Towards Deployable Quantum Communication: Integrated Quantum Key Distribution and Randomness Generation

De Matos Afonso Pereira, Maria Ana

UNIVERSITÉ DE GENÈVE
Section de Physique
Département de Physique Appliquée

FACULTÉ DES SCIENCES
Professeur Hugo Zbinden
Docteur Rob Thew

Towards Deployable Quantum Communication: Integrated Quantum Key Distribution and Randomness Generation

THÈSE
présentée à la Faculté des sciences de l'Université de Genève pour obtenir le grade de Docteur ès sciences, mention physique

par

Maria Ana de Matos Afonso Pereira
de Coimbra, Portugal

Thèse N° 5967

GENÈVE
Centre d'impression Uni Mail
2026

Se puderes olhar, vê. Se puderes ver, repara.

José Saramago

Soutenue le 21 Novembre 2025 devant le jury composé de:

Prof. Hugo Zbinden
Dr. Rob Thew
Prof. André Stefanov
Dr. Nino Walenta

Acknowledgements

I would like to express my deepest gratitude to all those who have supported and guided me throughout the course of this thesis. Firstly I would like to thank my supervisors Prof. Hugo Zbinden and Dr. Rob Thew for giving me the opportunity of undertaking this thesis. I would also like to thank Alberto, Fadri and Rebecka for the mentorship and guidance they provided at the beginning of this journey A special thanks goes to Davide Rusca, who has been an incredible source of support, advice and knowledge throughout the entire duration of this thesis. I am deeply grateful for Davide's patience and dedication in guiding me through the many challenges faced during this work, even long after he has left the group.

As progress in experimental physics relies heavily on high-level technical support. I want to express my particular thanks to Raphael Houlmann, who specializes in the hardware programming, and Claudio Barreiro, the greatest technician a group could wish for, who provided invaluable expertise in electronics and from whom I learned a great deal. I would also like to thank Giovanni Resta and David Cabrerizo for the support provided with the bonding of the integrated photonics in this thesis. I am also grateful to Tiff for having my back, for the countless discussions and brainstorming sessions.

Then I wish to thank my friends and colleagues from Geneva: Joey, Ana, Théo, Alexandre, Lorenzo, Towsif, Noshin, Alex and Daniela and many others for the great times we shared, both in and out of the lab. I would like to give a special mention to those closest to me, without whom this journey would have been much harder, maybe not even possible: Mingsong, beyond his scientific help and contributions, I am especially grateful for his exceptional personal support and friendship; Evi, for her unwavering friendship and for being the person I could always turn to, no matter the hour or the circumstance; and lastly Pauli, not only for agreeing to go to the gym with me at 5am, but also for all the for the guidance and belief he has shown in me over the years.

Finalmente chegou a altura de agradecer aos meus amigos de Portugal pelo apoio

e amizade ao longo destes anos. Ao Joando e ao Frazão por todos os momentos de descontração e risadas que tornaram esta jornada mais leve. À Bia, sempre presente, pela sua amizade incondicional. E à Lara, que me viu crescer, e foi sempre amiga das dezenas de *eus* que conheceu.
Por último, quero expressar a minha profunda gratidão à minha família pelo apoio e sacrifícios que fizeram para que eu pudesse chegar até aqui. Aos meus pais, pelo incentivo constante e por acreditarem em mim mesmo nos momentos em que eu duvidava de mim própria. Ao meu irmão, com quem eu sei que posso sempre contar. Ao meu avô Zé, que desde pequena me levou à escola, andou comigo para todo o lado e fez da sua casa um lugar de almoço quente. E por fim, à minha avó Duarte, que infelizmente não está cá para ver este momento, mas cuja personalidade forte, visão progressista e ensinamentos deixaram uma marca profunda em quem sou hoje. A todos, o meu mais sincero obrigado.

Abstract

Quantum technologies are rapidly transforming the landscape of secure communication and computation by exploiting the fundamental principles of quantum mechanics. Among these, Quantum Key Distribution (QKD) offers an unprecedented level of security in communications, provided by physical principles rather than computational assumptions. However, the realization of scalable and practical quantum networks remains challenging. Key technological hurdles include the need for high-sensitivity and high-speed single-photon detectors, stable and integrated optical platforms, and robust and fast true random number generators. Overcoming these challenges is essential to bridge the gap between laboratory demonstrations and the real-world deployment of quantum networks. This thesis addresses some of these critical issues through work in single-photon detection technologies, integrated QKD system implementation, and high-speed Quantum Random Number Generators (QRNGs).

Single-Photon Avalanche Diodes are fundamental detectors for quantum technologies, enabling high-sensitivity measurements at the single-photon level. This thesis investigates the operating principles, limitations, and advancements of Single-Photon Avalanche Diodes, with a focus on high-speed dual-anode designs. Their photon detection efficiency, dark count rate, afterpulsing probability, and timing jitter are analyzed, while optimized control electronics and capacitance management strategies are developed. The characterization of Single-Photon Avalanche Diodes demonstrates their suitability for quantum communication, particularly in quantum key distribution and asynchronous heralded photon sources, with pathways for further improvements in speed and integration.

Building on these detector technologies, a proof-of-principle integrated quantum key distribution system based on the BB84 protocol is presented. The system incorporates time-bin encoding, decoy-state methods, and photonic integrated circuits for both transmitter and receiver. Key challenges, such as chromatic dispersion and interferometer stability, are addressed through novel circuit and system-level de-

signs. System-level testing and field deployment validate the feasibility of compact, high-performance integrated quantum key distribution implementations.

Finally, the thesis explores quantum random number generation through semi-device-independent protocols. An experimental homodyne-based SDI-QRNG is realized, supported by custom transimpedance amplifier design and photonic integrated circuit integration. The setup achieves high bandwidth, strong quantum-to-classical noise separation, and stable operation, enabling practical random number generation rooted in fundamental quantum processes.

Overall, this work advances the development of integrated quantum technologies by improving room-temperature single-photon detection, demonstrating scalable QKD prototypes, and implementing a high-speed QRNG. Together, these contributions pave the way toward secure, integrated quantum communication systems and practical quantum information applications ready to deploy in real world networks.

Résumé

Les technologies quantiques transforment rapidement le paysage des communications et du calcul sécurisés en exploitant les principes fondamentaux de la mécanique quantique. Parmi celles-ci, la QKD offre un niveau de sécurité sans précédent dans les communications, garanti par des lois physiques plutôt que par des hypothèses computationnelles. Cependant, la réalisation de réseaux quantiques pratiques et évolutifs demeure un défi majeur. Les principaux obstacles technologiques incluent la nécessité de détecteurs de photons uniques à haute sensibilité et à grande vitesse, de plateformes optiques stables et intégrées, ainsi que de générateurs de nombres aléatoires véritables, rapides et robustes. Surmonter ces défis est essentiel pour combler le fossé entre les démonstrations en laboratoire et le déploiement réel des réseaux quantiques. Cette thèse aborde certains de ces enjeux critiques à travers des travaux sur les technologies de détection de photons uniques, la mise en œuvre d'un système intégré de QKD et le développement de QRNGs à haut débit.

Les diodes à avalanche déclenchées par un photon unique sont des détecteurs essentiels pour les technologies quantiques, car elles permettent des mesures extrêmement sensibles au niveau du photon unique. La recherche durant ma thèse a étudié en profondeur les principes de fonctionnement de ces dispositifