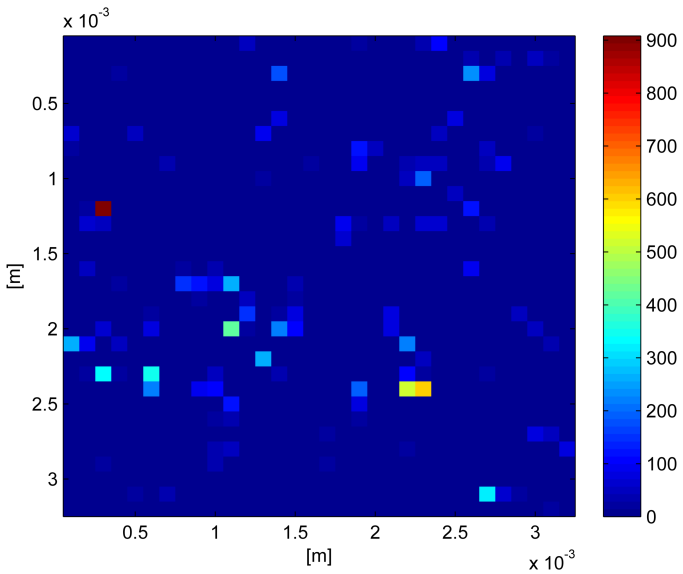


Figure 7.4: Image of successful crosstalk events relative to the lighting of a secondary pixel. Around the most intense pixels there are brighter zones indicating that these pixels too have triggered other successful crosstalk events.

To achieve a more precise calibration the system has been adjusted to account for events that triggered at least three cascading pixel. This way, the primary pixel will be highlighted among the rest giving a privileged direction to the cascading process. Fig. (7.5) shows the results for this setting. Iterating the process for "more cascades", the probability decreases rapidly in magnitude until no crosstalk events are detected. Fig. (7.6) shows the results for a three cascading events analysis. At the next level (four cascading events) no more pixels are selected. The quickness with which the probability ε decreases of magnitude for increasing cascading triggered pixels is shown in Fig. (7.7). Already after just one secondary triggered pixel the probability decreases by almost three orders of magnitude. The maximum value of the crosstalk probability has been used to not underrate the later entropy and random extraction analysis: $\varepsilon = 9.67 \times 10^{-2}$ ‰.

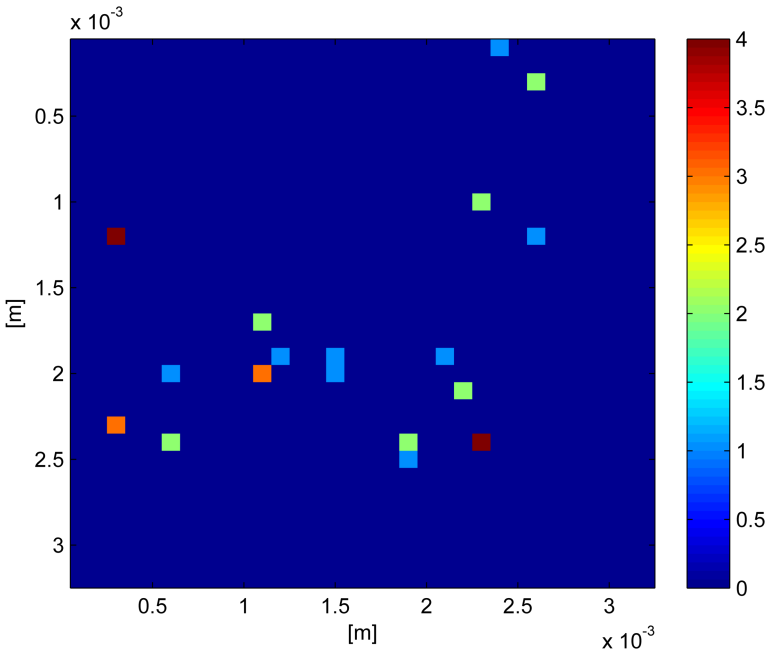


Figure 7.5: Image of successful crosstalk events relative to the lighting of two secondary pixels. The brighter zones of Fig. (7.4) are almost vanished. The remaining pixels are the effective primary pixels responsible for the cascading events.

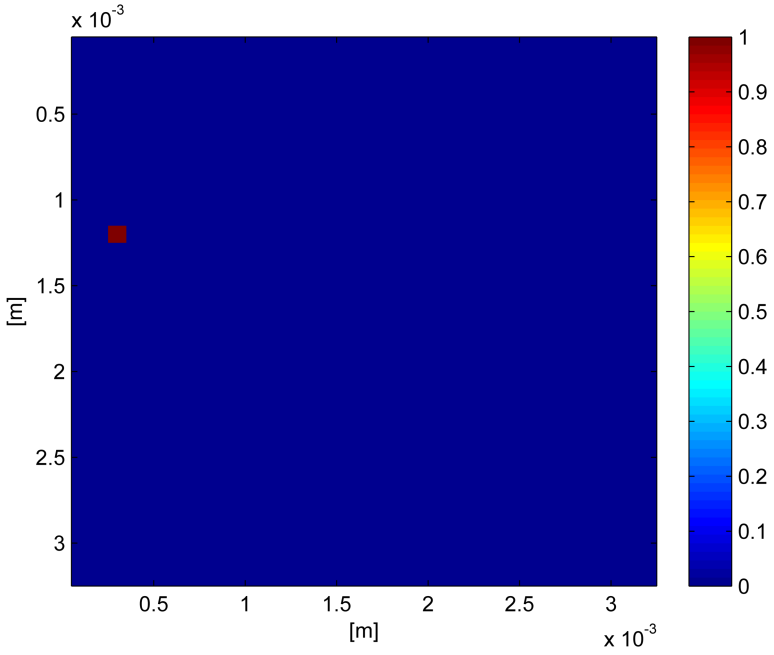


Figure 7.6: Image of successful crosstalk events relative to the lighting of three secondary pixels. Only one pixel meets the system's conditions.

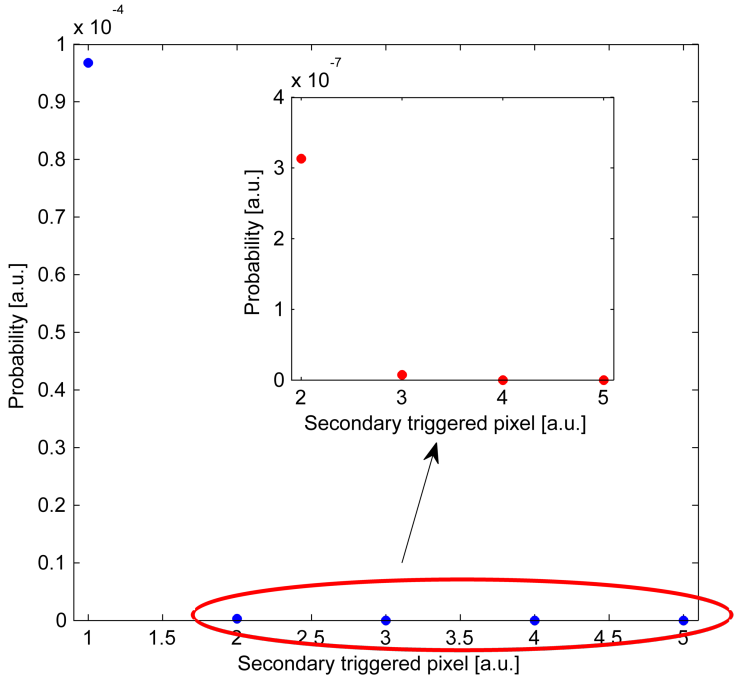


Figure 7.7: Relation between the crosstalk probability and the secondary triggered pixels. The probability decreases by almost three orders of magnitude going from one to two secondary triggered pixels.

7.2 Dark count

As depicted in Sec. (4.2), the dark count effect is relative to the generation of thermal carriers (electrons-holes). These particles may trigger an avalanche emission creating a false signal causing an increase of the electronic noise. Using the "background" frames it is possible to directly analyze the number of events produced by a false signal, i.e. dark count effect, thus determining the ratio $\frac{N_C}{N_G}$ (see Eq. (4.5)). Knowing that the integration time of a single frame is $\tau = 200\,ns$ and the total number of frames is $N_G = 131043$, the mean dark count rate is: $\mu = 27428$ cps. Fig. (7.8) shows the μ values for each pixel.

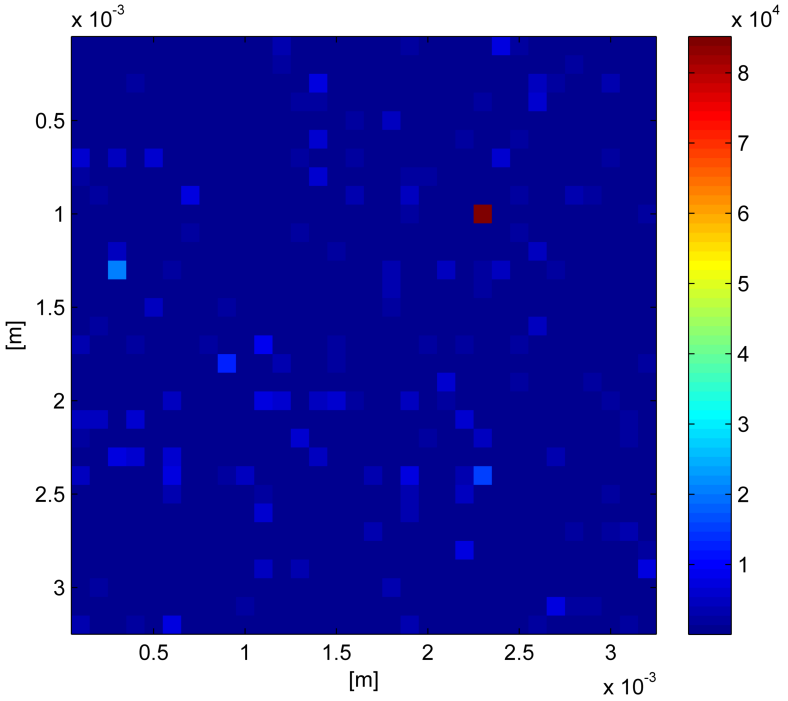


Figure 7.8: Dark count rate per pixel.

7.3 Afterpulsing

The afterpulsing effect comes from the inability of the device to quench an avalanche event before the next sampling. This results in the fact that different events are now correlated affecting the entire analysis. A similar process to that of the dark count rate permits to quantify the phenomenon. This time a temporal correlation substitutes the spatial correlation. The system is set so that, for each pixel, a double or triple triggered pixels in the next frames are selected and accounted for the afterpulse effect. The probability thus becomes:

$$\gamma = -\log\left(1 - \frac{N_A}{N_G}\right) \tag{7.1}$$

where N_A is the number of successful afterpulsing events. Fig. (7.9) shows the values γ for each pixel.

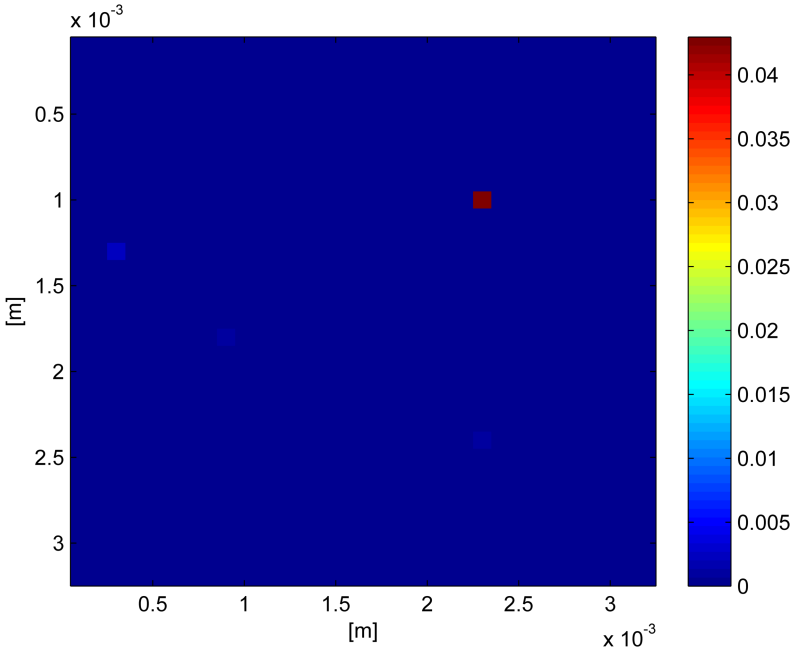


Figure 7.9: Afterpulsing probability per pixel.

As for the dark count rate, it has been chosen the maximum value of γ, that is: $\gamma = 4.29\,\%$.

8 Entropy

A parameter to test the quality of a QRNG is the *information entropy*. Its usage is to "quantify" a variable's randomness prior its measurement. The first section presents the entropy analysis made on the system focusing on the theoretical model expressed in Chap. (1). The second section instead presents the model for a generic number of detectors and its analysis.

8.1 Entropy analysis

It is possible to "quantify" the amount of information contained in a message through the usage of the *entropy* concept. The meaning of entropy may also be expressed as the quantity of uncertainty about a system *before* performing a measurement. In the Information Theory exists different type of entropy, but three are the most used in every-day experiments: the *Shannon Entropy*, the *Conditioned min-Entropy* and the *Classical min-Entropy*.[1]

The Shannon Entropy gives an upper limit to the length of a lossless compressed bit string defined as:

$$H_S(X) = -\sum_x P_X(x) \log_2 P_X(x) \tag{8.1}$$

where in the presence of side information, i.e. the variables that characterize the photons detector or the light source, Eq. (8.1) becomes:

$$H_S(X|W) = \sum_{w \in W} P_W(w) H(X|W = w) \tag{8.2}$$

where W represents a variable that accounts for all side information of the system.

The Conditioned min-Entropy instead provides a lower limit according to the following formula:

$$H_{min}(X|w) = -\log_2 \left[\max_x P_{X|W}(x|w) \right] \tag{8.3}$$

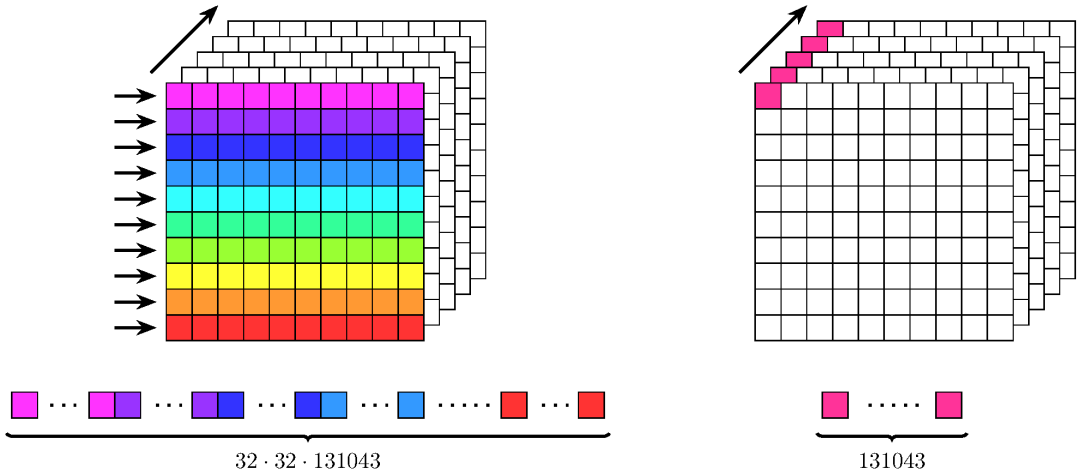
where W is the same variable of Eq. (8.2).

[1] Commonly the entropy concept is referred not only to string of bits but any type of system and information such as the letters in a message or the colors used in a painting.

The Classical min-Entropy[2] takes its name from the fact that no side information are taken into consideration and the system is considered completely deterministic. The system just computes the probability of all possible outputs, i.e. 0 or 1 for a bit string, and then defines its entropy as:

$$H_\infty(X) = -\log_2 \left[\max_x P_X(x)\right] \tag{8.4}$$

Making use of these functions and applying the "*Leftover Hash Lemma with Side Information*" (see Sec. (1.2)) permits to retrieve bit strings that are, with excellent approximation, truly random. Side information alters the system and the randomness extracted, allowing to perform two different "methods": a quantum and classical approach. From the point of view of the classical approach, the system may be compared to the simple "*flipping coin*" problem. The device, in fact, gives a sequence of 1s or 0s in correspondence of a detection or not. That is the same as looking at the result of a flipped coin: "*head*" or "*tail*". This configuration than quantify the uncertainty about the system itself. It is important, however, to stress out that Entropy *is not* information. Entropy just gives a knowledge of the system and helps to quantify the "*uncertainty about the source of information*". In other words, the less likely an event is, i.e. head or tail, 0 or 1, the more information it provides when it occurs. The method used to perform all acquisitions (see Sec. (5.3)) permits to split the analysis in two different ways: spatial and temporal.



(a) Spatial analysis

(b) Temporal analysis

Figure 8.1: Representation of the two different pixels extraction methods. The *spatial extraction* allows to extract each row of the 32x32 array of pixels from all 131043 frames. The *temporal extraction* allows to extract each pixel from the 131043 frames of each acquisition.

[2]The term "Classical" does not refer to a classical system but merely indicates that the entropy does not account for any side information that might exist.

Spatial analysis:

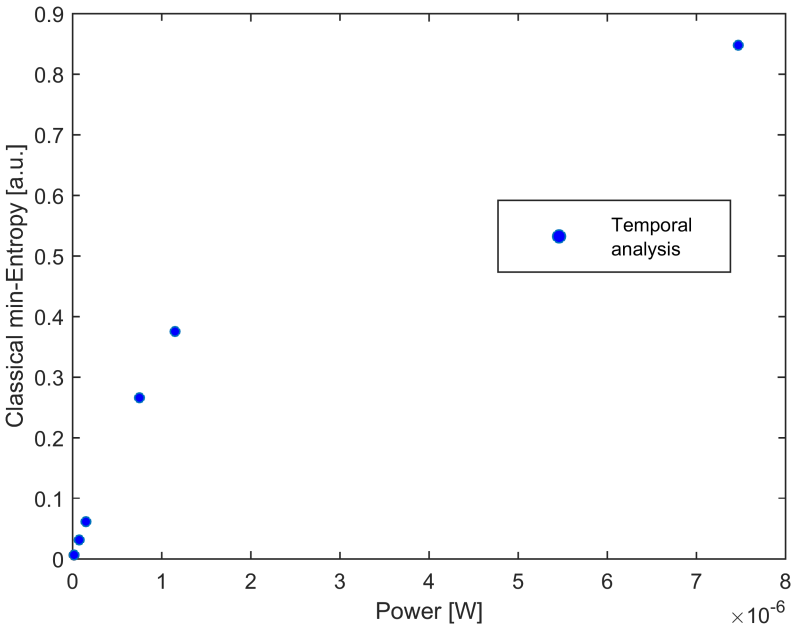
The spatial analysis consists in extracting each row of the 131043 arrays of pixels creating a single string of 32x32x131043 pixels. This way the system permits to relate the spatial photons distribution with the detection distribution of the entire active area. Fig. (8.1a) shows an example of spatial extraction used in a single acquisition. This kind of study is suitable only for the LED source because of its incoherent nature. The LED light, in fact, does not possess a defined spatial intensity profile and each pixel is independent of one another. The LASER distribution instead has a well defined spatial distribution as Eq. (6.7) indicates. If the LASER acquisitions were to be analyzed using this method, the request of independence would not be satisfied.

Temporal analysis:

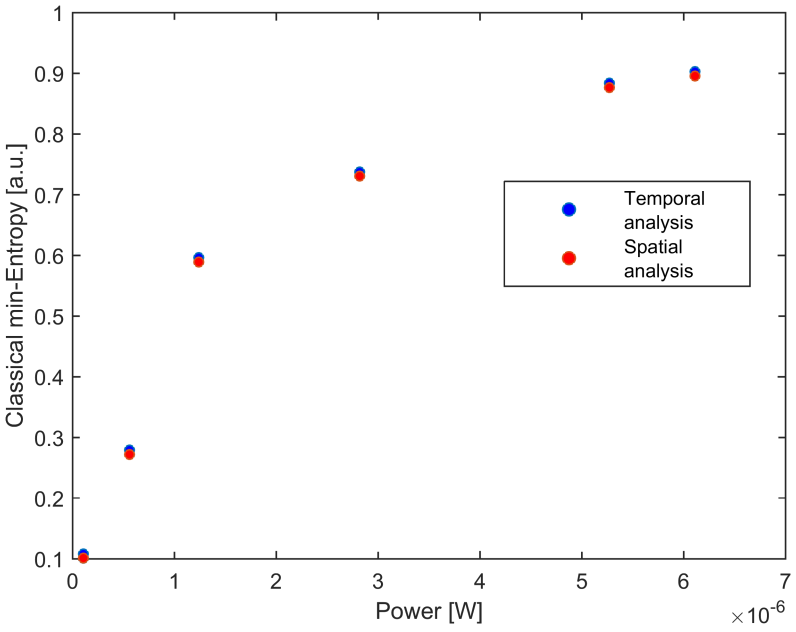
The temporal analysis consists on extracting each pixel separately in all the 131043 frames available for a single acquisition. This method relies on the fact that there is no temporal correlation between subsequent detections for a single pixel.[3] Fig. (8.1b) shows such extraction method.

Using both methods, it has been computed the Classical min-Entropy for varying light source and working powers with the use of Eq. (8.4). Fig. (8.2) shows the results of such calculus for both sources (LASER and LED). Observing Fig. (8.2a) and (8.2b) the entropy goes to zero for lower value of power ($\sim nW$) and then increase as the power increase. This accurately reproduces the theoretical model: for lower power values there is a higher probability of finding pixels with no detection, i.e. 0. In the limit case when the probability is exactly 1, $[P(X) = 1]$, the entropy is $H_\infty(X) = -\log_2(1) = 0$, as to say that the system does not provide any *more* information, i.e. it is no longer necessary to know if the system detected a photon or not because it is already known. When the power increases, the system mirrors the previous setting because now the probability of finding pixels with one detection is predominant. Entropy is maximum when there is equal probability of finding pixels with one detection or not: $P(X) = 0.5 \Rightarrow H_\infty(X) = -\log_2(0.5) = 1$. Fig. (8.3) shows the relation between Classical min-Entropy and the probability of finding pixels with no detection.

[3]This is not the case when afterpulsing effects are predominant.



(a) LASER source



(b) LED source

Figure 8.2: Classical min-Entropy for LASER and LED sources. The red data denote the *spatial extraction* while the blue the *temporal extraction*.

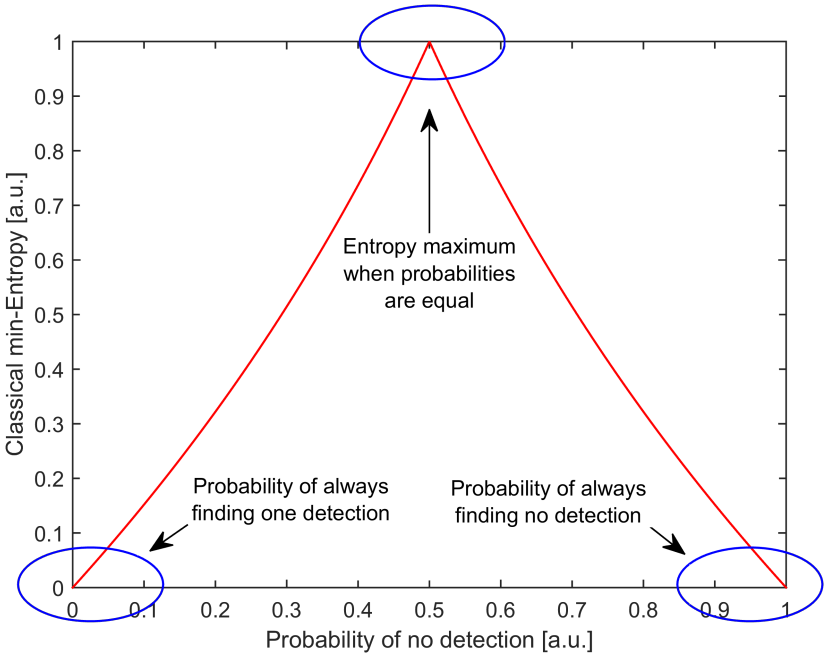


Figure 8.3: Relation between Classical min-Entropy and the probability of finding pixels with no detection. The entropy is minimum when it is absolutely known the output of a measurement, i.e. 1 or 0, $[P(0) = P(1) = 1]$. It is maximum when the probabilities are equal $[P(0) = P(1) = 0.5]$.

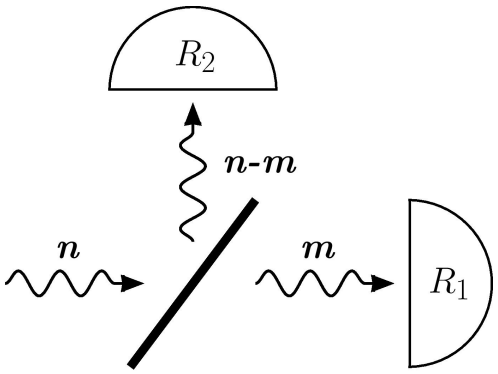


Figure 8.4: A PBS model of a QRNG. n incoming photons hit the PBS and are transmitted (m) or reflected (n-m) according to their polarization. These photons are then registered by two photodetectors (R_1, R_2).

The approaches so far do not account for the quantum nature of the system and do not consider the possible outcomes of the system measurement. Adjusting then the entropy analysis to the quantum system of the model described in Chap. (1), it is necessary to determine the probability of all possible output using the PBS (**P**olarising **B**eam **S**plitter) analogy shown in Sec. (1.3).

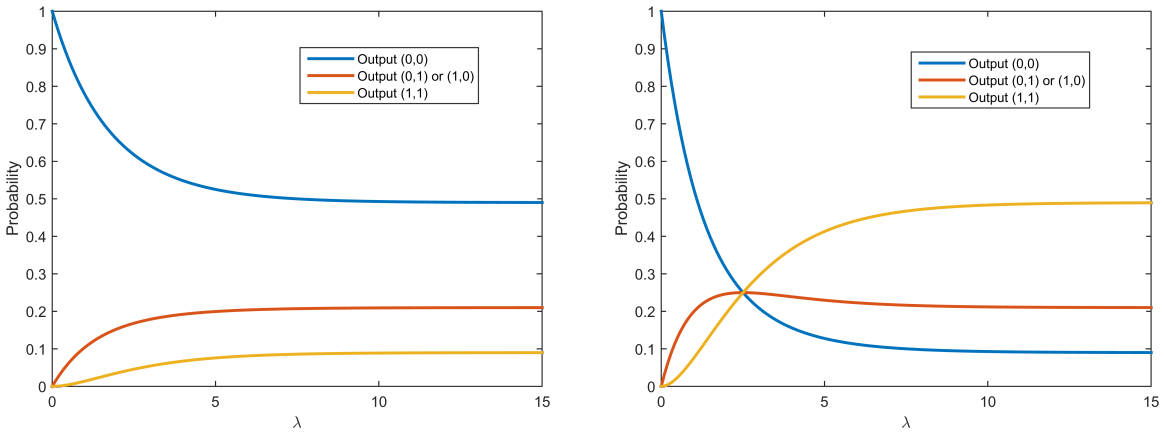
Incident photons are reflected or transmitted according to their polarization and then "seen" by two photodetectors (see Fig. (8.4)). Considering only two of the 1024 pixels of each frame permits to treat the system exactly as the PBS model. One of the assumptions made is that the incident photons, $n \in [0 \dots \infty]$, reflect a Poissonian distribution (see Eq. (6.5)). Enumerating the possible outcome of two photodetectors gives the following set of pairs: $\{(0,0),(0,1),(1,0),(1,1)\}$. Because the detectors are not ideal, a photon detection occurs with probability η. Assuming, then, that n incident photons hit the PBS, the probabilities of all possible outcome become:

(x_1, x_2)	$P_X(x_1, x_2)$
$(1,1)$	$\sum_{n=2}^{\infty} P_N(n)\left(1-2\left(\frac{1}{2}\right)^n\right)\eta^2$
$(1,0)$	$\frac{1}{2}P_N(1)\eta + \sum_{n=2}^{\infty} P_N(n)\left[\left(\frac{1}{2}\right)^n\eta^2 + \left(1-\left(\frac{1}{2}\right)^n\right)\eta(1-\eta)\right]$
$(0,1)$	$\frac{1}{2}P_N(1)\eta + \sum_{n=2}^{\infty} P_N(n)\left[\left(\frac{1}{2}\right)^n\eta^2 + \left(1-\left(\frac{1}{2}\right)^n\right)\eta(1-\eta)\right]$
$(0,0)$	$P_N(0) + P_N(1)(1-\eta) + \sum_{n=2}^{\infty} P_N(n)\left[2\left(\frac{1}{2}\right)^n(1-\eta) + \left(1-2\left(\frac{1}{2}\right)^n\right)(1-\eta)^2\right]$

Table 8.1: Probabilities relative to all possible outcome for a system of n incoming photons and two detectors. x_1, x_2 represent the output of detector R_1, R_2 respectively while P_X is the probability relative to that particular combination $\{(0,0),(0,1),(1,0),(1,1)\}$.

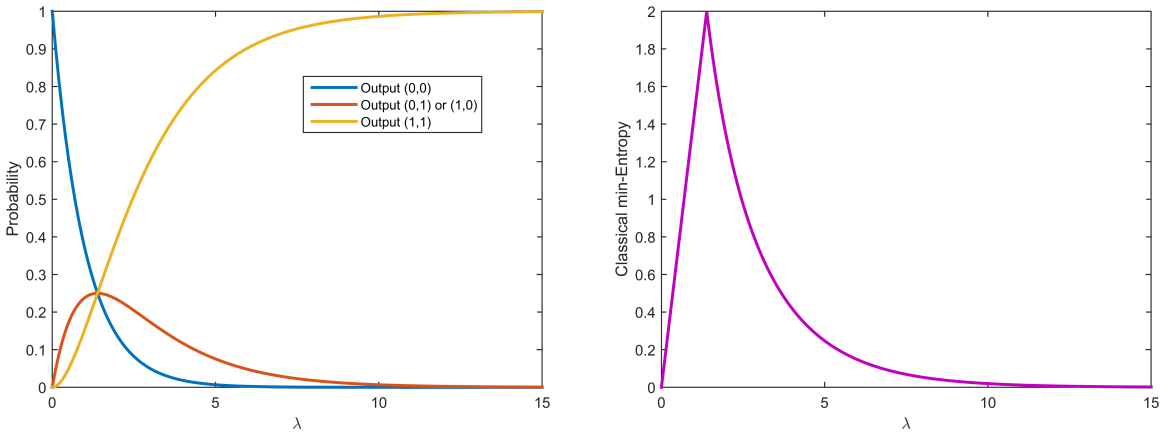
To see how these probabilities have been determined, it is simpler to focus on the expression for the $(1,1)$ output. This particular outcome may occur only when there are at least two incoming photons and that is why the summation starts from two and not 1. After the interaction with the PBS, photons must be present in both optical paths, and this happen with probability $1-2\left(\frac{1}{2}\right)^n$. This formulation arises from the fact that the probability, of finding all photons in the same optical path, occurs with $\left(\frac{1}{2}\right)^n$ probability. Assuming then that the detectors are not ideal, obtaining a double revelation occurs with η^2 probability. Adding all these together results exactly in the formula in Tab. (8.1). Now searching the maximum probability, for different values of λ of all the expressions in Tab. (8.1) gives the corresponding value of the Classical min-Entropy.[4] Fig. (8.5) shows the relation between probabilities of Tab. (8.1) and the parameter λ for different quantum efficiency ($\eta = 30\%$ and $\eta = 70\%$).

[4]Once again, the term "Classical" does not refer to a classical system but simply to a system that does not account for any side information.



(a) Probabilities for $\eta = 30\%$ (b) Probabilities for $\eta = 70\%$

Figure 8.5: Probabilities of Tab. (8.1) for different values of λ and quantum efficiency ($\eta = 30\%$ and $\eta = 70\%$).

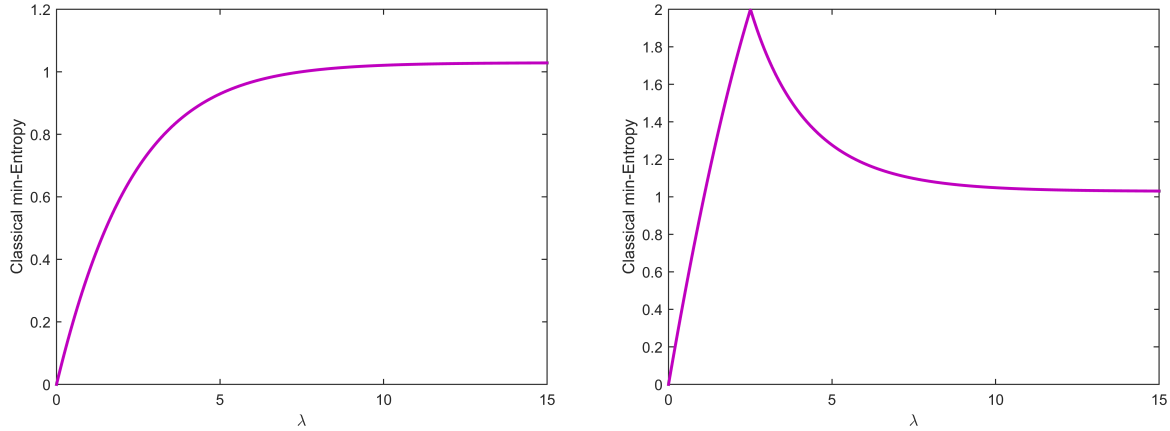


(a) Probabilities for $\eta = 100\%$ (b) Classical min-Entropy for $\eta = 100\%$

Figure 8.6: Probabilities of Tab. (8.1) and Classical min-Entropy for different values of λ for an ideal system ($\eta = 100\%$).

Observing Fig. (8.5) it is straightforward how the efficiency plays a fundamental role in the quantum system. This kind of behavior is easily explained. As the quantum efficiency increases to its maximum value ($\eta = 100\%$), the probability, of finding a signal in both detectors, increases as the mean of incident photons increases. Fig. (8.6) shows both the probabilities and the Classical min-Entropy for an ideal system ($\eta = 100\%$). The fact that the maximum of the Classical min-Entropy is 2 is consistent with the system's setting. In fact, when all possible outputs $\{(0,0),(0,1),(1,0),(1,1)\}$ are equally probable, the information gained from the system is equal to the number of bits available for one acquisition: 2. A further example on this matter is considering an 8 bit string. If all the extractable number with these bits, $\{0,\ldots,255\}$, are equally probable to come

out, the system's entropy is exactly 8. Applying Eq. (8.4) to the maximum of all the probabilities shown in Fig. (8.5), it is then extracted the corresponding Classical min-Entropy (see Fig. (8.7)).



(a) Classical min-Entropy computed for $\eta = 30\%$

(b) Classical min-Entropy computed for $\eta = 70\%$

Figure 8.7: Classical min-Entropy for different values of λ and quantum efficiency ($\eta = 30\%$ and $\eta = 70\%$).

This new approach well defines the quantum nature of the system, but the information extracted using the Classical min-Entropy does not ensure that an adversary may not predict the behavior of the system. In fact, he could be in posses of information about the system itself that allows him to change it as he sees fit. That is why any side information, accessible to anyone, have to be taken into account. The theoretical model provides these information in the contest of "Maximum Classical noise". The side information thus are: the number n of incoming photons and the two random variables, t_1 and t_2, that encode the information whether the two detectors are sensitive or not. Through the use of Eq. (8.3) and recalling that

$$2^{-H_{min}(X|NT_1T_2)} = \sum_{n,t_1,t_2} P_N(n) P_{T_1}(t_1) P_{T_2}(t_2) 2^{-H_{min}(X|nt_1t_2)}$$

where

$$H_{min}\left(X|nt_1t_2\right) = -\log_2 \left[\max_{x=(x_1,x_2)} P_{X|NT_1T_2}(x|nt_1t_2) \right]$$

the Conditioned min-Entropy becomes:

$$H_{min}\left(X|NT_1T_2\right) = -\log_2\left[P_N(0) + \sum_{n=1}^{\infty} P_N(n)\left(A+B+C\right)\right] \tag{8.5}$$

$$\text{where } \begin{cases} A = (1-\eta)^2 \\ B = 2\eta(1-\eta)\left(1-\left(\frac{1}{2}\right)^n\right) \\ C = \eta^2 \max\left[\left(\frac{1}{2}\right)^n, 1-2\left(\frac{1}{2}\right)^n\right] \end{cases}$$

For the upper bound of extraction instead, the Shannon Entropy becomes:

$$\begin{aligned} H_S(X|NT_1T_2) &= \sum_{n,t_1,t_2} P_N(n)P_{T_1}(t_1)P_{T_2}(t_2)H(X|nt_1t_2) \\ &= \sum_{n=1}^{\infty} P_N(n)\left\{2\eta(1-\eta)\left(-A\log_2 A - B\log_2 B\right)\right\} \\ &+ \sum_{n=1}^{\infty} P_N(n)\,\eta^2\left\{-2A\log_2 A - C\log_2 C\right\} \end{aligned} \tag{8.6}$$

$$\text{where } \begin{cases} A = \left(\frac{1}{2}\right)^n \\ B = \left(1-\left(\frac{1}{2}\right)^n\right) \\ C = \left(1-2\left(\frac{1}{2}\right)^n\right) \end{cases}$$

Fig. (8.8) shows both entropies, the Conditioned min-Entropy and the (conditioned) Shannon Entropy, in relation to λ (mean of the incident photons) for $\eta = 30\%$. Now, the value λ may be experimentally extracted using the notion expressed in Chap. (3) and (6). Because λ physically represents the mean of the incident photons on the system, this is exactly the $\overline{n}$ variable used to indicate the incoming photons from the LASER and LED sources. This way, the quantum efficiency becomes:

$$\eta = \frac{\overline{N}}{\overline{n}} \Rightarrow \eta = \frac{\overline{N}}{\lambda}$$

Picking two adjacent pixels from the 32x32 array permits to approximate the same mean of incident photons and then compute it thanks to the previous equation. Once obtained the λ value, it is inserted in Eqs. (8.5) and (8.6) to retrieve the entropy. Fig. (8.9) shows the relation between the experimental and theoretical values of both entropies for different values of λ of the LASER source. The corresponding λ values have been converted from the working power of the light source using the equations of Chap. (6). Tab.(8.2) shows the converted values.[5]

[5]The values refer to a particular pair of pixels chosen from all the 1024 of each acquisition.

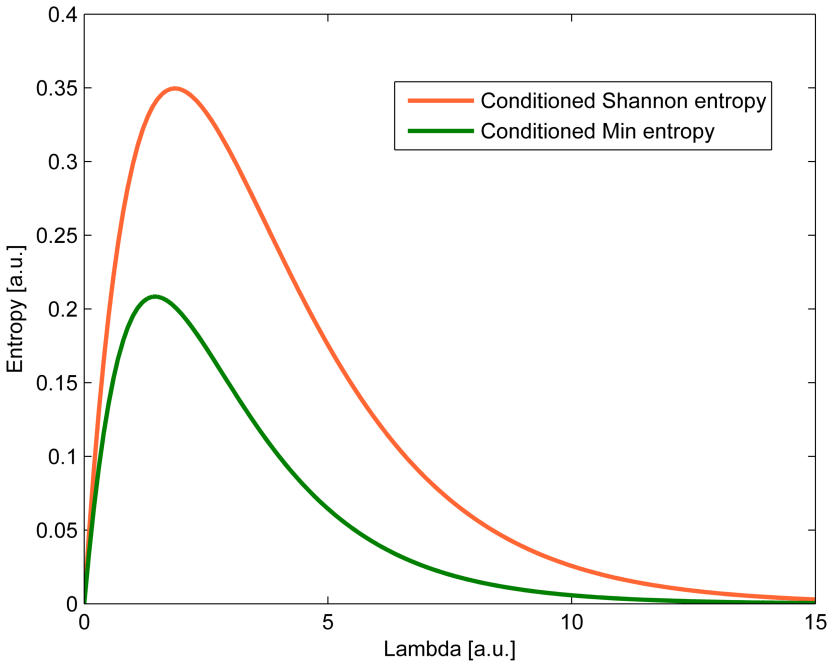


Figure 8.8: Relation of the Conditioned min-Entropy and the (conditioned) Shannon Entropy with the parameter λ for a quantum efficiency of $\eta = 30\%$.

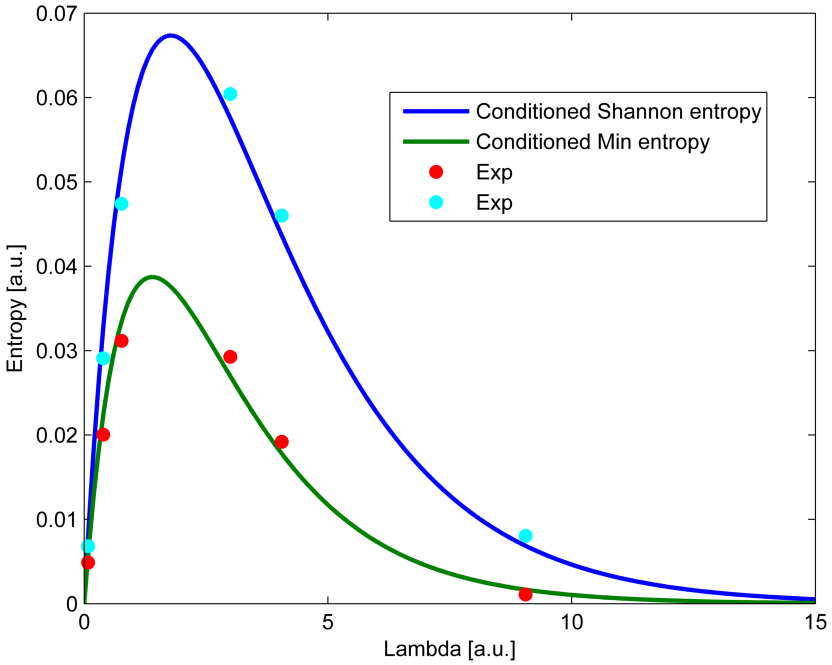


Figure 8.9: Relation between the experimental and theoretical values of both conditioned entropies for different values of λ for the LASER source. The colored lines refer to the theoretical values computed for a system of quantum efficiency $\eta = 5.47\%$.

There is good accordance between the theoretical and experimental values denoting the correct reasoning for the quantum system.

Power [W]	λ
14.83×10^{-9}	$0,093$
73.05×10^{-9}	$0,462$
148.8×10^{-9}	$0,909$
749.8×10^{-9}	$3,552$
1.154×10^{-6}	$4,801$
7.467×10^{-6}	$9,053$

Table 8.2: Table of converted power values of the LASER source used to determine the experimental data inserted in the graph of Fig. (8.9).

8.2 Generalized model

The theoretical model, so far, permits to quantify randomness in the presence of noise conditioning the entropy on such side information. This method allows then to extract variables that are uniformly distributed and noise-independent. The approach, however, relies only on the simple system of two detectors and a PBS that deflects incoming photons to the two devices. To develop the powerful model's capacity, it is possible to generalize the system to a generic l number of detectors and, of course, to n incoming photons. The first step is computing the probabilities of all possible outputs for a given number of detectors.

Suppose that there are l detectors, and that only k are 1 while the rest are 0. Suppose also that, within the classical model, a detector "clicks" with probability η corresponding to the quantum efficiency of the system. If n photons impinge on one of the l detectors, the probabilities of obtaining 0 or 1, $p_{0|n}$ and $p_{1|n}$ respectively, are:

$$\begin{aligned} p_{0|n} &= \delta_n + (1-\eta)(1-\delta_n) \\ p_{1|n} &= \eta(1-\delta_n) \end{aligned} \qquad \text{where} \begin{cases} \delta_n = 1 & \text{if } n = 0 \\ \delta_n = 0 & \text{if } n \neq 0 \end{cases}$$

If the incoming photons are distributed according to a Poissonian distribution, $\Upsilon_\mu(n)$, with mean μ, it is possible to define a specific "spatial position" for the jth detector with $j = 1, \ldots, l$. This way, each device is "hit" by a *Spatial Poissonian* $\Phi_{p_j\mu}(n)$ with mean $p_j\mu$, where p_j is the intrinsic probability of the jth detector. In this case, the probability that a detector produces a 0 is:

$$
\begin{aligned}
P_0^{(j)} &= \sum_{n=0}^{\infty} p_{0|n} \Phi_{p_j\mu}(n) \\
&= \Upsilon_{p_j\mu}(0) + (1-\eta) \sum_{n=1}^{\infty} \Upsilon_{p_j\mu}(n) \\
&= e^{-p_j\mu} + (1-\eta) \left[\underbrace{\sum_{n=0}^{\infty} \Upsilon_{p_j\mu}(n)}_{1} - e^{-p_j\mu} \right] \\
&= e^{-p_j\mu} + (1-\eta)(1-e^{-p_j\mu}) \\
&= 1 - \eta + \eta\, e^{-p_j\mu}
\end{aligned}
$$

Analogously the probability that a detector produces a 1 is:

$$
\begin{aligned}
P_1^{(j)} &= \sum_{n=0}^{\infty} p_{1|n} \Phi_{p_j\mu}(n) \\
&= \eta \sum_{n=1}^{\infty} \Upsilon_{p_j\mu}(n) \\
&= \eta \left[\underbrace{\sum_{n=0}^{\infty} \Upsilon_{p_j\mu}(n)}_{1} - e^{-p_j\mu} \right] \\
&= \eta\, (1 - e^{-p_j\mu})
\end{aligned}
$$

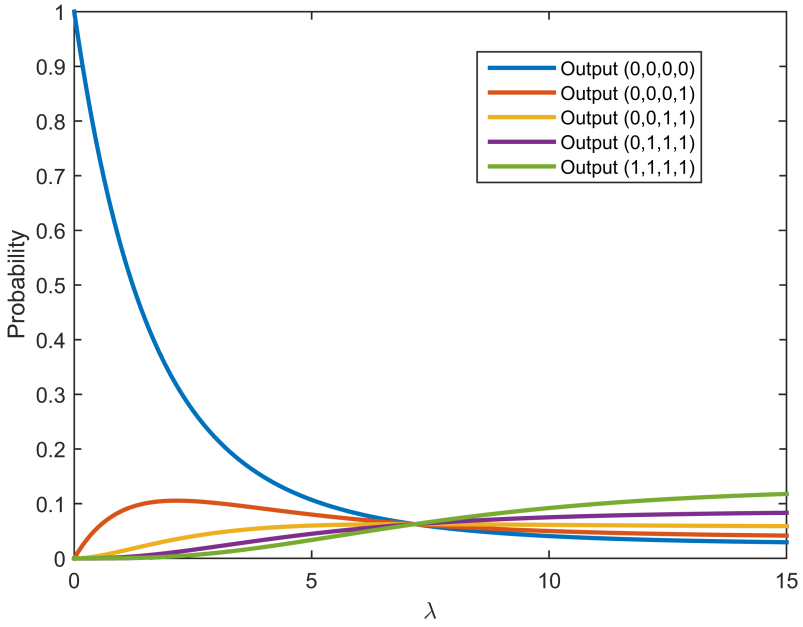
Denoting by $\mathcal{Q}_k$ the set of detectors that produced a 1 while $\mathcal{Z}_{l-k}$ the set of detectors that produced a 0, the probability that only k detectors produced a 1 is:

$$
P_{\mathcal{Q}_k} = \eta^k \prod_{j \in \mathcal{Q}_k} (1 - e^{-p_j\mu}) \prod_{u \in \mathcal{Z}_{l-k}} (1 - \eta + \eta\, e^{-p_j\mu})
$$

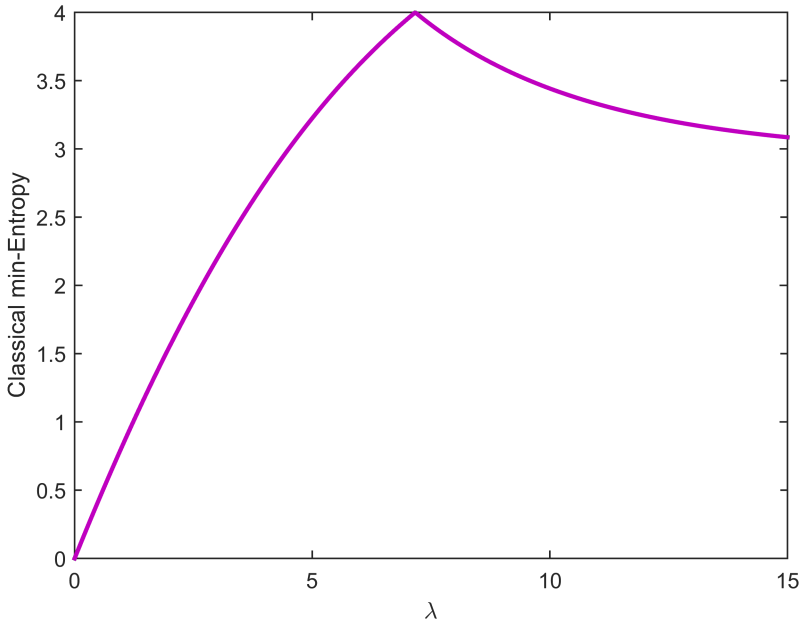
with the condition that $\mathcal{Q}_k + \mathcal{Z}_{l-k} = l$.

When $p_j = \frac{1}{l}\ \forall j$, the previous equation simplifies to:

$$
P_{\mathcal{Q}_k} = \eta^k (1 - e^{-p_j\mu})^k\, (1 - \eta + \eta\, e^{-p_j\mu})^{l-k} \tag{8.7}
$$



(a) Probabilities for $l = 4$ and $\eta = 60\,\%$



(b) Classical min-Entropy for $l = 4$ and $\eta = 60\,\%$

Figure 8.10: Probabilities and Classical min-Entropy for a system with $l = 4$ and quantum efficiency of $\eta = 60\,\%$. When all possible outputs are equally probable ($p = 1/16 = 0.0625$) the entropy reaches its maximum: 4.

The probabilities extracted using Eq. (8.7) corresponds to those of Tab. (8.1), this time for a generic number of detectors. Fig. (8.10a) shows the probabilities for $l = 4$ for a system of quantum efficiency $\eta = 60\%$. Using these probabilities, however, permits to retrieve the Classical min-Entropy, $H_\infty(X)$ (see Fig. (8.10b)), and not the Conditioned min-Entropy, $H_{min}(X|W)$, that is needed to account for the side information. Then, the next step is to compute the conditioned probabilities of all possible photons configuration for a particular set of sensitive or insensitive devices. The task, however, is not simple, and to show why it is, it is simpler to focus on a particular system so to have a better understand on the matter.

$l = 4$ DETECTORS

Assuming that there are four detectors, the first part of the analysis is to determine all possible configuration of Sensitive-Insensitive detectors. Tab. (8.4) summarizes the configurations. The total number of configurations is the maximum computable number with 4 bits: 16. The symmetry of the system, however, permits to use only a few of the table's entries because some of them produce the same results. For example consider all the configurations with only one insensitive detector. Because the detectors are all the same, the analysis' result using table's entry two will be the same as that of table's entries three, five and nine. In general, for s sensitive detectors and i insensitive detectors, the number of same configurations is retrieved from the binomial coefficient:

$$\binom{l}{s} = \frac{l!}{(l-s)!\,s!} = \frac{l!}{i!\,s!}$$

where the condition $(s + i = l)$ must be satisfied. Each of the table's entries designate a particular system from which determine all the probabilities of each output given the value of variables s, i, l and n. To understand how the process works, it is better to focus on a specific output, say $(1, 0, 1, 0)$. Of all the possible combination of Sensitive-Insensitive, only three [6] allows to retrieve the desired output. Tab. (8.3) shows these configurations. Splitting the problem in two parts permits to evaluate the wanted probabilities. The first part, as previously stated, regards the study of all the possible Sensitive-Insensitive configurations. The second part, instead focuses on the occupation numbers of the detectors and all the potential arrangements of a given configuration.

[6] As previously stated, it is possible to analyze some of all the configurations in order to remove redundancy on evaluating probabilities.

Cases	Detectors	Photons configuration
I	S S S S	$(>0, =0, >0, =0)$
II	S I S S	$(>0, \geq 0, >0, =0)$
III	S I S I	$(>0, \geq 0, >0, \geq 0)$

Table 8.3: Configurations of Sensitive-Insensitive detectors that allows to obtain the specific output $\{1, 0, 1, 0\}$. The photons configuration explain how the photons may dispose in the detectors always yielding the same output.

Sensitive-**I**nsensitive			
S	S	S	S
S	S	S	I
S	S	I	S
S	S	I	I
S	I	S	S
S	I	S	I
S	I	I	S
S	I	I	I
I	S	S	S
I	S	S	I
I	S	I	S
I	S	I	I
I	I	S	S
I	I	S	I
I	I	I	S
I	I	I	I

Table 8.4: Configurations of Sensitive-Insensitive detectors for $l = 4$. The total number of configurations correspond to the maximum computable number for 4 bit: 16.

Probability of case I

The probability relative to the case must be evaluated for all those configurations where $n_1 > 0$, $n_3 > 0$, $n_2 = 0$ and $n_4 = 0$.[7] Given n incoming photons, the configurations are of the kind $(n-m, 0, m, 0)$ with $m \in \{n-1, \cdots, 1\}$. This corresponds to the "integer partition" of n into two addends. For example, if $n = 5$, the appropriate configurations are:

$$\begin{aligned} 1\,0\,4\,0 &\leftrightarrow n = 5 \\ 2\,0\,3\,0 &\leftrightarrow n = 5 \\ 3\,0\,2\,0 &\leftrightarrow n = 5 \\ 4\,0\,1\,0 &\leftrightarrow n = 5 \end{aligned}$$

The total number of this kind of configurations is given by the number of ways to distribute n photons in $l^{'} = 2$ detectors decreased by the number of arrangements of all the photons in a single detector:

$$l^{'} = \binom{n + l^{'} - 1}{l^{'} - 1} - (l - l^{'})\binom{n + l^{'} - 1}{1}$$

Considering the photons to be distinguishable [45], the probability is given by a sort of *multinomial distribution*:

$$p_{\mathbf{I}}(1,0,1,0) = \sum_{n_3=1}^{n-1} \frac{n!}{(n-n_3)!\,n_3!}\frac{1}{4^n} \tag{8.8}$$

keeping in mind that each photon has intrinsic probability $p_i = 1/4$ to fall into a detector.[8] For the specific case of $n = 5$:

$$p_{\mathbf{I}}(1,0,1,0) = \frac{1}{1024}(5 + 10 + 10 + 5) = \frac{15}{512}$$

In a more compact form, the probability may be written as follow:

$$p_{\mathbf{I}}(1,0,1,0) = p(n_1 > 0 \cap n_2 = 0 \cap n_3 > 0 \cap n_4 = 0) \tag{8.9}$$

Probability of case II

In this case, one of the detectors is insensitive and may now receive any photons because it will never detect them and still give the same output. Now, the configurations of photons are those where $n_1 > 0$, $n_3 > 0$ and $n_4 = 0$ for the sensitive detectors while $n_2 \geq 0$ for the insensitive detector.

[7] The subscript refers to the specific detector.

[8] For a generic number l of detectors, the intrinsic probability is: $p_i = 1/l$.

Similarly to the previous case, the probability is:

$$p_{\mathbf{II}}(1,0,1,0) = \sum_{n_3=1}^{n-1} \sum_{n_2=0}^{n-n_3-1} \frac{n!}{(n-n_2-n_3)!\,n_2!\,n_3!}\frac{1}{4^n} \tag{8.10}$$

or in the more compact form

$$p_{\mathbf{II}}(1,0,1,0) = p(n_1 > 0 \cap n_2 \geq 0 \cap n_3 > 0 \cap n_4 = 0) \tag{8.11}$$

Considering the case with the fourth detector insensitive instead of the third would have given the same result thanks to the system's symmetry:

$$p(n_1 > 0 \cap n_2 \geq 0 \cap n_3 > 0 \cap n_4 = 0) \equiv p(n_1 > 0 \cap n_2 = 0 \cap n_3 > 0 \cap n_4 \geq 0)$$

Probability of case III

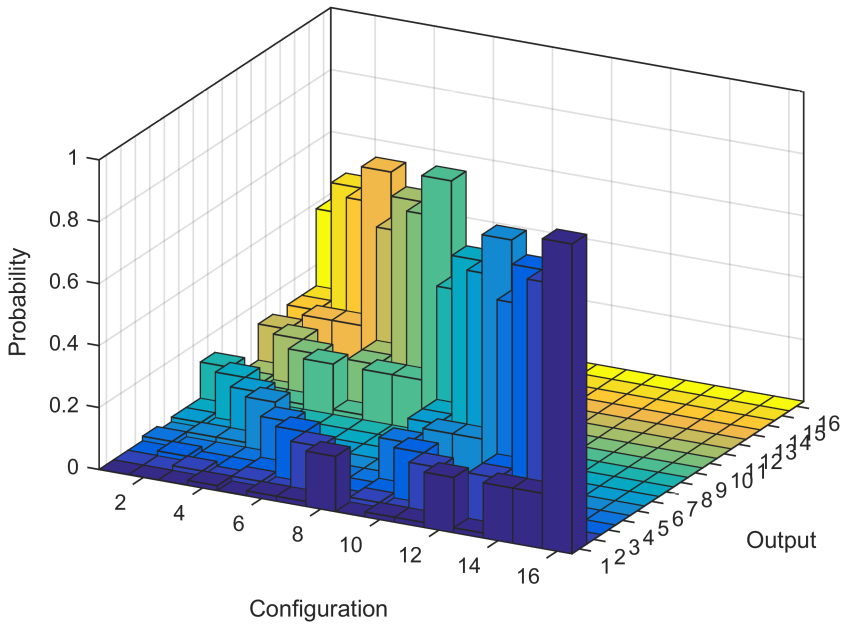
This time there are two insensitive detectors allowing both to receive any number of photons since they can not detect them. The probability then becomes:

$$p_{\mathbf{III}}(1,0,1,0) = \sum_{n_3=1}^{n-1} \sum_{n_2=0}^{n-n_3-1} \sum_{n_4=0}^{n-n_2-n_3-1} \frac{n!}{(n-n_2-n_3-n_4)!\,n_2!\,n_3!\,n_4!}\frac{1}{4^n} \tag{8.12}$$

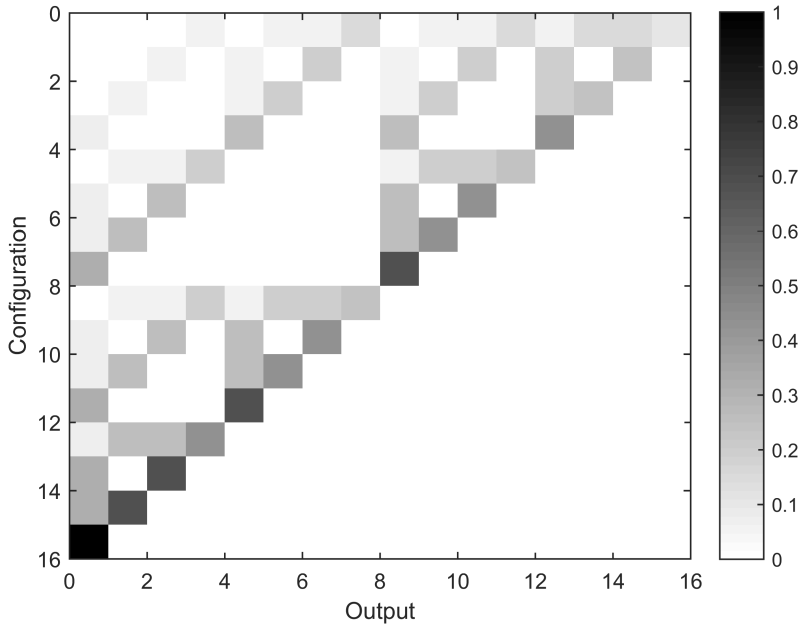
equivalently as

$$p_{\mathbf{III}}(1,0,1,0) = p(n_1 > 0 \cap n_2 \geq 0 \cap n_3 > 0 \cap n_4 \geq 0)$$

The theory and methodical approaches so far illustrate how to compute the probability for just one of all output made from $l = 4$ detectors. Computing all these probabilities for all possible configurations (Active-Inactive) and outputs $\{(0,0,0,0),(0,0,0,1),(0,0,1,0),(0,0,1,1),\ldots\}$ gives a "matrix" representation as shown in Fig. (8.11b). Fig. (8.11a) shows a 3D image where the height of each bar correspond to the value of the probability for a particular output and configuration. The two figures refer to a number of incoming photons of $n = 4$. The particular shape of the "matrix" resembles that of an upper triangular matrix. This is because in the lower-right zone the number of inactive detectors is greater than the number of 1s of a specific output. This means that, that particular entry (probability), is zero. Looking more carefully it is possible to notice some repetitive patterns like the repetition of a shape-like triangle of different size. Systems that presents this kind of behavior fall under the class of *fractals*. This particular pattern is called Sierpiński triangle. It is a fractal and attractive fixed set with the overall shape of an equilateral triangle, subdivided recursively into smaller equilateral triangles.



(a) 3D representation of conditioned probabilities for a system with $l = 4$ and $n = 4$



(b) Matrix representation of conditioned probabilities for a system with $l = 4$ and $n = 4$

Figure 8.11: 3D and matrix representation of all possible (Active-Inactive) configurations and outputs for a 4 detectors system. The second image resembles the Sierpiński triangle, a fractal with overall shape of an equilateral triangle, subdivided recursively into smaller equilateral triangles.

Originally constructed as a curve, this is an example of self-similar set, i.e. it is a mathematically generated pattern that can be reproducible at any magnification or reduction. The reason why the figure reproduces a fractal-like behavior is because of an internal system's symmetry. As previously stated, different sets produce the same probability. For example consider the case of the output $(0, 0, 0, 1)$ and the configurations (I S S S), (S I S S), (S S I S) and (S S S I). Tab. (8.5) resumes the probabilities for each configurations. The second column clearly shows that the criterions used to compute the probabilities are the same except for a mixing of the "$\geq$" symbol.

Output	
$(0, 0, 0, 1)$	
Configuration	**Probability**
(I S S S)	$p(n_1 \geq 0 \cap n_2 = 0 \cap n_3 = 0 \cap n_4 > 0)$
(S I S S)	$p(n_1 = 0 \cap n_2 \geq 0 \cap n_3 = 0 \cap n_4 > 0)$
(S S I S)	$p(n_1 = 0 \cap n_2 = 0 \cap n_3 \geq 0 \cap n_4 > 0)$
(S S S I)	$\varnothing$

Table 8.5: Table of redundant probabilities for the output $(0, 0, 0, 1)$. The criterions to compute the probabilities are the same except for a mixing of the "$\geq$" symbol. When the fourth detector is insensitive the relative probability is zero because the output can never be achieved.

This permits to reduce the number of configurations needed to compute every probability. The same reasoning can also be applied to the outputs. In fact, analyzing the case $(0, 0, 0, 1)$, a column like that of Tab. (8.5) may be attained considering the remaining same outputs: $(0, 0, 1, 0)$, $(0, 1, 0, 0)$ and $(1, 0, 0, 0)$. Through this simplification, the new matrix will have $(l+1)$ rows and columns, and each row is present in the previous figure (Fig. (8.11b)) exactly $\binom{l}{s}$ times while each column $\binom{l}{u}$, where u is the number of 1s in the specific output. Fig. (8.12) shows the new "matrix". Up until now the probabilities were calculated according to equations like Eqs. (8.8), (8.10) and (8.12).[9] Looking more closely to the system's symmetry, it is possible to recognize a specific relation between the rows and columns of the matrix of Fig. (8.12). In fact, each matrix's entry, M_{ij}, may be computed as

$$M_{ij} = \sum_{p=0}^{j} \binom{j}{p} (-1)^{j-p} (k+p)^n \frac{1}{l^n} \qquad \text{where } \begin{cases} 0 \leq i \leq l \\ 0 \leq j \leq l \end{cases} \tag{8.13}$$

This definition follows directly from the fact that the numbers of repetitions of rows and columns of the matrix in Fig. (8.11b) form a single row of the Pascal's triangle.[10]

[9] A complete table of conditioned probabilities for all configurations and outputs for a system with $l = 2$, $l = 3$ and generic n may be found in App. (C)

[10] Combining values for different l forms the Pascal's triangle.

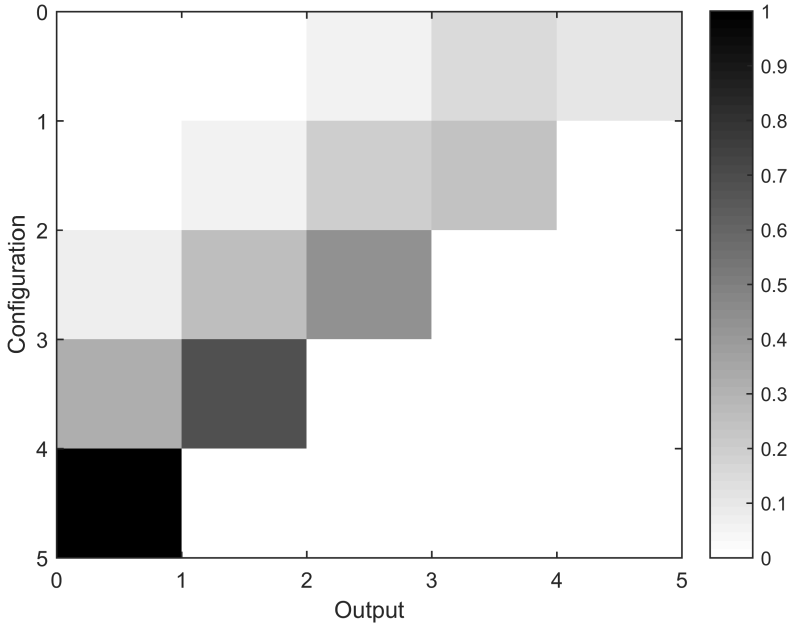


Figure 8.12: Simplified matrix representation for a 4 detectors system. In the general matrix of Fig. (8.11b) each row repeat itself exactly $\binom{l}{s}$ times while each column $\binom{l}{u}$ where u is the number of 1s in the specific output.

Tab. (8.6) shows a schematic example for $l = 4$ while Tab. (8.7) shows the Pascal's triangle for $l = 2, 3, \ldots, 10$. The bold row represents a 4 detectors system and the values are exactly those of Tab. (8.6).

Output				
$(0,0,0,0)$	$(0,0,0,1)$	$(0,0,1,1)$	$(0,1,1,1)$	$(1,1,1,1)$
$\binom{4}{0} = 1$	$\binom{4}{1} = 4$	$\binom{4}{2} = 6$	$\binom{4}{3} = 4$	$\binom{4}{4} = 1$

Table 8.6: Schematic example of how columns repetitions form a single row of the Pascal's triangle. The same relation follows even for all possible Active-Inactive configurations.

Because the matrix of Fig. (8.12) has been made considering the system's symmetry, combining configurations with outputs is, somehow, similar to computing the *binomial expansion* or the *binomial theorem*. The theorem allows to compute the algebraic expansion of powers of a binomial according to the following formula:

$$(x+y)^q = \sum_{r=0}^{q} \binom{q}{r} x^{q-r} y^r \equiv \sum_{r=0}^{q} \binom{q}{r} x^r y^{q-r} \tag{8.14}$$

It is easily observed a close analogy between Eqs. (8.13) and (8.14).

$$
\begin{array}{c}
1, 2, 1 \\
1, 3, 3, 1 \\
\mathbf{1,4,6,4,1} \\
1, 5, 10, 10, 5, 1 \\
1, 6, 15, 20, 15, 6, 1 \\
1, 7, 21, 35, 35, 21, 7, 1 \\
1, 8, 28, 56, 70, 56, 28, 8, 1 \\
1, 9, 36, 84, 126, 126, 84, 36, 9, 1 \\
1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1
\end{array}
$$

Table 8.7: Example of Pascal's triangle for $l = 2, 3, \ldots, 10$. The bold row corresponds to a 4 detectors system.

Now, the model's purpose is to retrieve the maximum probability for a particular configuration for all outputs and use it to compute the Conditioned min-Entropy. Making use of Eq. (8.3) it is possible to generalize its formula as follows:

$$
H_{min}(X|NT_1 \ldots T_l) = -\log_2 \left[P_N(0) + \sum_{n=1}^{\infty} \sum_{t_1 \ldots t_l} P_N(n) \left(D \cdot E \cdot F \right) \right] \tag{8.15}
$$

$$
\text{where} \begin{cases} D &= \binom{l}{s} \\ E &= \eta^s (1-\eta)^i \\ F &= \max\limits_x P_{X|NT_1 \ldots T_l}(x|nt_1 \ldots t_l) \end{cases}
$$

where the variables $(T_1 \ldots T_l)$ assume the values s or i according to the specific combination. The binomial coefficient accounts for all repetitive configurations. It is fundamental, however, to stress out that each maximum probability is retrieved from a specific Sensitive-Insensitive combination but generic output x. Fig. (8.13) shows the Conditioned min-Entropies relative to a system with $l = 3, 4, 5, \ldots, 16$ detectors for varying λ values. The figure shows that for increasing number of detectors the entropy's maximum shifts to higher values of λ. This behavior is due to the increasing probability of events where most of the detectors register a photon. Because the entropy reaches its maximum when the probabilities of all possible outcomes are equal, as the number of detectors increases it is necessary that a larger number of photons impinge on the detectors. This way, the probability that all detectors register an event becomes equal to all others so to increase the entropy.

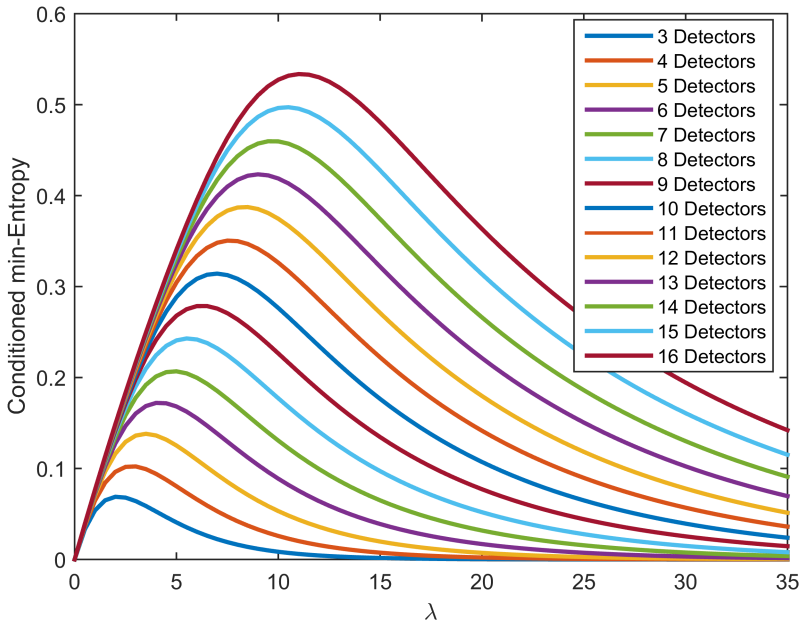


Figure 8.13: Conditioned min-Entropies relative to a system with $l = 3, 4, 5, \ldots, 16$ detectors. All of them refer to a quantum efficiency of $\eta = 5.47\,\%$.

8.3 Generalized model with detector noise

The model presented in Chap. (1) showed a simplified system where, except for the quantum efficiency of the detectors, no noise effect were taken into account. In more realistic system, there are always secondary effects that disturb a measurement. As introduced in Chap. (4), the main influences found in this kind of devices are: *Crosstalk*, *Dark count* and *Afterpulses*.

As in Example 2, it is possible to define a state and measurements such that the side information corresponding to the noise is encoded in a maximum classical noise model. Because those definitions are straightforward, it is possible to directly introduce the random variables resulting from the model.

- **Incoming photons**

As for the previous model, the number of photons emitted by a source is encoded as a random variable N with outcomes $n \in [0, \ldots \infty]$ distributed according to a Poisson distribution

$$P_N(n) = e^{\lambda} \frac{\lambda^n}{n!}$$

- **Detector's sensitivity**

The sensitivity of the detectors corresponds to the quantum efficiency μ and is modeled, for each detector $R_{1,2,\dots,l}$, by a random variable $T_{1,2,\dots,l}$ with outcomes $t_{1,2,\dots,l} \in [0,1]$. The distribution is then

$$P_{T_{1,2,\dots,l}}(t_{1,2,\dots,l}) = \mu^s$$

where s corresponds to the number of sensitive detectors for a specific configuration of $t_{1,2,\dots,l}$.

- **Noise**

Dark counts, Afterpulses and Crosstalk correspond to the side information whether a detector "clicks" independently of any incoming photons. In the same way for the sensitivity, this too can be modeled by a random variable $S_{1,2,\dots,l}$ with outcomes $s_{1,2,\dots,l} \in [0,1]$, where for $s_{1,2,\dots,l} = 1$ corresponds to such deterministic "click". The distribution is then expressed as follow

$$P_{S_{1,2,\dots,l}}(s_{1,2,\dots,l} = 1) = 1 - (1 - p_\varrho)(1 - p_\gamma)(1 - p_\delta)$$

where p_ϱ is the Dark count probability, p_γ is the Afterpulses probability and p_δ is the Crosstalk probability.

Because the joint distribution $P_{S_1,S_2,\dots,S_l}(s_1, s_2, \dots, s_l)$ is in general not equal to the product of the distributions $P_{S_1}(s_1)P_{S_2}(s_2)\dots P_{S_l}(s_l)$, it is necessary to compute each conditioned probability. However, looking more closely to the system's settings, it is possible to recognize that the probabilities are actually the same as those of the simpler model. In fact, the criterions used to compute them do not change. This way, after some rearrangements, it is possible to retrieve a matrix similar to that of Fig. (8.12). Now, for the calculation of $H_{min}(X|NS_1S_2\dots S_lT_1T_2\dots T_l)$ it has been minimized over all $P_{S_1S_2\dots S_l}(s_1s_2\dots s_l, y)$ subject to the constraint $P_{S_{1,2,\dots,l}}(s_{1,2,\dots,l} = 1) \coloneqq p$ where y is the free parameter $(0 \leq y \leq p)$. This is the same as placing the system in the "worst case scenario".

The Conditioned min-Entropy is then

$$H_{min}(X|NS_1S_2\dots S_lT_1T_2\dots T_l) = \min_y \left\{ -\log_2 \left[P_N(0) + \sum_{n=1}^{\infty} P_N(n) \left(\sum_{\substack{s_{1,2,\dots,l} \\ t_{1,2,\dots,l}}} A \cdot B \cdot C \right) \right] \right\} \quad (8.16)$$

$$\text{where } \begin{cases} A = P_{T_1}(t_1)P_{T_2}(t_2)\dots P_{T_l}(t_l) \\ B = P_{S_1S_2\dots S_l}(s_1s_2\dots s_l, y) \\ C = \max_x P_{X|S_1S_2\dots S_lT_1T_2\dots T_l}(x|s_1s_2\dots s_lt_1t_2\dots t_l) \end{cases}$$

Fig. (8.14) shows the new Conditioned min-Entropy for a system with $l = 3, 4, 5, \ldots, 16$ detectors and quantum efficiency of $\eta = 5.47\,\%$.

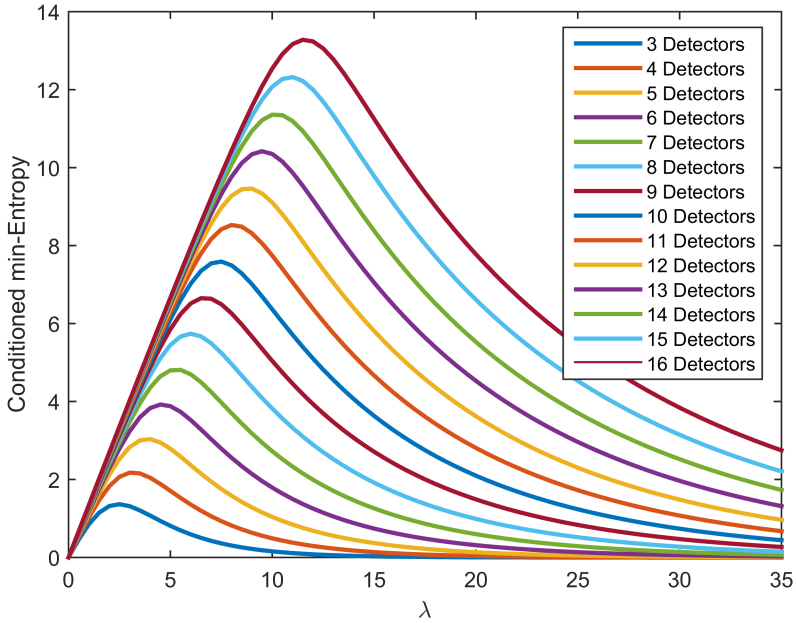


Figure 8.14: Conditioned min-Entropies relative to a system with $l = 3, 4, 5, \ldots, 16$ detectors in the presence of noises ($p_\varrho = 0.204$, $p_\gamma = 0.0429$, $p_\delta = 9.67 \times 10^{-5}$). All of them refer to a quantum efficiency of $\eta = 5.47\,\%$.

Making use of Eq. (8.16) and the information from Chap. (7), it is possible to extract uniformly distributed variables conditioned on side information. Having available a 32x32 matrix of pixels from the acquisition camera, it has been subdivided the entire frame in four smaller matrices of 256 pixels each. Figs. (8.15) and (8.16), show the respective calculated values of λ and H_{min} for each matrix subsequently printed directly on the 32x32 frame. This kind of divisions is possible only for the LED acquisitions. For the LED source, the photon distribution is uniform and then, the value of λ, should be the same for each pixel of each sub-matrix. A LASER acquisition, on the contrary, has a characteristic photon distribution so that, each pixel of each sub-matrix, is different. This leads, then, to different values of λ (and consequently H_{min}) for all four matrices.

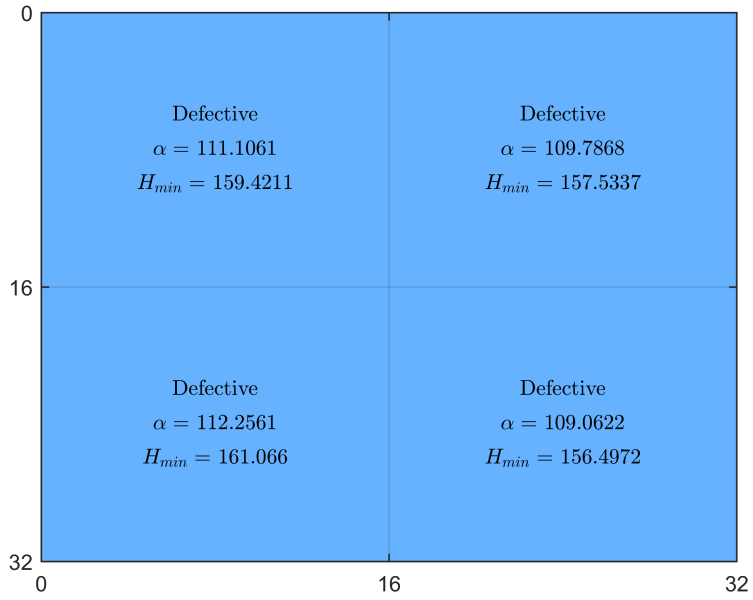


Figure 8.15: Division of a 32x32 frame in four sub-matrices of 256 pixels. On each matrix is printed the respective value of λ and H_{min}. The system's quantum efficiency is $\eta = 41.4\,\%$ and the frame refers to a LED acquisition of working power $P = 560\,nW$.

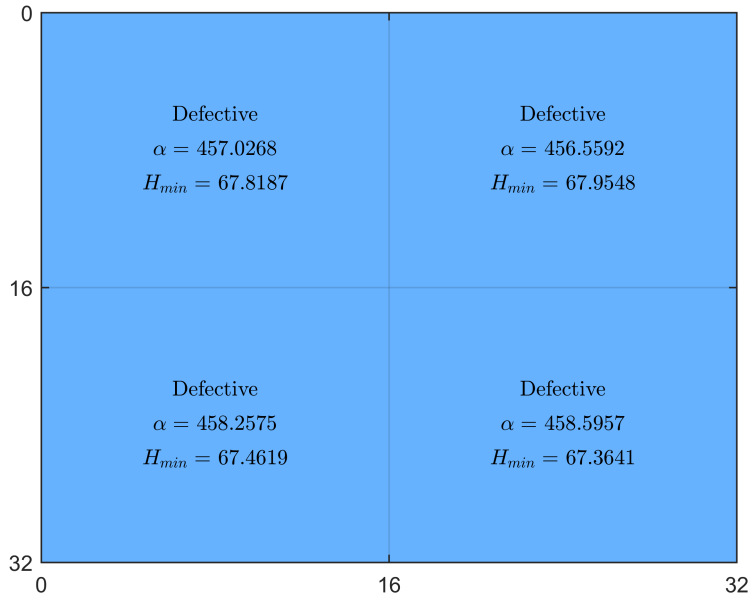


Figure 8.16: Division of a 32x32 frame in four sub-matrices of 256 pixels. On each matrix is printed the respective value of λ and H_{min}. The system's quantum efficiency is $\eta = 41.4\,\%$ and the frame refers to a LED acquisition of working power $P = 5.269\,\mu W$.

9 Randomness extractors

The output of a generic generator shows a lack of uniformity and independence. Even if the physical process and extraction mechanism are theoretically strong, the loss of randomness is unavoidable because of the employment of not ideal electronics and devices. The first section presents classical random extractors applied to the raw data extracted from the system while second section shows a quantum extractors combined with the concept of entropy discussed in Chap. (8).

9.1 Classical randomness extractors

In TRNGs is common to involve a stage of post-processing of the acquired raw data from a physical system. Postprocessing applies mathematical functions in order to enhance random properties of bit strings. Computer Theory develops these functions, also called *randomness extractors*. The theoretical framework considers the fact that sources of "high entropy", i.e. sources that hold the genuine physical randomness at the origin, produce raw bits. The usage of extractors required to compensate the loss of uniformity and independence of bits caused by electronics that make the source "weak". [46, 47, 48] The first classical randomness extractor is the one developed by J. von Neumann.

Von Neumann Extractor

The *Von Neumann Extractor* is the most renown extractor in Computer Theory. J. von Neumann presented this extraction method in his famous paper "*Various techniques for use in connection with random digits*" in 1951. [5] The procedure applies a mapping to each pairs of an input sequence $x_1, x_2, \ldots, x_n$ generated by a process $X_v(p)$ which chooses x_n from $\{0, 1\}$ with independence and uniform bias. For all n bits, $x_1 = 1$ with probability p and $x_i = 0$ with probability $q = 1 - p$, where p is unknown but fixed $(0 < p < 1)$. The mapping then extracts bits in the following manner:

$$\begin{aligned} 00 &\rightarrow \varnothing \\ 01 &\rightarrow 0 \\ 10 &\rightarrow 1 \\ 11 &\rightarrow \varnothing \end{aligned}$$

Tab. (9.1) shows how the extract works for a given bit string. The method is useful because, given p, the chance of the pair (01) is the same as that of the pair (10) since

$P(01) = qp = pq = P(10)$. So, using the symmetry, it is possible to produce a perfect unbiased string without knowing the probability p.

Bit string										
10001110001001110101										
	Von Neumann Extractor									
Biased	10	00	11	10	00	10	01	11	01	01
Unbiased	1		1		1	0		0	0	

Table 9.1: Representation of the Von Neumann Extractor. For a given bit string the algorithm takes pairs of bits and produces a bit whenever they meet the algorithm's condition.

Von Neumann defined the efficiency of the procedure as the expected number of output digits per input digit. For each input pair the probability of generating non-null output digit is $pq + qp = 2qp$, so the efficiency is just $^{2qp}/_{2} = qp$, which is $^{1}/_{4}$ when $p = q = {}^{1}/_{2}$ and less elsewhere. This way, it is true that the map creates string independent of the probability p, but the efficiency is not. Besides this fact, the implementation requires independent bits to guarantee a true random output since, in the opposite case, the symmetry of pairs probabilities could be compromised.

The second classical extractor widely implemented for its great efficiency is the Elias Extractor made by Peter Elias.

Elias Extractor

P.Elias presented his extraction method in the paper *The efficient construction of an unbiased random sequence* in 1927. [4] The aim of this extractor is to achieve a great efficiency maintaining the request of independence and uniformity of a bit string. The method is indeed simple and efficient. Given a biased N-bit long string, the system divides the set of 2^N possible input sequences into the $N+1$ composition classes S_k with $0 \leq k \leq N$, containing the $\binom{N}{k}$ sequences of length N which have k ones and $N-k$ zeros. The next step is to consider the binary expansion of $\binom{N}{k}$:

$$\binom{N}{k} = \xi_{n_k}\, 2^{n_k} + \xi_{(n_k-1)}\, 2^{(n_k-1)} + \cdots + \xi_0\, 2^0$$

with

$$n_k = \left\lfloor \log_2 \binom{N}{k} \right\rfloor$$

where $\lfloor \cdot \rfloor$ denotes the largest integer so to ensure that $\xi_{n_k}, \xi_{(n_k-1)}, \cdots \xi_0$ is the binary expansion of $\binom{N}{k}$ with $\xi_{n_k} = 1$, $\xi_j = 0$ or 1 for $0 \leq j < n_k$. For each non-vanishing ξ_j, $0 \leq j \leq n_k$, the system assigns the 2^j possible output binary sequences of length j to 2^j distinct members of S_k which have not already been assigned. One member of S_k will be assigned to no output if $\xi_0 = 1$ so that S_k is odd. S_0 and S_N have only one member each, which is therefore assigned to no output.

For a practical example, suppose that α is a random, but biased, number, β another random number grater than α ($\beta > \alpha$) and $T = \lfloor \log_2 \beta \rfloor$. If $\alpha < 2^T$, T bits may be extracted; when $2^T \leq \alpha < 2^T + \xi_{(T-1)}\, 2^{(T-1)}$, $T-1$ bits may be extracted and so on till $\alpha = \beta - 1$ and $\xi_0 = 1$. In the latter no string is assigned.

The last classical extractors used in this research work is based on the work of Yuval Peres [7] on iterating the Von Neumann extractor. In particular, M. Mitzenmacher developed a recursive method to Peres' system called **A**dvanced **M**ulti**L**evel **S**trategy (AMLS). [6]

Advanced MultiLevel Strategy (AMLS) Extractor

The AMLS extractor comes from the essential principle of Peres' unbiasing technique to reiterate v. Neumann method on the discarded bits. Its efficiency lies in the use of additional symmetries with respect to the v. Neumann one. The first symmetry is between the case (0011) and (1100). In the original extraction scheme, no fair bits are obtained. However, if the system produces a 0 in the first case and a 1 in the other, the probability symmetry is still maintained ($p^2q^2 = q^2p^2$), while increasing the chances of producing a fair bit (see Tab. (9.2)). A simple way to analyze this is to build a new bit string in the following way: whenever there is a pair of bits that are the same in the original sequence, the system introduce a new bit of that type into the new sequence and then performs the Von Neumann Extractor. Tab. (9.3) shows the process for a given string.

Rules		Bit String
$11 \to 00 = 1$	Biased	10**11**01**0000**100**11**
$00 \to 11 = 0$	Unbiased	1 0

Table 9.2: Representation of the system's symmetry. For a given biased sequence, a new one is generated applying the rules of the first column.

Defining $B(p)$ a function that represents the average number of fair bits retrieved its possible to understand how this implementation improves the number of fair bits

Rules		Bit string
$11 = 1$ $00 = 0$	Biased	1011010000100111
	Numerical	1 0 0 1
	Unbiased (v. N.)	1 0

Table 9.3: Representation of the AMLS *Numerical* procedure. A new sequence is generated applying the rules of the first column. Subsequently the system applies the Von Neumann Extractor.

obtained from the initial bit string. A pair of same bits (00 or 11) have probabilities q^2 and p^2 respectively, so, on average, the system provides $(p^2+q^2)/2$ additional recursive bits per original bit. Also, each bit at the new sequence, is 1 with probability $p^2/(p^2+q^2)$ and 0 with probability $q^2/(p^2+q^2)$. This way, $B(p)$ becomes:

$$B(p) = pq + \frac{p^2+q^2}{2} B\left(\frac{p^2}{p^2+q^2}\right)$$

when $p = q = 1/2$, the equation gives $B(1/2) = 1/4 + 1/4\, B(1/2) \Rightarrow B(1/2) = 1/3$.

To improve the system, the AMLS procedure implement a secondary symmetry. The cases (1110) or (1011), produce one fair bit (v. Neumann procedure) but do not account for the order in which these two events happen. Considering this fact, the system can produce another fair bit thanks to the symmetry (see Tab. (9.4)). To extract the extra bit, the method creates an additional sequence of bits again. From the original sequence, whenever two bits are the same a 1 bit is produced while if two bits are different, a 0 is produced. Subsequently the Von Neumann Extractor is applied to the new sequence as well as the original one. Tab. (9.5) shows the process for a given string.

Rules		Bit String
$\begin{cases}0100\\0001\end{cases}$	$= 0$ Biased	1011010000100111
$\begin{cases}1110\\1011\end{cases}$	$= 1$ Unbiased	1 1 0 1

Table 9.4: Representation of the system's symmetry. For a given biased sequence, a new one is generated applying the rules of the first column.

As done before, it is possible to determine a long-term average number of bits produced. The equation is similar to the previous case, except now for every two bits, the system gets an additional bit in one of the derived sequences. This bit is 1 with probability $p^2 + q^2$ since it comes up 1 whenever the pair of bits are the same.

Rules		Bit string
11 or 00 = 0 10 or 01 = 1	Biased	1011010000100111
	Alphabetical	1 0 1 0 0 1 1 0
	Unbiased (v. N.)	1 1 0 1

Table 9.5: Representation of the AMLS *Alphabetical* procedure. A new sequence is generated applying the rules of the first column. Subsequently the system applies the Von Neumann Extractor.

Hence the resulting equation is:

$$B(p) = pq + \frac{p^2+q^2}{2} B\left(\frac{p^2}{p^2+q^2}\right) + \frac{1}{2} B\left(p^2+q^2\right)$$

when $p = q = 1/2$, the equation gives $B(1/2) = 1/4 + 1/4\, B(1/2) + 1/2\, B(1/2) \Rightarrow B(1/2) = 1$.

Now the procedure retrieves 1 fair bit, accordingly to a perfectly uniform and independent system. In fact, in the limit, this process extracts the maximum number of fair bits possible for every value of p. Naturally, the first and the second procedure, called respectively *Numerical* and *Alphabetical*, may be applied recursively to the newly generated sequences. Tab. (9.6) shows a figurative example.

Level	Bit string	Von Neumann Extractor
Biased	11101100101110111110011 1	111110
N_{Bias}	1 1 0 1 1 1 1	0
$N[N_{Bias}]$	1 1	
$A[N_{Bias}]$	0 1 0	0
A_{Bias}	0 1 0 0 1 0 1 0 0 1 1 0	01101
$N[A_{Bias}]$	0	
$A[A_{Bias}]$	1 0 1 1 1 1	1
$N[A[A_{Bias}]]$	1 1	
$A[A[A_{Bias}]]$	1 0 0	1
Unbiased		11110000110111

Table 9.6: Representation of a recursive implementation of the AMLS Extractor. Both *Numerical* and *Alphabetical* procedure are applied to each new sequence retrieving an unbiased bit string made from the sequences elaborated with the Von Neumann Extractor.

9.2 Quantum randomness extractor

Differently from classical extractors, quantum extractors do not use particular complex and difficult to predict algorithm in order to produce unbiased bit string. The potency of a quantum approach is to exploit the very probabilistic nature of a physical process. Chap. (1) showed how it was possible to extract a truly random output from a physical process conditioning on all side information available to an adversary. The *Conditioned min-Entropy* plays a significant role in the extraction because it is an extremely useful tool to quantify and study intrinsic randomness of a system. The algorithm to extract a truly random output is rather straightforward. The first step is to obtain the value of $H_{min}(X|E)$ relative to a particular bit string and multiply it by the original bit string length. The *Quantum Leftover Hash Lemma* ensures that whenever the new sequence length l is smaller than the Conditioned min-Entropy, $l < H_{min}(X|E)$, the output of an hash function $f(x)$ is uniform and independent of the system E, except with probability $\epsilon_{hash} < 1$. It is sufficient to choose the hash function once, using randomness that is independent of the device. The Lemmas and definitions in Chap. (1) holds if the hash function is extracted from the bigger family of *Two-universal hashing functions*. A simple and computationally fast to implement function is the modulo. The system determines a suitable bit length l, for a given bit string of length n, and produces a random matrix, $(l \times n)$, c_{ij}, through which define the two-universal hashing function $Y = f(x)$. The extraction then continues easily performing a matrix product between the matrix c_{ij} and the bit string X_j $(j = 1, \ldots n)$ modulo 2:

$$Y_i = \sum_{j=1}^{n} \mathrm{mod}_2 \left(c_{ij} X_j\right) \tag{9.1}$$

with Y_i with $i = 1, 2, \ldots l$.

The new bit string Y is indeed truly random and independent of any "physical noise" presents in the system. It is straightforward to see that Y is always smaller than the original bit sequence X. This procedure uses a two-universal function that relies on an external (and independent) source of randomness and therefore $f(x)$ may be selected already when manufacturing a QRNG device and hardcoded on the instrument. [20]

10 Statistical tests

Statistical tools are essential to assess the degree of randomness of a number. These tools belong to a distinct branch of Statistics, *Hypothesis Testing*, and are commonly used in works regarding generic RNGs. The first section of the chapter presents the concepts of test statistics, $\mathcal{P}$-*values* and statistical distributions. The second section shows different tests examples focusing on the tests of the NIST (**N**ational **I**nstitute of **S**tandard and **T**echnology) suite. [49]

10.1 Statistical Hypothesis Testing

The model proposed in this work takes the task to study the processes that generate random numbers and ask to consider them truly random if the outcomes are uniformly distributed *and* independent of any information available in advance. Studying the bits produced by applying a broad statistical analysis permits to judge how well the generator fits its model. A test's aim is to evaluate a *trait* of a bit sequence and show if it is *likely* that an ideal generator may reproduce the same string. Another way to put it, is to say that the test tries to *see* if a generator is compatible with the requirements of producing string of bits uniform and independent. The test, of course, does not provide an actual response of the sort *yes* or *no*, but gives a probabilistic evaluation of the hypothesis. [50] The assumption that a bit string is random is called the *null-hypothesis* symbolized with $\mathcal{H}_0$. A successful result from the test, or test "passed", means that $\mathcal{H}_0$ is *not rejected* for that string, while a failure, or test "failed" means *accepting* the alternative hypothesis, $\mathcal{H}_1$. So, in this case, if the test passed the string is assumed random, however if the test failed the string is not considered random. Because the number of tests is unlimited as the number of *traits*, it is impossible to test a bit string exhaustively. It is commonly recognized, however, that if different types of tests pass, the more evidences do support the *null-hypothesis* $\mathcal{H}_0$.

Test statistic, or *test variable*, is a specific value that permits to link a particular *trait* of a bit string to the probabilities' framework. Let D_N be the set of binary bit strings, d_N, of length N. The way a test verify the plausibility of $\mathcal{H}_0$ for a string d_N consists in defining a *test function* q which maps the strings to real numbers $\mathbb{R}$:

$$\begin{aligned} q:\ D_N &\longrightarrow \mathbb{R} \\ d_N &\longmapsto q(d_N) \end{aligned} \tag{10.1}$$

where $q(d_N)$ is the *test statistic* or *test variable*.

As a matter of fact, $q(d_N)$ is a random variable but, if $\mathcal{H}_0$ holds, it follows a specific and *known* probability distribution. This means that if d_N is truly random, the function q produces test statistics distributed according to a particular distribution i.e. normal, chi-square etc., specific to the trait tested and the chosen function q. When the particular distribution is confirmed, the statistical testing determines a *critical value* to compare with the test statistics. If the test statistic value exceeds the critical value $\mathcal{H}_0$ is rejected, otherwise $\mathcal{H}_1$ is rejected. The reason why the statistical testing works is because the reference distribution and the critical value depend on and generate under the unconfirmed assumption of randomness. For a given data sample, if the randomness assumption is indeed correct, the relative calculated test statistic will have a low probability (e.g. 0.01%) of exceeding the critical value. Indeed, because of their random nature, the test statistics, now referred to as X, assume possible values x that are more or less likely to appear according to the probability distribution $f(x)$ associated with X. The way a test provides this evidence consists in estimating the probability that, under $\mathcal{H}_0$, the use of q on d_N can return a value incompatible with x, i.e. $P\left(X \geq x | \mathcal{H}_0\right)$. This probability, estimated on x and indicated as $\mathcal{P}(x)$, is named $\mathcal{P}$-*value*. Its mathematical definition is as follow:

$$\mathcal{P}(x) = P\left(X \geq x \,|\, \mathcal{H}_0\right) = \int_x^{\infty} f(x^{'})\, dx^{'} \tag{10.2}$$

Another way to express the meaning of the $\mathcal{P}$-*value* is to consider it as the value that summarize the strength of the evidence against the null-hypothesis. [51, 52]

Statistical hypothesis testing provides a two possible outcome systems: accept $\mathcal{H}_0$ or $\mathcal{H}_1$. Tab. (10.1) shows the potential conclusions of the analysis given the exact status of the given data.

SITUATION	**CONCLUSION**	
	Accept $\mathcal{H}_0$	Accept $\mathcal{H}_1$
Data random ($\mathcal{H}_0$ is true)	No error	Type I error
Data not random ($\mathcal{H}_1$ is true)	Type II error	No error

Table 10.1: Representation of the possible results after applying a statistical hypothesis testing to a given data set.

Looking at Tab. (10.1) there are some combinations that provide two different type of errors. The *Type I error* corresponds to the event where the null-hypothesis is rejected, even though the data is truly random. The *Type II error* instead is when the null-hypothesis is accepted, even though, the data is not random. The probability of Type I error is usually called the *level of significance* of the test and denoted as α.

This parameter then indicates the probability that a bit string is not considered random when it really is. The probability of Type II error instead is called β and, unlike α, does not have a fixed value. This is because there are an infinite number of ways that a bit string may be "non-random" and each of them yields a different β. It is desirable, of course, to minimize the probability of this error, and that is achieved by pre-selecting the sample size n of the data and the value α because α, β and n, are strictly related to each other so that knowing two of them automatically determine the other. Thanks to this relation, after applying the test statistic for a sample size n and a value for α, a critical value is selected so to produce the smallest β possible. These two types of error may be expressed in terms of conditioned probability. Using the same notation as Eq. (10.2), and denoting c the critical value, the Type I and II error becomes:

$$\begin{aligned} P(X > c \,|\, \mathcal{H}_0 \text{ is true}) &= P(\text{reject } \mathcal{H}_0 \,|\, \mathcal{H}_0 \text{ is true}) && \text{Type I error} \\ P(X \leq c \,|\, \mathcal{H}_0 \text{ is false}) &= P(\text{accept } \mathcal{H}_0 \,|\, \mathcal{H}_0 \text{ is false}) && \text{Type II error} \end{aligned}$$

In view of the above explanation, the definition of the $\mathcal{P}$-*value* should be restated as follow: the $\mathcal{P}$-*value* is the probability that a perfect RNG would have produced a sequence less random than the one that is being tested. If a $\mathcal{P}$-*value* is equal to 1, then the bit string appears to have "perfect randomness". If it is equal to 0, instead, the string appears to be absolutely not random. The limit value of $\mathcal{P}$-*value* that defines if a string should be considered random is the α parameter itself. If $\mathcal{P}$-*value* $\geq \alpha$, then the null-hypothesis is accepted, i.e. the bit string "appears" to be random. Otherwise, if $\mathcal{P}$-*value* $< \alpha$, the alternative hypothesis is accepted and the string "appears" to be non-random. The verb "*appear*" has been used because of the statistical and probabilistic nature of the Hypothesis Testing. In cryptography, ordinary values of α lie in the range [0.001, 0.01] corresponding to a *confidence level* of 99.9 % and 99 % respectively.

As stated above, test statistics follow a specific and known statistical distribution according to the chosen test function and the validity of the randomness assumption. There are two principal distributions that most of the test statistics follow: the *standard normal* distribution and the *chi-square* distribution, or equivalently χ^2.

Normal distribution:

A random variable X follows the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2 if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{10.3}$$

If the test statistic X is normally distributed then the closer the value x is to μ, the "better" random traits the bit string has. Measuring by how much the test statistic

deviates from the expected value μ, the test *quantifies* the evidence of no-randomness of the bit string. Given the symmetric profile of the distribution both $x > \mu$ and $x < \mu$ provide the same probabilities and the $\mathcal{P}$-*value* is:

$$\mathcal{P}(x) = P\left(X > |x|\right) = 2P\left(X > x\right) = 2\int_{x}^{\infty} f(x')\,dx' \qquad -\infty < x < \infty$$

where this type of test is called a *two tails test*.

For a practical example, setting a level of significance of 1% and using a standard normal distribution $\mathcal{N}(0,1)$, all test statistics with $x \geq |2.576|$ are rejected since

$$2\int_{2.576}^{\infty} f(x)\,dx = 0.01$$

Chi-Square or χ^2 distribution

A variable X follows a chi-square distribution with d degrees of freedom, mean $\mu = d$ and variance $\sigma^2 = 2d$, if its probability density function is

$$f(x) = \frac{1}{\Gamma(d/2)\,2^{d/2}} x^{d/2-1} \exp\left[-\frac{x}{2}\right] \qquad 0 \leq x < \infty \tag{10.4}$$

The system usually performs a *one tail test* with a level of significance set only at the rightmost side of the distribution. This way, a bit string appears to be not random if the test statistic value is far away from the mean value.

10.2 RNG Tests

There exist many different statistical tests available to verify a bit string randomness. Most of them are contained in various *Testing Suite* freely accessible online. A *suite* is a collection of tests specifically developed by scientific institution or researchers in the field of RNGs and cryptography. Depending on the purpose of the analysis made on the system, even suites may be subdivided into two main branches: the cryptography and the statistical branch. The first applies a "bit-approach" considering the bit string as a sequence of 0s and 1s, while the latter applies a "number-approach" studying the qualities of numbers extracted from the same string taking more bits at a time (usually 32). This research project focuses on the randomness analysis especially for cryptographic purpose and so it was decided to perform statistical tests by the usage of the NIST suite.

The Statistical Test Suite developed by the American **N**ational **I**nstitute of **S**tandard and **T**echnology (NIST) applies stringent tests in order to validate the reliability of PRNGs and TRNGs (in which QRNGs appear) for cryptographic applications. The suite is widely used and thus counted as a standard. Tab. (10.2) shows the tests implemented

in the suite. Test statistics retrieved from the suite follow the normal and chi-square distributions. The test statistic for the normal distribution is of the form

$$z = \frac{x - \mu}{\sigma}$$

where x is the sample test statistic, μ and σ^2 are the expected value and variance of the test statistic respectively. The chi-square distribution (i.e. a left-skewed curve) compares the goodness-of-fit of the observed frequencies of a sample measure to the corresponding expected frequencies of the hypothesized distribution. The test statistic is of the form:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

where o_i and e_i are the observed and expected frequencies of occurrence of the measure respectively.

Statistical Test Suite NIST	
Nr.	Test
1	Frequency (Monobit)
2	Frequency within a Block
3	The Runs
4	Longest-Run-of-Ones in a Block
5	Binary Matrix Rank
6	Discrete Fourier Transform (Spectral)
7	Non-overlapping Template Matching
8	Overlapping Template Matching
9	Maurer's "Universal Statistical"
10	Linear Complexity
11	Serial
12	Approximate Entropy
13	Cumulative Sums (Cusums)
14	Random Excursions
15	Random Excurions Variant

Table 10.2: List of tests implemented in the NIST suite.

When applying the NIST suite, the developers specify that the size of a sequence length must be large (of the order of 10^3 to 10^7) [53, 54, 55, 56] so to justify the derivation of the asymptotic reference distributions. Most of the test are still applicable for a smaller sequence length, but the asymptotic reference distributions would be unsuitable and would have to be replaced by the exact distributions commonly hard to calculate. The following list analyzes each single test and explains what its purpose is and how it retrieves the $\mathcal{P}$-*value*.[1] The mathematical functions employed for the computation of the $\mathcal{P}$-*value* are summarize in App. (B). All the tests refer to a bit string of length n.

1. Frequency (Monobit) Test

The purpose of this test is to determine whether the number of ones and zeros in a sequence is approximately the same as expected for a truly random sequence. [57] All subsequent tests depend on the passing of this test. The reference distribution for the test statistic is half normal (for large n). If the sequence is random, then the plus and minus ones will tend to cancel one another out so that the test statistic will be about 0. If there are too many ones or too many zeroes, then the test statistic will tend to be larger than zero. The $\mathcal{P}$-*value* is computed as follow:

$$\mathcal{P}\text{-}value = \mathbf{erfc}\left(\frac{s_{\text{obs}}}{\sqrt{2}}\right)$$

where **erfc** is the complementary error function and s_{obs} is the observed test statistic.

2. Frequency within a Block Test

The purpose of this test to determine whether the frequency of ones in an M-bit block is approximately $M/2$ as would be expected under an assumption of randomness. [58] For block size $M = 1$, the test degenerates to the Frequency (Monobit) test. The reference distribution for the test statistic is a chi-square distribution. The $\mathcal{P}$-*value* is computed as follow:

$$\mathcal{P}\text{-}value = \mathbf{igamc}\left(\frac{N}{2}, \frac{\chi^2(obs)}{2}\right)$$

where **igamc** is the incomplete gamma function, N is the number of M-bit blocks and $\chi^2(obs)$ is the observed chi-square statistic. The latter is a measure of how well the observed proportion of ones within a given M-bit block match the expected proportion ($1/2$).

[1] The description of each test is taken from the "***NIST Special Publication 800-22rev1a***" from the internet website: *http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html*.

The parameter is computed as follow:

$$\chi^2(obs) = 4M \sum_{i=1}^{N} (\pi_i - 1/2)^2$$

where π_i is the proportion of ones in the ith block retrieved from the formula

$$\pi_i = \frac{1}{M} \sum_{j=1}^{M} \varepsilon_{(i-1)M+j} \qquad \text{for } 1 \leq i \leq N$$

where ε_k is the kth element of the bit string.

3. Runs Test

The focus of this test is the total number of runs in the sequence, where a run is an uninterrupted sequence of identical bits. A run of length k consists of exactly k identical bits, bounded before and after with a bit of the opposite value. The purpose of the runs test is to determine whether the number of runs of ones and zeros of various lengths is as expected for a random sequence. [59] In particular, this test determines whether the oscillation between such zeros and ones is too fast or too slow. This test also performs the $\mathcal{P}$-*value* using the **erfc** function:

$$\mathcal{P}\text{-}value = \mathbf{erfc}\left(\frac{|V_n(obs) - 2n\pi(1-\pi)|}{2\sqrt{2n}\pi(1-\pi)}\right)$$

where π is the proportion of ones in the entire bit string and $V_n(obs)$ is the observed test statistic

$$V_n(obs) \sum_{k=1}^{n-1} r(k) + 1$$

where $r(k) = 0$ if $\varepsilon_k = \varepsilon_{k+1}$ and $r(k) = 1$ otherwise.

4. Longest Run of Ones in a Block Test

The purpose of this test is to determine whether the length of the longest run of ones within the tested sequence is consistent with the length of the longest run of ones that would be expected in a random sequence. An irregularity in the expected length of the longest run of ones implies that there is also a deviation in the expected length of the longest run of zeroes. For this reason, only a test for ones is necessary. [60] The reference distribution for the test statistics is a chi-square distribution, where the test statistic value itself is computed as follow:

$$\chi^2(obs) = \sum_{i=0}^{K} \frac{(\nu_i - N\pi_i)^2}{N\pi_i}$$

where the program provides the values for pi_i as for the values of the parameters K and N. ν_i is the frequency of the longest runs of ones in the ith block retrieved according to a particular algorithm implemented in the suite. As for some of the previous tests, the function **igamc** retrieves the $\mathcal{P}$-*value*:

$$\mathcal{P}\text{-}value = \textbf{igamc}\left(\frac{K}{2}, \frac{\chi^2(obs)}{2}\right)$$

5. Binary Matrix Rank Test

The focus of the test is the rank of disjoint sub-matrices of the entire sequence. The purpose of this test is to check for linear dependence among fixed length substrings of the original sequence. This test also appears in the DIEHARD battery of tests [61]. The original bit sequence is divided into $M \cdot Q$-bit disjoint blocks and subsequently computed their binary ranks, R_l with $l = 1, \ldots, N$. Defining F_m, F_{M-1} and $N - F_M - F_{M-1}$ respectively as the number of matrices with rank $R_l = M$, $R_l = M - 1$ and $R_l =$ number of matrices remaining, the test statistic becomes:

$$\chi^2(obs) = \frac{(F_M - 0.288N)^2}{0.288N} + \frac{(F_{M-1} - 0.5776N)^2}{0.5776N} + \frac{(N - F_M - F_{M-1} - 0.1336N)^2}{0.1336N}$$

The corresponding $\mathcal{P}$-*value* then is:

$$\mathcal{P}\text{-}value = \textbf{igamc}\left(1, \frac{\chi^2(obs)}{2}\right)$$

6. Discrete Fourier Transform (Spectral) Test

The focus of this test is the peak heights in the Discrete Fourier Transform of the sequence. The purpose of this test is to detect periodic features (i.e., repetitive patterns that are near each other) in the tested sequence that would indicate a deviation from the assumption of randomness. The intention is to detect whether the number of peaks exceeding the 95% threshold is significantly different than 5%. [62] Applying a Discrete Fourier Transform (DFT) on the bit string produce a sequence of complex variables that represents periodic components found in the sample data. The test then uses a modulo function on the first $n/2$ elements producing a sequence of peak heights. These values are confronted with the 95% peak height threshold value T defined as

$$T = \sqrt{\log\left(\frac{1}{0.05}\right) n}$$

The test statistic d which is the normalized difference between the observed and the expected number of frequency components that are beyond the 95% threshold follow a normal distribution and is defined as follow:

$$d = \frac{\sqrt{4}(N_1 - N_0)}{\sqrt{n \cdot 0.95 \cdot 0.05}}$$

The corresponding $\mathcal{P}$-*value* is:

$$\mathcal{P}\text{-}value = \mathbf{erfc}\left(\frac{|d|}{\sqrt{2}}\right)$$

7. Non Overlapping Template Matching Test

The purpose of this test is to detect generators that produce too many occurrences of a given non-periodic (aperiodic) pattern. For this test and the Overlapping Template Matching Test, an m-bit window is used to search for a specific m-bit pattern. If the pattern is not found, the window slides one bit position. If the pattern is indeed found, the window is reset to the bit after the observed pattern, and the search resumes. [63] The test subdivides the bit string in N smaller string of length M and then determines the values W_j as the number of times that a template B occurs within the jth block. The template B is an m-bit string of zeros and ones that is defined in a template library of non-periodic patterns contained within the suite. Under an assumption of randomness, the theoretical mean μ and variance σ^2 are:

$$\mu = \frac{M - m + 1}{2^m} \qquad \sigma^2 = M\left(\frac{1}{2^m} - \frac{2m - 1}{2^{2m}}\right)$$

The test statistic follows a chi-square distribution

$$\chi^2(obs) = \sum_{j=1}^{N} \frac{(W_j - \mu)^2}{\sigma^2}$$

and, as the previous tests, gives a $\mathcal{P}$-*value*:

$$\mathcal{P}\text{-}value = \mathbf{igamc}\left(\frac{N}{2}, \frac{\chi^2(obs)}{2}\right)$$

It is important to stress out that the test computes a $\mathcal{P}$-*value* for each template providing up to 148 and 284 $\mathcal{P}$-*values* for $m = 9$ and $m = 10$ respectively.[2]

[2]The test code has been written to provide templates with $m = 10$ at maximum.

8. Overlapping Template Matching Test

The focus of the Overlapping Template Matching Test is the number of occurrences of pre-specified target strings. As for the Non Overlapping Template Matching Test, this test also uses an m-bit window to search for a particular pattern. [64] If the pattern is not found, the window slides one bit position. The difference between this test and the previous one is that, when the pattern is found, the window slides only one bit before resuming the search. Analogously to the Non Overlapping Test, this one computes two parameters, λ and η, later used to determine the theoretical probabilities π_i according to a predefined algorithm within the suite. These parameters are defined as follow:

$$\lambda = \frac{(M - m + 1)}{2^m} \qquad \eta = \frac{\lambda}{2}$$

Even in this case, the test statistic follow a chi-square distribution

$$\chi^2(obs) = \sum_{i=0}^{5} \frac{(\nu_i - N\pi_i)^2}{N\pi_i}$$

where

$$\begin{aligned} \pi_0 &= 0.364091 \\ \pi_1 &= 0.185659 \\ \pi_2 &= 0.139381 \\ \pi_3 &= 0.100571 \\ \pi_4 &= 0.070432 \\ \pi_5 &= 0.139865 \end{aligned}$$

The $\mathcal{P}$-*value* is then retrieved:

$$\mathcal{P}\text{-}value = \mathbf{igamc}\left(\frac{5}{2}, \frac{\chi^2(obs)}{2}\right)$$

9. Maurer's "Universal Statistical" Test

The purpose of the test is to detect whether or not a bit sequence may be significantly compressed without loss of information. An evident compressible sequence is considered to be non-random. [50, 65, 66] The test partitions the n-bit sample in two subsets: an initialization segment consisting of $Q \cdot L$-bit non overlapping blocks and a test segment consisting of $K \cdot L$-bit non overlapping blocks. The remaining bits at the end that do not form a complete L-bit block are discarded. The recommended values for L lies in the region $6 \leq L \leq 16$. Through the usage of a sophisticated algorithm, the test creates

a table, T, of the decimal representation of the contents of each L-bit block and then computes the following test statistic:

$$f_n = \frac{1}{K} \sum_{i=Q+1}^{Q+K} \log_2(i - T_j)$$

where T_j is the jth entry of the table T. f_n is the sum of the $\log_2$ distances between matching L-bit templates, i.e. the sum of the number of digits in the distance between L-bit templates. As for the Frequency Test, the test statistic follows a half normal distribution. The $\mathcal{P}$-*value* then is retrieved using the **erfc** function:

$$\mathcal{P}\text{-}value = \mathbf{erfc}\left(\left|\frac{f_n - \mu(L)}{\sqrt{\vartheta(L)}}\right|\right)$$

where $\mu(L)$ and $\sigma(L)$ take values from a table of precomputed values.[3] Under an assumption of randomness, the sample mean, $\mu(L)$, is the theoretical expected value of the computed statistic for the given L-bit length. The theoretical standard deviation is given by

$$\sigma = c\sqrt{\frac{\vartheta(L)}{K}}$$

where

$$c = 0.7 - \frac{0.8}{L} + \left(4 + \frac{32}{L}\right)\frac{K^{-3L}}{15}$$

10. Linear Complexity Test

The purpose of this test is to determine whether or not a sequence is complex enough to be considered random. Random sequences are characterized by long LFSRs (**L**inear **F**eedback **S**hift **R**egister). An LFSR that is too short implies non-randomness. The test splits the n-bit sequence in N independent blocks of M bits. It then uses the Berlekamp-Massey algorithm determining the linear complexity L_i of each of the N blocks. [67] L_i is the length of the shortest linear feedback shift register sequence that generates all bits in the block i. Within any L_i-bit sequence, some combination of the bits, when added together modulo 2, produces the next bit in the sequence (bit $L_i + 1$). Under an assumption of randomness, the theoretical mean μ is:

$$\mu = \frac{M}{2} + \frac{\left(9 + (-1)^{M+1}\right)}{36} - \frac{1}{9}\frac{(3M+2)}{2^M}$$

[3]The values are taken from the "*Handbook of Applied Cryptography*".

In the next step, the test compute the value T_i:

$$T_i = (-1)^M (L_i - \mu) + \frac{2}{9}$$

According to a specific criterion, if the ith value of T is between a set of predefined boundaries, the system increments a counter ν_j with $j = 1, \ldots, 6$. The test statistic follows a chi-square distribution:

$$\chi^2(obs) = \sum_{i=0}^{K} \frac{(\nu_i - N\pi_i)^2}{N\pi_i}$$

where the π_i, as for those of the Overlapping Template Matching Test, are extracted according to a predefined function and K is the number of degrees of freedom.[4] The $\mathcal{P}$-*value* is then computed:

$$\mathcal{P}\text{-}value = \mathbf{igamc}\left(\frac{K}{2}, \frac{\chi^2(obs)}{2}\right)$$

11. Serial Test

The purpose of this test is to determine whether the number of occurrences of the 2^m m-bit overlapping patterns is approximately the same as expected for a random sequence. Random sequences have uniformity; that is, every m-bit pattern has the same chance of appearing as every other m-bit pattern. [68] For $m = 1$, the Serial Test is equivalent to the Frequency Test. The test first extend the original bit sequence by appending the first m-1 bits to its end, where m is the length in bits of each block. It then determines the frequencies, $\nu_{i_1} \ldots \nu_{i_m}$, of all possible overlapping m-bit blocks of the m-bit patterns $i_1 \ldots i_m$. The same procedure is performed to the (m-1) and (m-2)-bit pattern. The next step is to compute the respective ψ^2_m, ψ^2_{m-1} and ψ^2_{m-2} functions:

$$\begin{aligned}
\psi^2_m &= \frac{2^m}{n} \sum_{i_1 \ldots i_m} \nu^2_{i_1 \ldots i_m} - n \\
\psi^2_{m-1} &= \frac{2^{m-1}}{n} \sum_{i_1 \ldots i_{m-1}} \nu^2_{i_1 \ldots i_{m-1}} - n \\
\psi^2_{m-2} &= \frac{2^{m-2}}{n} \sum_{i_1 \ldots i_{m-2}} \nu^2_{i_1 \ldots i_{m-2}} - n
\end{aligned}$$

[4] The value $K = 6$ has been hard coded into the test.

The test statistics, which follow a chi-square distribution, are then computed as:

$$\begin{aligned} \nabla\psi_m^2 &= \psi_m^2 - \psi_{m-1}^2 \\ \nabla^2\psi_m^2 &= \psi_m^2 - 2\psi_{m-1}^2 + \psi_{m-2}^2 \end{aligned}$$

and the corresponding $\mathcal{P}$-*values*:

$$\begin{aligned} \mathcal{P}\text{-}value_1 &= \mathbf{igamc}\left(2^{m-2}, \nabla\psi_m^2\right) \\ \mathcal{P}\text{-}value_2 &= \mathbf{igamc}\left(2^{m-3}, \nabla^2\psi_m^2\right) \end{aligned}$$

12. Approximate Entropy Test

The purpose of the test is to compare the frequency of overlapping blocks of two consecutive/adjacent lengths (m and m+1) against the expected result for a random sequence. [69] The test performs the same first steps of the algorithm of the Serial Test to a bit string of length n. The system then counts all the possible m-bit, (m+1)-bit, values and represents them as C_i^m, where i is the m-bit value:

$$C_i^m = \frac{\#\, i}{n}$$

The test then computes the functions $\phi^{(m)}$ and $\phi^{(m+1)}$ to extract the test statistic:

$$\begin{aligned} \phi^{(m)} &= \sum_{i=0}^{2^m-1} C_i^{(m)} \log_2 C_i^{(m)} \\ \phi^{(m+1)} &= \sum_{i=0}^{2^{m+1}-1} C_i^{(m+1)} \log_2 C_i^{(m+1)} \\ \chi^2(obs) &= 2n\left[log_2 - ApEn(m)\right] \end{aligned}$$

where

$$ApEn(m) = \phi^{(m)} - \phi^{(m+1)}$$

The $\mathcal{P}$-*value* thus is:

$$\mathcal{P}\text{-}value = \mathbf{igamc}\left(2^{m-1}, \frac{\chi^2(obs)}{2}\right)$$

13. Cumulative Sums (Cusum) Test

The purpose of the test is to determine whether the cumulative sum of the partial sequences occurring in the tested sequence is too large or too small relative to the expected behavior of that cumulative sum for random sequences. The cumulative sum may be considered as a random walk. [60] For a random sequence, the excursions of the random walk should be near zero. For certain types of non-random sequences, the excursions of this random walk from zero will be large. The test converts the zeros and ones of the input sequence into -1 and +1 respectively. The system then computes partial sums S_i of successively larger subsequences starting from the first value (*mode 0*) or last value (*mode 1*).

$$\begin{cases} S_k &= S_{k-1} + X_k & \textit{for mode 0} \\ S_k &= S_{k-1} + X_{n-k+1} & \textit{for mode 1} \end{cases}$$

where X_k identifies the kth element of the converted bit string. The test then computes the test statistic as

$$z = \max_{1 \le k \le n} |S_k|$$

where $\max_{1 \le k \le n} |S_k|$ is the largest of the absolute values of the partial sums S_k. This variable follows a normal distribution and the corresponding $\mathcal{P}$-*value* is:

$$\begin{aligned} \mathcal{P}\textit{-value} \quad = \quad 1 \quad - \quad & \sum_{k=4\left(\frac{-n+z}{z}\right)}^{4\left(\frac{n-z}{z}\right)} \left[\Phi\left(\frac{(4k+1)z}{\sqrt{n}} \right) - \Phi\left(\frac{(4k-1)z}{\sqrt{n}} \right) \right] + \\ & \sum_{k=4\left(\frac{-n-3z}{z}\right)}^{4\left(\frac{n-z}{z}\right)} \left[\Phi\left(\frac{(4k+3)z}{\sqrt{n}} \right) - \Phi\left(\frac{(4k+1)z}{\sqrt{n}} \right) \right] \end{aligned}$$

where Φ is the Standard Normal Cumulative Probability Distribution Function.

14. Random Excursions Test

The focus of this test is the number of cycles having exactly K visits in a cumulative sum random walk. The cumulative sum random walk derives from partial sums after the $(0, 1)$ sequence is transferred to the appropriate $(-1, +1)$ sequence. A cycle of a random walk consists of a sequence of steps of unit length taken at random that begins at and return to the origin. The purpose of this test is to determine if the number of visits to a particular state within a cycle deviates from what one would expect for a random sequence. [70] This test is actually a series of eight tests (and conclusions), one test and conclusion for each of the states: $(x = -4, -3, -2, -1)$ and $(x = +1, +2, +3, +4)$. The test computes the partial sums S_i of successively larger sequences of the n-bit string and

uses their values to initialize a new one, $S^{'}$, adding a zero before and after the set of values. The test then computes a test statistic for each value of x:

$$\chi^2(obs) = \sum_{k=0}^{5} \frac{(\nu_k(x) - J\pi_k(x))^2}{J\pi_k(x)}$$

where $\pi_k(x)$ is the probability that the state x occurs k times in a random distribution calculated according to predefined functions. $\nu_k(x)$ is the total number of cycles in which state x occurs exactly k times among all cycles, for $k = 0, 1, \ldots, 5$ ($k = 5$ counts also all frequencies ≥ 5). J is the total number of zeros crossings in the new string $S^{'}$ and it is also the number of cycles in $S^{'}$, where a cycle is a subsequence consisting of an occurrence of zero, followed by no-zero values, and ending with another zero. For each x the $\mathcal{P}$-*value* is:

$$\mathcal{P}\text{-}value = \mathbf{igamc}\left(\frac{5}{2}, \frac{\chi^2(obs)}{2}\right)$$

15. Random Excursions Variant Test

The purpose of this test is to detect deviations from the expected number of visits to various states in the random walk. [71] This test is actually a series of eighteen tests (and conclusions), one for each of the states: $-9, -8, \ldots, -1$ and $+1, +2, \ldots, +9$. The test proceeds in the same way of the Cumulative Sums and then, after obtaining the string $S^{'}$, computes the test statistic $\xi(x)$ for each of the eighteen non-zero states of x. $\xi(x)$ is the total number of times that the state x occurred across all J cycles. Because the test statistics follows an half normal distribution, the xth $\mathcal{P}$-*value* is computed as:

$$\mathcal{P}\text{-}value = \mathbf{erfc}\left(\frac{|\xi(x) - J|}{\sqrt{2J(|x| - 2)}}\right)$$

The NIST suite is, of course, not the only existing testing suite. In the world of cryptographic suite there exists three more "standard" suites: the AIS31, the DIEHARD and the TESTU01. The AIS31 is a cryptographic suite developed by the German **B**undesamt für **S**icherheit in der **I**nformationstechnik (BSI) (Federal Office for Security in Information Technology) on the theoretical work made by W. Schindler in order to define a reliable method for the evaluation of TRNGs. [72, 73] The tests described by Schindler are divided into two classes: the first class checks that the random bits do not present

conspicuous statistical traits while the second class verifies that they are *practically impossible to determine even if the predecessors or successors are known.* The DIEHARD battery includes tests developed to test PRNGs and represent a standard due to the difficulty generators have of passing them. The DIEHARD suite is also used for TRNGs [74, 75, 76], but that could be attributed more to its fame rather than the suitability of the tests to reveal weakness introduced by hardware problems. The TESTU01 is a library of tests developed by L'Ecuyer [77, 78]. The suite is the most recent and comprises the largest spectrum of tests presently available. As for the DIEHARD suite, most of its tests are oriented for the analysis of PRNGs, but the suite provides a specific battery, the *Alphabit battery*, designed primarily to test hardware random bits generators. This battery directly analyzes bits and applies tests sensitive to the typical problems of bit uniformity and independence.

11 Tests results

Test statistics provide information about the degree of randomness of a number. Normally, the $\mathcal{P}$-*values* associated to these values, are analyze in order to asses, with sufficient confidence, if the bit string *may* be considered random. Two specific studies permit to decide if the test statistics really satisfy the *null-hypothesis*: uniformity of $\mathcal{P}$-*values* and threshold limit for the sub-strings proportion. The first section shows the results obtained from strings generated by classical extractors while the second section shows the results from the quantum extractor.

11.1 Classical testing results

The NIST suite, after the application of the tests, prints a detailed report, for every test, of the distribution of $\mathcal{P}$-*values* calculated for every sub-string, the $\mathcal{P}$-*value* on the uniform distribution of the $\mathcal{P}$-*values* and the proportion of bit strings that passed the test. A test is failed if: the proportion of sub-strings is below a given threshold calculated on the basis of the number of strings analyzed or if the distribution of $\mathcal{P}$-*values* is not uniform. Starting with the first criterion, it is simpler to understand the process considering the following example. Suppose that the bit string analyzed is subdivided in $m = 1000$ sub-strings. The significance level α gives the number of strings expected to fail the test (for $\alpha = 0.01$ it means that 1 string, out of 100, is expected to fail the test). Out of all the m sub-strings, suppose that 996 passed the test, that is $\mathcal{P}$-*values* $\geq \alpha$, then the proportion is $^{996}/_{1000} = 0.996$. The upper and lower threshold, that specify the range of acceptable proportions, are determined as:

$$\hat{p} \pm 3\sqrt{\frac{\hat{p}(1-\hat{p})}{m}} \tag{11.1}$$

where $\hat{p} = 1 - \alpha$ and m is the sample size.

These levels arise from the normal approximation of the binomial distribution. If the sample size m is such that $n\,(1-\alpha) > 5$ and $\alpha\, m > 5$, then the expected number of strings with a $\mathcal{P}$-*value* larger than α is $m\,(1-\alpha)$ with standard deviation $\sigma = \sqrt{\alpha\,(1-\alpha)\,m}$. [79, 80] The expected proportion then may be lowered by the right hand side of Eq. (11.1). Fig. (11.1) shows the plot of proportion relative to the tests performed on a bit string extracted from raw data. Fig. (11.3) represents the result of the analysis made on a 100Mb file obtained by applying the Elias Extractor while Fig. (11.4) display the results of a 100Mb file obtained by applying the AMLS Extractor. The red lines represents the upper and lower threshold computed following Eq. (11.1). Looking at the plot of the

raw data, it is clear that the system is heavily biased. Because of the implementation of inefficient devices, the system produce more 0s than 1s, offsetting the entire bit string.

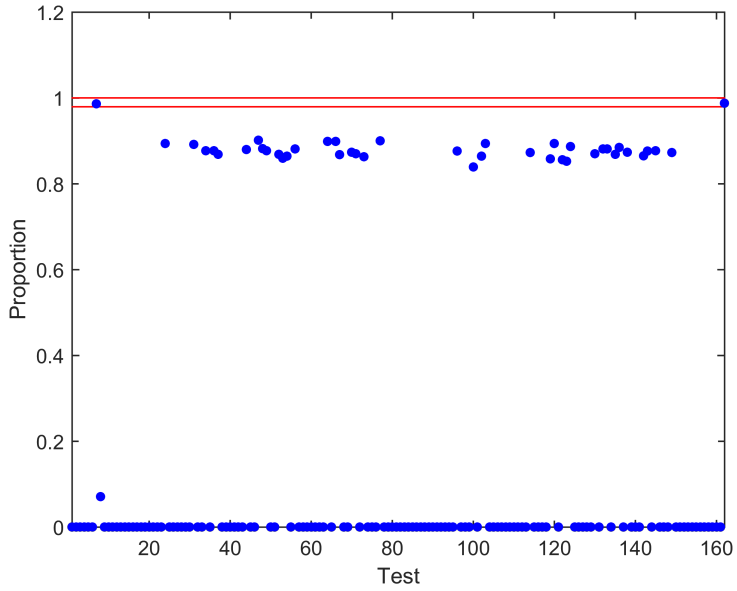


Figure 11.1: Proportion plot relative to a 20Mb file of raw data directly extracted from the acquisition frames. Only the Rank and Linear Complexity tests passed.

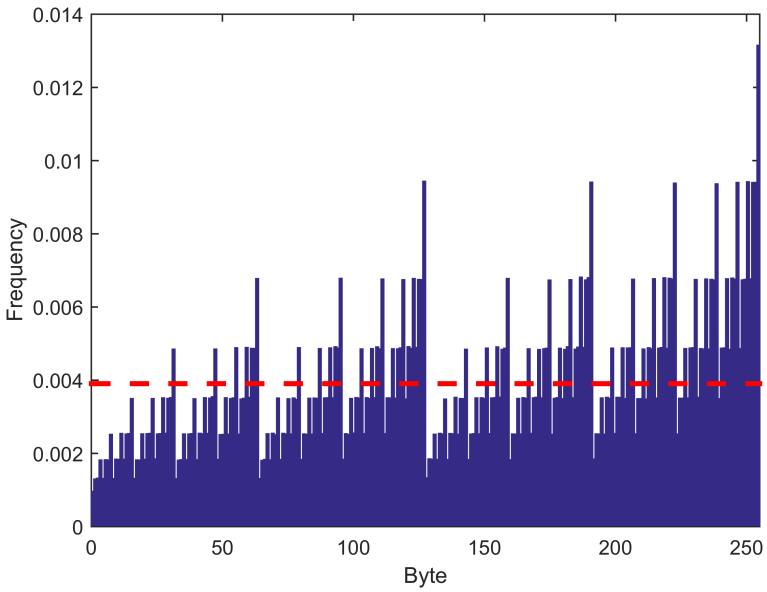


Figure 11.2: Byte distribution for the raw data file showing an evident repetitive pattern. The dashed red line represents the constant value of $1/256$, corresponding to a perfectly uniform distribution.

A second way to assess the impossibility of using numbers extracted from this type of bit sequences is to examine the bytes distribution. If the bit string is indeed random, the possible byte outputs (8-bit sequence corresponding to unsigned integer values $\{0, \ldots, 255\}$) are uniformly distributed. Any evident deviation from the expected behavior is an indicator of no randomness. Fig. (11.2) shows the bytes distribution for the raw data file. As shown in the graphic, the distribution is far from uniform. There is a clear repetitive pattern symptom of more frequent identical bytes in the bit sequence. This is an obvious example of why it is necessary to implement a stage of post processing in a TRNG. Figs. (11.3) and (11.4) instead, show that the tests passing the first condition for asserting that the bit string is truly random are practically all of them.

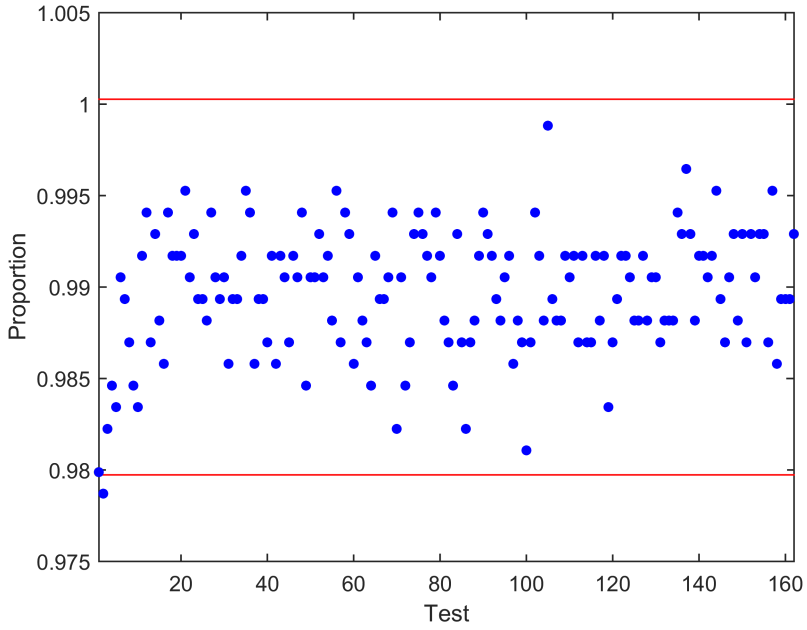
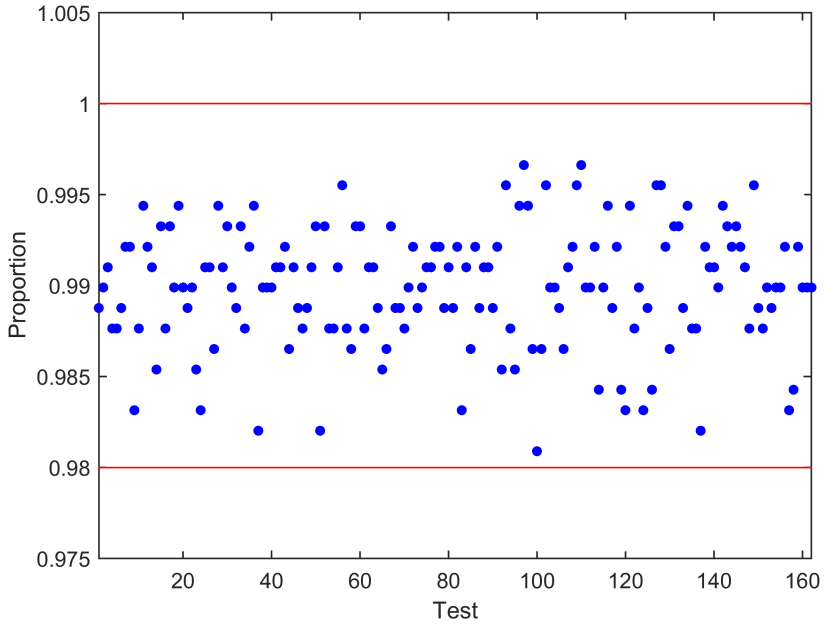
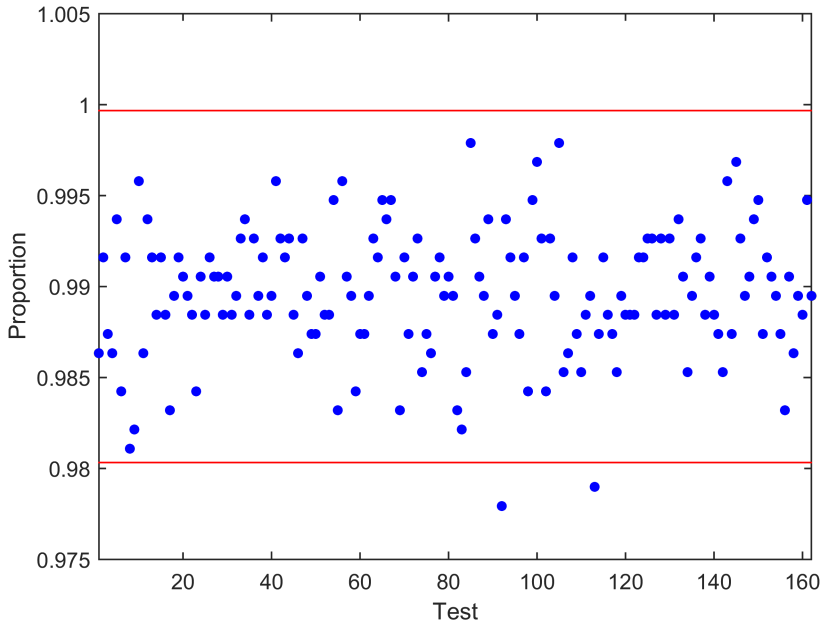


Figure 11.3: Proportion plot relative to a 100Mb file obtained by applying the Elias Extractor to a LASER acquisition at working power of $1.24\,\mu W$. Only one point exceeds the lower threshold, but to ascertain that the test did not pass, it is necessary to examine the relative $\mathcal{P}$-*values* distribution.

Indeed, there are very few tests (two for the *temporal* AMLS and only one for the Elias) that did not meet the passing condition, but it is important to stress out that its the second criterion that more strongly rules the passing of a test. In fact, if the $\mathcal{P}$-*values* distribution is not uniform, then this is a strong indicator that the analyzed bit string does not present randomness properties even if the relative proportion is inside the acceptance range (red lines). [81, 82] This way, it is necessary to study the $\mathcal{P}$-*values* distribution for each test. A simple solution to how illustrate the analysis is to use an histogram plot. The interval [0,1], that is the range of possible $\mathcal{P}$-*values*, is divided into 10 sub-intervals, and the $\mathcal{P}$-*values* that lie within each sub-interval are counted and displayed.



(a) Results relative to a file extracted via the spatial method for a LED acquisition of $1.24\,\mu W$



(b) Results relative to a file extracted via the temporal method for a LASER acquisition at working power of $7.46\,\mu W$

Figure 11.4: Proportion plots relative to a 100Mb files obtained by applying the AMLS Extractor to a bit string retrieved via the spatial and temporal method for a LED acquisition at working power of $1.24\,\mu W$ and a LASER acquisition at working power of $7.46\,\mu W$. There are only two points outside the lower threshold, but to ascertain that both tests did not pass, it is necessary to examine the relative $\mathcal{P}$-*values* distribution.

Naturally there is the need to mathematically, and not just figuratively, assert that the distribution is indeed uniform. To do that a chi-square test is made, extracting $\mathcal{P}$-$value_T$, corresponding to the *Goodness of Fit Distributional Test* on the $\mathcal{P}$-*values* obtained for an arbitrary statistical test, i.e. a $\mathcal{P}$-*value* of the $\mathcal{P}$-*values*. [83, 84]

This is accomplished by computing:

$$\chi^2 = \sum_{i=1}^{10} \frac{(F_i - S/10)^2}{S/10}$$

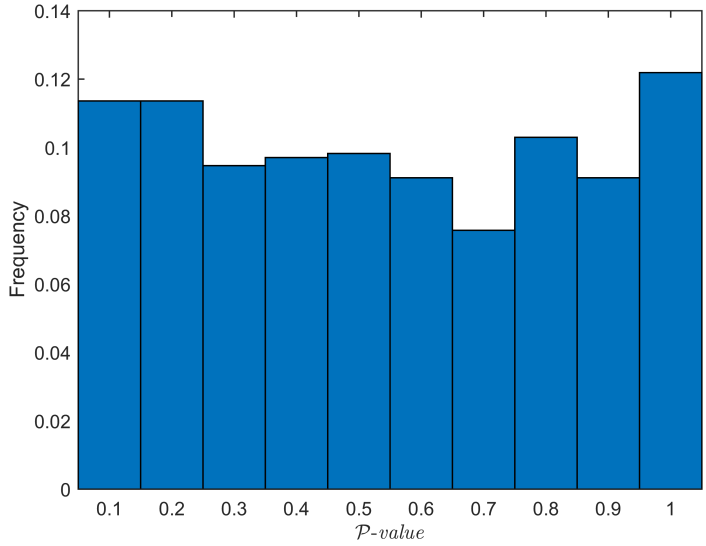
where F_i is the number of $\mathcal{P}$-*values* in sub-interval i and S is the sample size. The $\mathcal{P}$-$value_T$ is then calculated as:

$$\mathcal{P}\text{-}value_T = \mathbf{igamc}\left(\frac{9}{2}, \frac{\chi^2}{2}\right)$$

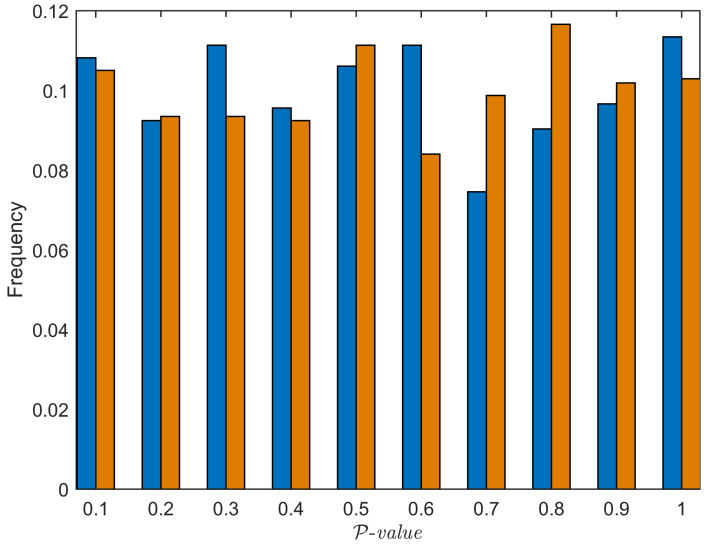
If $\mathcal{P}$-$value_T \geq 10^{-4}$, then the sequence may be considered to be uniformly distributed. Additionally, to provide statistically meaningful results at least 55 sequences must be processed. [49] To verify that, the three test that in Figs. (11.3) and (11.4) did not pass the first criterion are in reality just statistical fluctuations, it has been computed the respective $\mathcal{P}$-$value_T$ of their $\mathcal{P}$-*values*. Fig. (11.5) shows the histograms of the $\mathcal{P}$-*values* distributions while Tab. (11.1) shows the $\mathcal{P}$-$values_T$. Looking at the table's entries, all of them satisfy the condition $\mathcal{P}$-$value_T \geq 10^{-4}$. This proves that the reason why the three tests did not pass the proportion criterion was because of statistical fluctuations. This is perfectly normal and common in statistical testing because of its statistic nature.

Files	$\mathcal{P}$-$value_T$
"Elias"	0.122
"AMLS"	0.159
	0.532

Table 11.1: $\mathcal{P}$-$values_T$ of the three data proportions that do not satisfy the passing condition of Eq. (11.1). The entries for the "AMLS" file are relative to the temporal method extraction. All of them satisfy the condition $\mathcal{P}$-$value_T \geq 10^{-4}$ confirming that the reason for not passing the proportion test was due to statistical fluctuations.

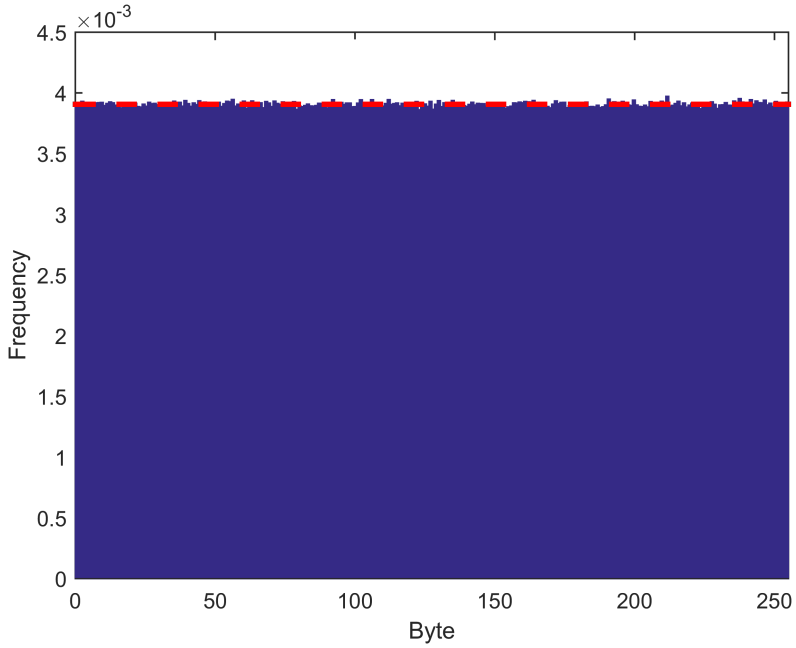


(a) $\mathcal{P}$-*values* distribution for the "Elias" file relative to the test that did not pass the proportion criterion

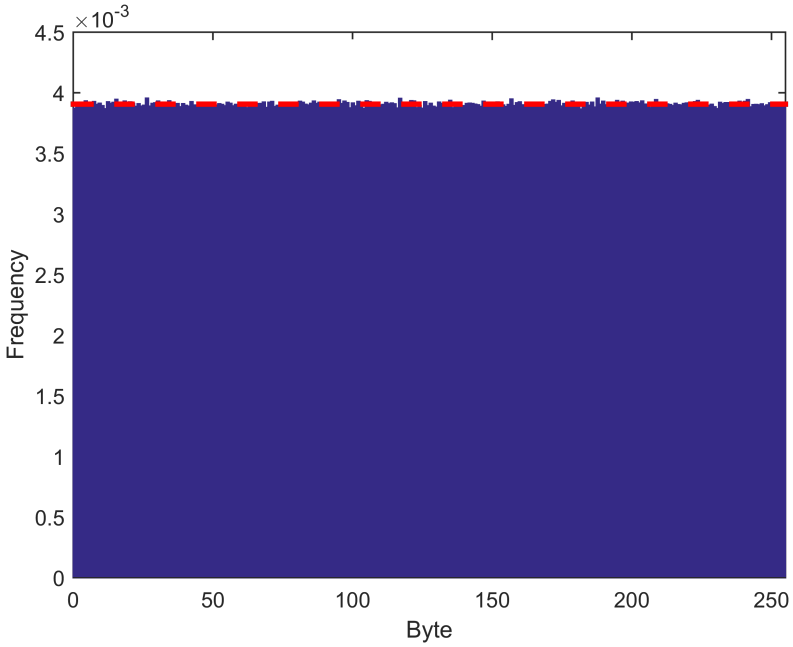


(b) $\mathcal{P}$-*values* distributions for the "AMLS" file relative to the tests that did not pass the proportion criterion

Figure 11.5: $\mathcal{P}$-*values* distributions for all three tests that did not pass the proportion criterion. The distributions do not provide a direct visual proof if they are uniform or not. However, looking at Tab (11.1) entries, the chi-square analysis verifies that they are indeed uniform and then confirming that the system presents statistical fluctuations.



(a) Byte distribution for the "Elias" file



(b) Byte distribution for the "AMLS" file

Figure 11.6: Bytes distributions relative to the "Elias" and "AMLS" files. The frequency with which every byte is present in the original bit sequence is normalized so that the expected value is $f_M = 0.0039 = 1/256$. The red dashed lines represent this value.

As previously stated, normal hardware and software implementation require the usage of random numbers instead of bit strings. Performing a byte (8-bit long sequence) distribution on a given bit string permits to assert if the sequence is indeed random. Contrarily to the byte distribution shown in Fig. (11.2), this time it is expected that all the possible values $\{0, \ldots, 255\}$ that a byte may produce, are uniformly distributed. Fig. (11.6) shows the bytes distributions for the "Elias" and "AMLS" files. Looking at these histograms it is clear that the distributions are perfectly uniform. The normalized frequency at which every byte is present in the bit string is exactly $f_M = 0.0039 = 1/256$ represented with a red dashed line in the plots.

11.2 Quantum testing results

The application of the quantum extractor yields bit strings that are really small compared to the raw data. This is because the value of the $H_{min}(x|E)$ for a given number of detectors is always less than the number of extractable bit sequence. For example, consider the case of three detectors and a value of λ that maximize the Conditioned min-Entropy for a quantum efficiency of $\eta = 5.47\%$. According to Fig. (8.13) the corresponding value is $H_{min} = 0.0690$. This means that, selecting only 0.0690 bits out of a 4-bit sequence, provides a truly random sequence conditioned on all side information. Because of this low "bit-rate" extraction, the implemented program used to obtained the sequences glues together more strings in order to retrieve a sufficiently longer one to test. Fig. (11.7) shows the tests proportions for a 100Mb file. This time the proportions are distributed more closely to the expected value $\hat{p}$ symptom that more strings passed the statistical testing. Naturally to fully consider passed all the tests, it is necessary to control the $\mathcal{P}$-*values* distribution and check if they are uniformly distributed. Fig. (11.8) shows the distribution. Computing the chi-square test gives $\chi^2 = 0.0577$ that produces $\mathcal{P}\text{-}value_T = 1.00$. Because $\mathcal{P}\text{-}value_T$ is greater than 10^{-4}, the sequence can be considered uniformly distributed. The fact that the $\mathcal{P}\text{-}value_T$ is one is mostly because the programs rounds values up to the hundredth digit and not because of the extremely degree of uniformity. Even if that is the case, this proves that a bit sequence extracted via quantum approach is (almost) perfectly random. Fig. (11.9) shows the bytes distribution relative to the 100Mb file.

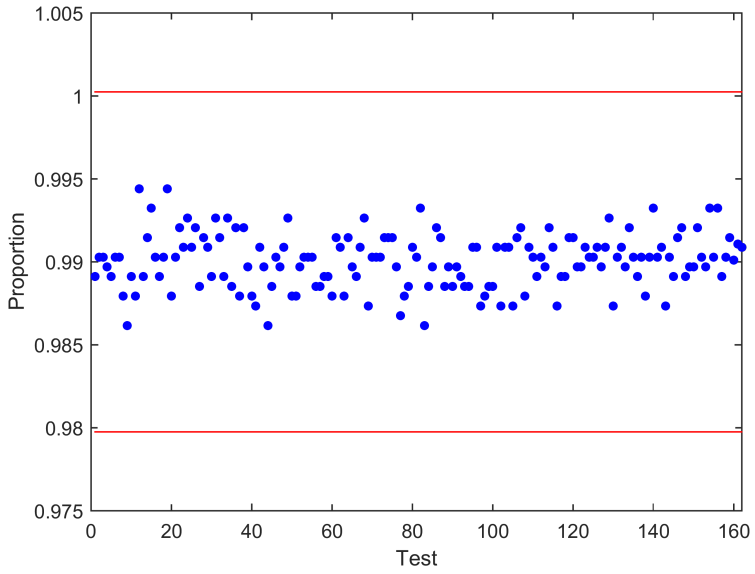


Figure 11.7: Proportion plots relative to a 100Mb files obtained by applying the Quantum Extractor to a bit string retrieved from a LASER acquisition at working power of $1.15\,\mu W$. The points are distributed closely to the expected value $\hat{p} = 0.99$ meaning that more sub-strings passed the tests.

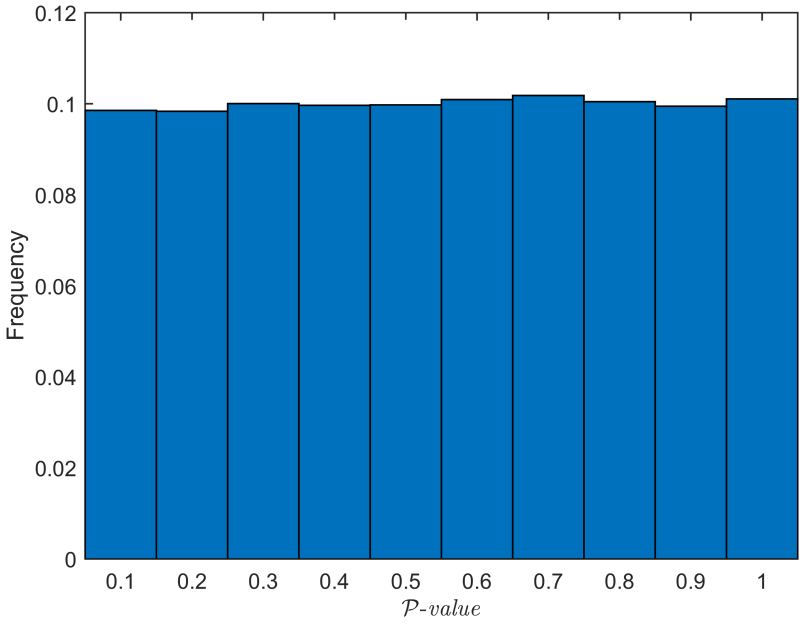


Figure 11.8: $\mathcal{P}$-*values* distribution for all tests applied to the bit sequence extracted via the Quantum Extractor. The χ^2 test on the distribution produces an extreme $\mathcal{P}$-$value_T$: $\mathcal{P}$-$value_T = 1$.

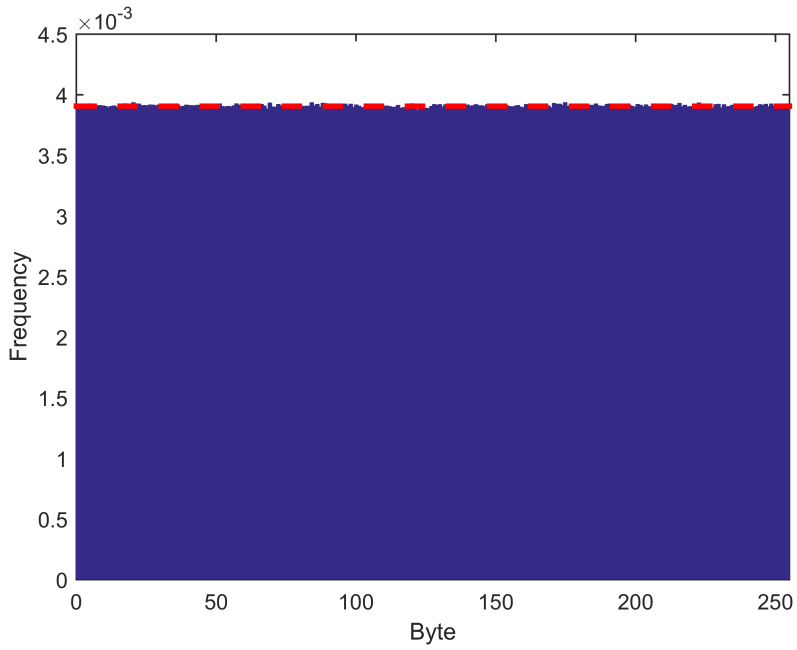


Figure 11.9: Byte distribution relative to a LASER acquisition at working power of $1.15\mu W$ after applying the Quantum Extractor. The frequency with which every byte is present in the original bit sequence is normalized so that the expected value is $f_M = 0.0039 = 1/256$. The red dashed lines represent this value.

Conclusions

The thesis focuses on the generation of truly random numbers from quantum physical processes, QRNGs (**Q**uantum **R**andom **N**umber **G**enerators), in this case the PBS model, through the usage of particular post-processing. A light source (LASER and LED) emits photon that impinge on a camera made of 32x32 SPADs (**S**ingle **P**hoton **A**valanche **D**iodes) with a "single photon" sensitivity. The signal is then digitalized and sent to the computer through a USB cable directly connected to the camera. Different acquisitions were made varying the working power of the light sources and the relative distance between the camera and the sources themselves.

The theoretical model developed by D. Frauchiger, R. Renner and M. Troyer *"True randomness from realistic quantum devices"* [20], permits to apply a sophisticated post-processing capable of extracting *true* random bits even from heavily biased random sources. Instead of concentrating on the actual random number, the model studies the physical process that produces the number.

Exploiting the spatial chaos of photons detection by the camera, the simpler model of [20] was expanded to a general one with an indefinite number of detectors. Generalizing the multinomial distribution (rising from the calculus of all possible outputs for a given set of detectors) allows to perform fractals, binomial expansions, and probabilities analysis. For a more realistic physical system, Dark count, Crosstalk and Afterpulsing effects were taken into account.

The thesis studies in parallel PRNGs (**P**seudo **R**andom **N**umber **G**enerators) using sophisticated algorithms. The main difference between QRNGs and PRNGs is that the first exploit the very probabilistic and aleatory nature of quantum processes while the latter uses mathematical approaches that are fated to produce the same numbers after a specific time. The work focuses on three principal, and most sophisticated, extractors: Von Neumann, Elias and AMLS (**A**dvance **M**ulti**L**evel **S**trategy) (The Von Neumann extractor is not sophisticated, but it is one of the most famous).

Statistical tests permit to verify randomness and statistically ensure that numbers are actually random. Both classical and quantum extracted numbers pass these tests guaranteeing the requests of uniformity and independence.

Cameras made of 32x32 arrays of SPADs are expensive and thus are not always affordable. The model, however, permits to reduce the number of required pixels to the one desired and still produce truly random numbers. This way it is possible to make use of less refined and complex cameras.

As for all QRNGs that relies on "PBS models", even this one does not permit to obtain a fast generation rate (in this case 48 Mbit/s). Nonetheless, the post-processing ensures that the random numbers are *truly random*.

A Coherent states

The most commonly "*single mode*" state does not correspond to an individual "*number state*" $|n\rangle$, but to a linear superposition of them. There exist many different kinds of superposition states, but the most important and of practical use is the "*coherent state*". This state plays a significant role in the quantum theory of light. The coherent states take the Greek letter $|\alpha\rangle$. The corresponding electric field variation approximates that of the classical wave of stable amplitude ad fixed phase in the limit of high excitation. Their importance does not only dwell on the fact that their properties most closely resemble those of a classical electromagnetic wave, but also because a single mode LASER operated above threshold generates a coherent state excitation. In general terms, a coherent state is expressed as following:

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle \tag{A.1}$$

In Eq. (A.1), α is a complex number and the coherent state defined forms a double continuum corresponding to the continuous ranges of values of the real and imaginary part of α. It may be easily verified that the state $|\alpha\rangle$ is normalized:

$$\begin{aligned}
\langle\alpha|\alpha\rangle &= e^{-|\alpha|^2} \sum_{m,n=0}^{\infty} \frac{{\alpha^*}^m \alpha^n}{\sqrt{m!}\sqrt{n!}} \langle m|n\rangle \\
&= e^{-|\alpha|^2} \sum_{m,n=0}^{\infty} \frac{{\alpha^*}^m \alpha^n}{\sqrt{m!}\sqrt{n!}} \delta_{m,n} \\
&= e^{-|\alpha|^2} \sum_{n=0}^{\infty} \frac{|\alpha|^{2n}}{n!} \\
&= e^{-|\alpha|^2} e^{|\alpha|^2} \\
&= 1
\end{aligned} \tag{A.2}$$

An important fact about coherent states is that, differently from the $|n\rangle$ states, they are not necessarily orthogonal:

$$\langle\alpha|\beta\rangle = \exp\left[-\frac{1}{2}|\alpha|^2 - \frac{1}{2}|\beta|^2\right] \sum_{n=0}^{\infty} \frac{{\alpha^*}^n \beta^n}{n!} = \exp\left[-\frac{1}{2}|\alpha|^2 - \frac{1}{2}|\beta|^2 + \alpha^*\beta\right] \tag{A.3}$$

This way, the $|\alpha\rangle$ state in Eq. (A.1) is eigenstate of the destruction operator with eigenvalue α:

$$\begin{aligned}\hat{a}\,|\alpha\rangle &= e^{-|\alpha|^2/2}\sum_{n=1}^{\infty}\frac{\alpha^n}{\sqrt{n!}}\sqrt{n}\,|n-1\rangle \\ &= e^{-|\alpha|^2/2}\sum_{n=1}^{\infty}\frac{\alpha^n}{\sqrt{(n-1)!}}\,|n-1\rangle \\ &= e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}\,|n\rangle \\ &= \alpha\,|\alpha\rangle \end{aligned} \tag{A.4}$$

Using the relations that define the destructive and creation operator

$$\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle \tag{A.5}$$

$$\hat{a}^\dagger\,|n\rangle = \sqrt{n+1}\,|n+1\rangle \tag{A.6}$$

Eq. (A.1) becomes:

$$|\alpha\rangle = e^{-|\alpha|^2/2}\sum_{n=0}^{\infty}\frac{\left(\alpha\hat{a}^\dagger\right)^n}{n!}\,|0\rangle = \exp\left[\alpha\hat{a}^\dagger - \frac{1}{2}|\alpha|^2\right]|0\rangle \tag{A.7}$$

Generally this result is rewritten in a more compact form:

$$|\alpha\rangle = \hat{D}(\alpha)\,|0\rangle \tag{A.8}$$

where the "*coherent state displacement*" operator

$$\hat{D}(\alpha) = \exp\left(\alpha\hat{a}^\dagger - \alpha^*\hat{a}\right) \tag{A.9}$$

is equivalent to a creation operator for the complete state, analogous to the particle number operator $\hat{n}$. The operator satisfies the following relations:

$$\hat{D}^\dagger(\alpha)\hat{D}(\alpha) = \hat{D}(\alpha)\hat{D}^\dagger(\alpha) = 1 \tag{A.10}$$

$$\hat{D}^\dagger(\alpha)\hat{a}\hat{D}(\alpha) = \hat{a} + \alpha \tag{A.11}$$

$$\hat{D}^\dagger(\alpha)\hat{a}^\dagger\hat{D}(\alpha) = \hat{a}^\dagger + \alpha^* \tag{A.12}$$

The coherent state expectation value for the $\hat{n}$ operator is obtained using properties from Eqs. (A.4 - A.6):

$$\langle n\rangle = \langle\alpha|\hat{n}|\alpha\rangle = \left\langle\alpha|\hat{a}^\dagger\hat{a}|\alpha\right\rangle = |\alpha|^2 \tag{A.13}$$

Analogously the second momentum may be easily computed redefining the dependence of $\hat{n}$ on $\hat{a}$ and $\hat{a}^\dagger$ as following:

$$\hat{n}^2 = \hat{a}^\dagger\hat{a}\hat{a}^\dagger\hat{a} = \hat{a}^\dagger\left(\hat{a}^\dagger\hat{a}+1\right)\hat{a} = \hat{a}^\dagger\hat{a}^\dagger\hat{a}\hat{a} + \hat{a}^\dagger\hat{a} \tag{A.14}$$

using Eq. (3.15). The result is:

$$\left\langle n^2\right\rangle = \left\langle\alpha|\hat{n}^2|\alpha\right\rangle = |\alpha|^4 + |\alpha|^2 = \langle n\rangle^2 + \langle n\rangle \tag{A.15}$$

and consequently:

$$(\Delta n)^2 = |\alpha|^2 = \langle n\rangle \tag{A.16}$$

The projective measurement operators are used to calculate the probability of finding n photons in the system:

$$P(n) = |\left\langle n|\alpha\right\rangle|^2 \tag{A.17}$$

obtaining:

$$\begin{aligned} P(n) &= e^{-|\alpha|^2}\frac{|\alpha|^{2n}}{n!} \\ &= e^{-\langle n\rangle}\frac{\langle n\rangle^n}{n!} \end{aligned} \tag{A.18}$$

Eq. (A.18) is the Poisson distribution relative to $\langle n\rangle$ mean value. Fig. (A.1) shows the distribution's shape for different values of $\langle n\rangle$. As the mean value increases, the distribution tends to a Gaussian.

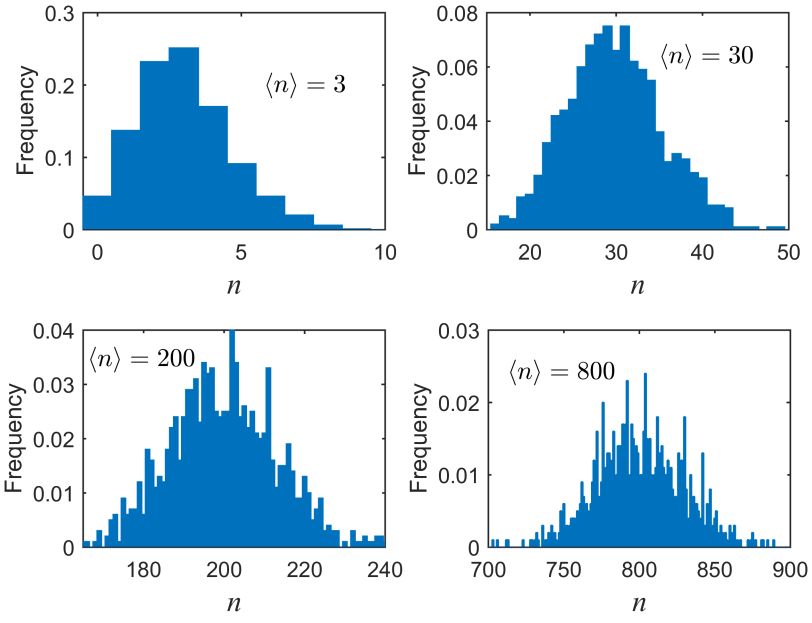


Figure A.1: Poisson distributions for increasing mean values. The distributions tend to a Gaussian as the mean value $\langle n\rangle$ increases.

B Test statistics functions

For each binary sequence, an individual statistical test must produce at least one $\mathcal{P}$-*value*. $\mathcal{P}$-*values* are based on the evaluation of special functions, which must be as accurate as possible on the target platform. The NIST suite provides log files for each statistical test reporting $\mathcal{P}$-*values* with six digits of precision. However, if greater precision is desired, it is possible to modify the code test in each statistical test accordingly. During the testing phase, NIST commonly evaluates sequences on the order 106; hence, results are based on this assumption. If the user wishes to choose longer sequence lengths, then he should be aware that numerical computations may be inaccurate due to machine or algorithmic limitations. Tab. (B.1) and Tab. (B.2) show sample parameter values and corresponding special function values for illustrative purposes. Tab. (B.1) compares the results for the upper incomplete gamma function for selected parameter values of a and x. The definitions for the gamma function, $\Gamma(z)$, and the upper Incomplete Gamma Function, $\mathbf{igamc}\,(a,x)$, are defined, respectively, as:

$$\Gamma(z) = \int_0^\infty t^{(z-1)} e^{-t}\, dt \tag{B.1}$$

$$\mathbf{igamc}\,(a,x) \equiv Q(a,x) = \frac{\Gamma(a,x)}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_x^\infty t^{(a-1)} e^{-t}\, dt \tag{B.2}$$

where $Q(a,0) = 1$ and $\lim_{x\to\infty} Q(a,x) = 0$.

Since the algorithm used in the test suite implementation of the incomplete gamma function is based on the numerical recipe codes, it is evident that the function is accurate to at least the seventh decimal place. For large values of a, the precision will degrade, as will confidence in the result (unless a computer algebra system is employed to ensure high precision computations). Tab. (B.2) compares the results for the Complementary Error Function for selected parameter values of x. The definition for the complementary error function is:

$$\mathbf{erfc}\,(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-u^2}\, du \tag{B.3}$$

To reduce the likelihood for obtaining an inaccurate $\mathcal{P}$-*value* result, NIST has prescribed recommended input parameters.

Parameters	$Q(a, x)$
a=x=600	0.4945710331
a=x=800	0.4952983876
a=x=1000	0.4957947558
a=x=10000	0.4986701917
a=x=100000	0.4995794778
a=x=1000000	0.4998670205

Table B.1: Selected input parameters for the Incomplete Upper Gamma Function $Q(a, x)$ and its values.

Parameter	**erfc** (x)
x=0.00	1.000000000000000
x=0.50	0.479500122186953
x=1.00	0.157299207050285
x=1.50	0.033894853524689
x=2.00	0.004677734981047
x=2.50	0.000406952017445
x=3.00	0.000022090496999
x=3.50	0.000000743098372

Table B.2: Selected input parameters for the Complementary Error Function **erfc** (x) and its values.

C Example of conditioned probabilities for 2 and 3 detectors

Tabs. (C.1) and (C.4) show the conditioned probabilities for $l = 2$ and $l = 3$ respectively and $n \geq 1$. Each row corresponds to a specific Sensitive-Insensitive configuration while each column to a possible output for the given number of detectors l. It is straightforward to see that each row sum up to unity. That is because summing each row is the same as summing all conditioned probabilities for a given configuration over all possible outputs. The relation follow directly from the very definition of conditioned probability:

$$\sum_x P_{X|NT_1T_2\ldots T_l}(x|nt_1t_2\ldots t_l) = 1$$

where $P_{X|NT_1T_2\ldots T_l}(x|nt_1t_2\ldots t_l)$ is the conditioned probability for a generic number of detectors l and incoming photons n. Tabs. (C.2) and (C.3), instead, show the reduced tables according to Chap. (8)

The case when $n = 0$ is simply a table with every entry zero except for the one in the $(2^l - 1)$th row and 1st column which is always 1. In fact, when no photons impinge on the detectors, the probability of obtaining the output $(\underbrace{0, 0, 0, \ldots, 0}_{l})$, when all detectors are insensitive, is exactly one.

		Outputs			
		(0,0)	**(0,1)**	**(1,0)**	**(1,1)**
Configurations	**(S S)**	0	$\left(\frac{1}{2}\right)^n$	$\left(\frac{1}{2}\right)^n$	$1 - 2\left(\frac{1}{2}\right)^n$
	(S I)	$\left(\frac{1}{2}\right)^n$	0	$1 - \left(\frac{1}{2}\right)^n$	0
	(I S)	$\left(\frac{1}{2}\right)^n$	$1 - \left(\frac{1}{2}\right)^n$	0	0
	(I I)	1	0	0	0

Table C.1: Conditioned probabilities for $l = 2$ and $n \geq 1$ for all possible Sensitive-Insensitive configurations and outputs.

		Outputs		
		(0,0)	**(0,1)**	**(1,1)**
Configurations	**(S S)**	0	$\left(\frac{1}{2}\right)^n$	$1-2\left(\frac{1}{2}\right)^n$
	(I S)	$\left(\frac{1}{2}\right)^n$	$1-\left(\frac{1}{2}\right)^n$	0
	(I I)	1	0	0

Table C.2: Conditioned probabilities for $l = 2$ and $n \geq 1$ for reduced configurations and outputs. To retrieve Tab. (C.1) is sufficient to create a new matrix with each row and column repeated exactly $\binom{l}{s}$ and $\binom{l}{u}$ times respectively, where s is the number of sensitive detector for the specific configuration while u is the number of 1s relative to a given output. As illustrated in Chap. (8), the new table has $l + 1$ rows and columns.

		Outputs			
		(0,0,0)	**(0,0,1)**	**(0,1,1)**	**(1,1,1)**
Configurations	**(S S S)**	0	$\left(\frac{1}{3}\right)^n$	$\left(\frac{2}{3}\right)^n\left(1-2\left(\frac{1}{2}\right)^n\right)$	$3\left(\frac{1}{3}\right)^n-3\left(\frac{2}{3}\right)^n+1$
	(I S S)	$\left(\frac{1}{3}\right)^n$	$\left(\frac{1}{3}\right)^n(2^n-1)$	$1-2\left(\frac{2}{3}\right)^n+\left(\frac{1}{3}\right)^n$	0
	(I I S)	$\left(\frac{2}{3}\right)^n$	$1-\left(\frac{2}{3}\right)^n$	0	0
	(I I I)	1	0	0	0

Table C.3: Conditioned probabilities for $l = 3$ and $n \geq 1$ for reduced configurations and outputs. To retrieve Tab. (C.4) is sufficient to create a new matrix with each row and column repeated exactly $\binom{l}{s}$ and $\binom{l}{u}$ times respectively, where s is the number of sensitive detector for the specific configuration while u is the number of 1s relative to a given output. As illustrated in Chap. (8), the new table has $l + 1$ rows and columns.

		Outputs							
		(0,0,0)	**(0,0,1)**	**(0,1,0)**	**(0,1,1)**	**(1,0,0)**	**(1,0,1)**	**(1,1,0)**	**(1,1,1)**
Configurations	**(S S S)**	0	$\left(\frac{1}{3}\right)^n$	$\left(\frac{1}{3}\right)^n$	$\left(\frac{2}{3}\right)^n\left(1-2\left(\frac{1}{2}\right)^n\right)$	$\left(\frac{1}{3}\right)^n$	$\left(\frac{2}{3}\right)^n\left(1-2\left(\frac{1}{2}\right)^n\right)$	$\left(\frac{2}{3}\right)^n\left(1-2\left(\frac{1}{2}\right)^n\right)$	$3\left(\frac{1}{3}\right)^n-3\left(\frac{2}{3}\right)^n+1$
	(S S I)	$\left(\frac{1}{3}\right)^n$	0	$\left(\frac{1}{3}\right)^n\left(2^n-1\right)$	0	$\left(\frac{1}{3}\right)^n\left(2^n-1\right)$	0	$1-2\left(\frac{2}{3}\right)^n+\left(\frac{1}{3}\right)^n$	0
	(S I S)	$\left(\frac{1}{3}\right)^n$	$\left(\frac{1}{3}\right)^n\left(2^n-1\right)$	0	0	$\left(\frac{1}{3}\right)^n\left(2^n-1\right)$	$1-2\left(\frac{2}{3}\right)^n+\left(\frac{1}{3}\right)^n$	0	0
	(S I I)	$\left(\frac{2}{3}\right)^n$	0	0	0	$1-\left(\frac{2}{3}\right)^n$	0	0	0
	(I S S)	$\left(\frac{1}{3}\right)^n$	$\left(\frac{1}{3}\right)^n\left(2^n-1\right)$	$\left(\frac{1}{3}\right)^n\left(2^n-1\right)$	$1-2\left(\frac{2}{3}\right)^n+\left(\frac{1}{3}\right)^n$	0	0	0	0
	(I S I)	$\left(\frac{2}{3}\right)^n$	0	$1-\left(\frac{2}{3}\right)^n$	0	0	0	0	0
	(I I S)	$\left(\frac{2}{3}\right)^n$	$1-\left(\frac{2}{3}\right)^n$	0	0	0	0	0	0
	(I I I)	1	0	0	0	0	0	0	0

Table C.4: Conditioned probabilities for $l = 3$ and $n \geq 1$ for all possible Sensitive-Insensitive configurations and outputs.

D Matlab code

```
%-------------------------------------------------------------------------
%Computation of Conditioned min-Entropy for the generalized model including
%Dark count, Crosstalk and Afterpulsing effects
%-------------------------------------------------------------------------
mu = 5.47;      %efficiency (5.47% for LASER - 41.4% for LED)
mu = mu/100;
alphamin = 0;
alphamax = 35;
alpha_step = .5;
start_dec = 3;
end_dec = 16;
count_t = 0;

prob_dark = 0.204;          %Dark count probability
prob_cross = 9.67*10^-5;    %Crosstalk probability
prob_puls = 0.0429;         %Afterpulsing probability
prob = 1 - (1-prob_dark)*(1-prob_cross)*(1-prob_puls);  %total probability
l_step = .01;

total = zeros(floor(prob/l_step) + 1,1,...
    (alphamax-alphamin)/alpha_step + 1,end_dec-start_dec + 1);

for t = start_dec:end_dec

    count_t = count_t + 1;
    %-------------------------------------------------------
    %Extract the photon matrices for different values of n
    %-------------------------------------------------------
    if t <= 9
        string_1 = strcat('F:\Laboratorio\photons\00',num2str(t),'riv\');
        cd(string_1);
        files = dir(strcat('riv_0',num2str(t),'_*.dat'));
        nchoosek_files = dir(strcat('nchoosek_',num2str(t),'.dat'));
    elseif t >= 10 && t <= 99
        string_2 = strcat('F:\Laboratorio\photons\0',num2str(t),'riv\');
        cd(string_2);
        files = dir(strcat('riv_',num2str(t),'_*.dat'));
        nchoosek_files = dir(strcat('nchoosek_',num2str(t),'.dat'));
    elseif t >= 100
        string_3 = strcat('F:\Laboratorio\photons\',num2str(t),'riv\');
        cd(string_3);
        files = dir(strcat('riv_',num2str(t),'_*.dat'));
        nchoosek_files = dir(strcat('nchoosek_',num2str(t),'.dat'));
```

```
end
disp(['Analizing system with ' num2str(t) ' detectors.............'])
count_alpha = 0;

%-------------------------------------------------------
%Execute system analysis for different values of
%incident photons mean (alpha)
%-------------------------------------------------------
for alpha = alphamin:alpha_step:alphamax

    count_alpha = count_alpha + 1;
    count_l = 0;

    for l = 0:l_step:prob
        count_l = count_l + 1;
        branch1 = 0;
        branch = 0;
        disp(['Prob: ' num2str(count_l) '/' num2str(...
            floor(prob/l_step) + 1)])
        for a = 1:length(files)

            file_curr = files(a).name;
            detectors = t;
            file_ID = fopen(files(a).name);
            test_table = fread(file_ID,files(a).bytes,'double');
            table = zeros(detectors+1);
            count = 0;

            for b =1:detectors+1:size(test_table,1)
                count = count + 1;
                table(count,:) = test_table(b:b+detectors,1);
            end

            clear tabella_test

            photons = a;

            %-------------------------------------------------------
            %Create all possible active-inactive configurations
            %for a given number of detectors
            %-------------------------------------------------------
            goofy_test = 0;
            inactive = zeros(detectors+1,detectors);

            for c = 2:detectors+1
                goofy_test = goofy_test + 1;
                for d = 1:goofy_test
                    inactive(c,d) = 1;
                end
            end

            maximum = 0;
```

```
%-------------------------------------------------------------------------
%Section for the Conditioned min-Entropy calculus
%-------------------------------------------------------------------------

        prob_max = zeros(size(table,1),1);

        for e = 1:size(table,1)
            prob_max(e,1) = max(table(e,:));
        end

            file_ID_nchoosek = fopen(nchoosek_files.name);
            nchoosek_table = fread(file_ID_nchoosek,...
                nchoosek_files.bytes,'double');

            for f = 1:size(prob_max,1)
                maximum = maximum + prob_max(f,1)*mu^(sum(...
                    inactive(f,:)))*(1-l)^(detectors-...
                    sum(inactive(f,:)))*l^(sum(inactive(f,:)))*...
                    nchoosek_table(f,1);
            end
%-------------------------------------------------------------------------

                branch1 = maximum*poisspdf(photons,alpha);

                branch = branch + branch1;

                fclose('all');
                clear ans
            end

            total(count_l,1,count_alpha,count_t) = ...
                -log2(poisspdf(0,alpha) + branch);

        end
    end
end
%-----------------------------------------------------------
%Calculate minimum over all possible values of "prob"
%given the condition "0 <= l <= prob"
%-----------------------------------------------------------
minimum_dummy = min(total);
minimum(1:size(minimum_dummy,3),1:size(minimum_dummy,4)) = ...
    minimum_dummy(1,1,:,:);
clear minimo_dummy
%-----------------------------------------------------------
%Plot the figure for all chosen detectors
%-----------------------------------------------------------
figure
plot(alphamin:alpha_step:alphamax,minimum)
```

Bibliography

[1] N METROPOLIS and S ULAM. The Monte Carlo method. *Journal of the American Statistical Association*, 44:335–341, 1949.

[2] Ralph C. Merkle. Protocols for public key cryptosystems. In *IEEE Symposium on Security and Privacy*, pages 122–133, 1980.

[3] A Uchida. Review on ultra-fast physical random number generators based on optical random phenomena. *The Review of Laser Engineering*, 39(7):508–514, 2011.

[4] Peter Elias. The efficient construction of an unbiased random sequence. *The Annals of Mathematical Statistics*, 43:6, 1972.

[5] J Von Neumann. Various techniques used in connection with random digits, 1951.

[6] Michael Mitzenmacher. Advanced Multilevel Strategy (AMLS). 1996.

[7] Yuval Peres. Iterating Von Neumann's Procedure for Extracting Random Bits, 1992.

[8] ACM Transactions on Modeling and Computer Simulation (TOMACS).

[9] Ian Goldberg and David Wagner. Randomness and the netscape browser. *Dr Dobb's Journal*, 1996.

[10] Satoshi Sunada, Takahisa Harayama, Kenichi Arai, Jun Muramatsu, Kazuyuki Yoshimura, Ken Tsuzuki, Peter Davis, and Atsushi Uchida. Theory and Experiments of Fast Non-Deterministic Random Bit Generation Using On-Chip Chaos Lasers. *Procedia IUTAM*, 5:190–194, 2012.

[11] Atsushi Uchida, Kazuya Amano, Masaki Inoue, Kunihito Hirano, Sunao Naito, Hiroyuki Someya, Isao Oowada, Takayuki Kurashige, Masaru Shiki, Shigeru Yoshimori, Kazuyuki Yoshimura, and Peter Davis. Fast physical random bit generation with chaotic semiconductor lasers Supplementary Material. *Nature Photonics*, 2(12):Supplementary Material, 2008.

[12] I. Reidler, Yaara Aviad, Michael Rosenblum, and Ido Kanter. Ultrahigh-Speed Random Number Generation Based on a Chaotic Semiconductor Laser. *Physical Review Letters*, 103:1–4, 2009.

[13] Kunihito Hirano, Kazuya Amano, Atsushi Uchida, Sunao Naito, Masaki Inoue, Shigeru Yoshimori, Kazuyuki Yoshimura, and Peter Davis. Characteristics of fast physical random bit generation using chaotic semiconductor lasers. *IEEE Journal of Quantum Electronics*, 45:1367–1379, 2009.

[14] Pu Li, Yun-Cai Wang, and Jian-Zhong Zhang. All-optical fast random number generator. *Optics express*, 18:20360–20369, 2010.

[15] A Alkassar, T Nicolay, and M Rohe. Obtaining true-random binary numbers from a weak radioactive source. *Computational Science and Its Applications - Iccsa 2005, Pt 2*, 3480:634–646, 2005.

[16] Markus Rohe. RANDy-A True-Random Generator Based on Radioactive Decay. pages 1–36, 2003.

[17] Thomas Jennewein, Ulrich Achleitner, Gregor Weihs, Harald Weinfurter, and Anton Zeilinger. A Fast and Compact Quantum Random Number Generator. 2008.

[18] Wayne M. Itano, J. C. Bergquist, Randall G. Hulet, and D. J. Wineland. Radiative decay rates in Hg^+ from observations of quantum jumps in a single Ion. *Physical Review Letters*, 59:2732–2735, 1987.

[19] Th Sauter, W. Neuhauser, R. Blatt, and P. E. Toschek. Observation of quantum jumps. *Physical Review Letters*, 57:1696–1698, 1986.

[20] Daniela Frauchiger, Renato Renner, and Matthias Troyer. True randomness from realistic quantum devices. *arXiv.org*, quant-ph:12, 2013.

[21] Roger Colbeck and Renato Renner. No extension of quantum theory can have improved predictive power. *Nature communications*, 2:411, 2011.

[22] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential Separation for One-Way Quantum Communication Complexity, with Applications to Cryptography, 2009.

[23] Robert König and Renato Renner. Sampling of min-entropy relative to quantum knowledge. *IEEE Transactions on Information Theory*, 57:4760–4787, 2011.

[24] Robert König, Ueli Maurer, and Renato Renner. On the power of quantum memory. *IEEE Transactions on Information Theory*, 51:2391–2401, 2005.

[25] Renato Renner. Security of Quantum Key Distribution. *International Journal Of Quantum Information*, 06:1, 2005.

[26] Renato Renner and Robert Koenig. Universally composable privacy amplification against quantum adversaries. (20):14, 2004.

[27] Robert König, Renato Renner, and Christian Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55:4337–4347, 2009.

[28] Charles H. Bennett, Gilles Brassard, and Jean-Marc Robert. Privacy Amplification by Public Discussion, 1988.

[29] Russell Impagliazzo, Leonid A Levint, and Michael Luby. Pseudo-random generation from one-way functions (Extended Abstract). In *STOC '89 Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 12–24, 1989.

[30] R. Impagliazzo and D. Zuckerman. How to recycle random bits. *30th Annual Symposium on Foundations of Computer Science*, 1989.

[31] Charles H. Bennett, Gilles Brassard, Claude Crepeau, and Ueli M. Maurer. Generalized privacy amplification. In *IEEE International Symposium on Information Theory - Proceedings*, page 350, 1994.

[32] J Lawrence Carter and Mark N Wegman. Universal classes of hash functions. *Journal Of Computer And System Sciences*, 18:143–154, 1979.

[33] Mark N. Wegman and J.Lawrence Carter. New hash functions and their use in authentication and set equality. *Journal of Computer and System Sciences*, 22:265–279, 1981.

[34] Douglas R Stinson. Universal hash families and the leftover hash lemma, and applications to cryptography and computing. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 42:3–31, 2002.

[35] Giacomo Mauro D'Ariano, Paoloplacido Lo Presti, and Paolo Perinotti. Classical randomness in quantum measurements. 5979:8, 2004.

[36] David McMahon. Quantum Measurement Theory. pages 121–146, 2007.

[37] N. Wheeler. Generalized Quantum Measurement Imperfect meters and POVMs. (September):1–31, 2012.

[38] Mark Fox. *Quantum Optics - An Introduction*, volume 1. Oxford University Press, Oxford, 10 edition, 2006.

[39] Rodney Loudon. *The Quantum Theory of Light*, volume 1. Oxford University Press, Oxford, 3 edition, 2001.

[40] Ivan Rech, Antonino Ingargiola, Roberto Spinelli, Ivan Labanca, Stefano Marangoni, Massimo Ghioni, and Sergio Cova. Optical crosstalk in single photon avalanche diode arrays: a new complete model. *Optics express*, 16(12):8381–8394, 2008.

[41] L Gallego, J Rosado, F Blanco, and F Arqueros. Modeling crosstalk in silicon photomultipliers. *Journal of Instrumentation*, 8:P05010–P05010, 2013.

[42] S Cova, M Ghioni, A Lacaita, C Samori, and F Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics*, 35:1956–1976, 1996.

[43] R. Ben-Michael, M.a. Itzler, and B. Nyman. Afterpulsing Effects in 1.5 micrometers Single Photon Avalanche Photodetectors. *LEOS 2006 - 19th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, pages 783–784, 2006.

[44] Micro Photon Devices. SPC2 Series - Single Photon Counting Camera.

[45] Ulf Leonhardt. Quantum Physics of Simple Optical Instruments. 2003.

[46] M. Santha and U.V. Vazirani. Generating Quasi-Random Sequences From Slightly-Random Sources. *25th Annual Symposium onFoundations of Computer Science, 1984.*, 1984.

[47] William J. Morokoff and Russel E. Caflisch. Quasi-Random Sequences and Their Discrepancies, 1994.

[48] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, 1985.

[49] S Rukhin, A., Soto, J., Nechvatal, J., Smid M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, A., Vo. Statistical Test Suite for Random and Pseudorando Number Generators for Cryptographic Applications. *Special Publication 800-22 National Institute Standart Technology*, 2010.

[50] Ueli M. Maurer. A universal statistical test for random bit generators. *Journal of Cryptology*, 5:89–105, 1992.

[51] Soto Juan. Statistical Testing of Random Number Generators. 2, 1999.

[52] Andrew Rukhin, Juan Soto, James Nechvatal, Smid Miles, Elaine Barker, Stefan Leigh, Mark Levenson, Mark Vangel, David Banks, Alan Heckert, James Dray, and San Vo. A statistical test suite for random and pseudorandom number generators for cryptographic applications. *National Institute of Standards and Technology*, 800:131, 2010.

[53] M Stipcevic and B Medved Rogina. Quantum random number generator. *Time*, page 7, 2006.

[54] F. James. A review of pseudorandom number generators, 1990.

[55] M. Fiorentino, C. Santori, S. M. Spillane, R. G. Beausoleil, and W. J. Munro. Secure self-calibrating quantum random-bit generator. *Physical Review A - Atomic, Molecular, and Optical Physics*, 75, 2007.

[56] Miloš Drutarovskỳ and Pavol Galajda. A robust chaos-based true random number generator embedded in reconfigurable switched-capacitor hardware. *Radioengineering*, 16:120–127, 2007.

[57] Oliver Knill. Probability Theory and Stochastic Processes with Applications. page 373, 2009.

[58] Milton Abramowitz and Irene Stegun. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards Applied Mathematics Series 55, 10th edition, 1972.

[59] Jean Dickinson Gibbons and Douglas A Wolfe. Nonparametric Statistical Inference. *Technometrics*, pages 185–194, 2003.

[60] Pal Revesz. *Random walk in random and non-random environments*, volume 1. World Scientific Publishing Co. Pte. Ltd., 2th edition, 2005.

[61] G Marsaglia. A current view of random number generators. In *Computer Science and Statistics. Proceedings of the Sixteenth Symposium on the Interface*, pages 3–10, 1985.

[62] R N Bracewell. *The Fourier Transform and its Applications*, volume 42. 1986.

[63] A.D. Barbour, L Holst, and S Janson. *Poisson approximation*. Oxford University Press, 1992.

[64] O. Chrysaphinou and S Papastavridis. A Limit Theorem on the Number of Overlapping Appearances of a Pattern in a Sequence of Independent Trials. *Probability Theory and Related Fields*, 79:129–143, 1988.

[65] Js Coron and David Naccache. An accurate evaluation of Maurer's universal test. *Selected Areas in Cryptography*, 1999.

[66] Adi Shamir. On the Security of {DES}. In *Advances in Cryptology—CRYPTO˜'85*, number 1560, pages 280–281, 1985.

[67] Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of Applied Cryptography*, volume 106. CRC Pres, 1997.

[68] I.J. Good. *The serial test for sampling numbers and other tests for randomness*. Proc. Cambridge Philos. Soc., 1953.

[69] S Pincus and B H Singer. Randomness and degrees of irregularity. *Proceedings of the National Academy of Sciences of the United States of America*, 93:2083–2088, 1996.

[70] Frank Spitzer. *Principles of Random Walk*. Springer New York, 1964.

[71] M Baron and Andrew Rukhin. Distribution of the Number of Visits For a Random Walk. *Communications in Statistics: Stochastic Models*, 15, 1999.

[72] Werner Schindler and Wolfgang Killmann. Evaluation Criteria for True (Physical) Random Number Generators Used in Cryptographic Applications. *Informationstechnik*, 2523:431–449, 2003.

[73] Werner Schindler. Efficient online tests for true random number generators. *Hardware and Embedded Systems-CHES 2001*, pages 103–117, 2001.

[74] Joohyun Han, Yongyun Cho, and Jaeyoung Choi. *Computational Science and Its Applications – ICCSA 2005*, volume 3481. 2005.

[75] G. Marsaglia and A. Zaman. Monkey tests for random number generators, 1993.

[76] George Marsaglia and Wai Wan Tsang. Some difficult-to-pass tests of randomness. *Journal Of Statistical Software*, 7:1–9, 2002.

[77] Pierre L'Ecuyer and Richard Simard. TestU01: A C library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, 33:22, 2007.

[78] Pierre L'Ecuyer, Richard Simard, and Stefan Wegenkittl. Sparse Serial Tests of Uniformity for Random Number Generators. *SIAM Journal on Scientific Computing*, 24(2):652–668, 2002.

[79] R. E. Barlow. Theoretical Statistics, 1966.

[80] D. R. Cox and D. V. Hinkley. *Theoretical statistics*. Chapman and Hall, London, 1974.

[81] Thorsten Dickhaus. Randomized p-values for multiple testing of composite null hypotheses. *Journal of Statistical Planning and Inference*, 143:1968–1979, 2013.

[82] Alessio Farcomeni. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Statistical methods in medical research*, 17:347–388, 2008.

[83] Joshua D. Habiger and Edsel A. Peña. Randomised P-values and nonparametric procedures in multiple testing. *Journal of nonparametric statistics*, 23(3):583–604, 2011.

[84] John W. Pratt. Length of Confidence Intervals. *Journal of the American Statistical Association*, 56:549–567, 1961.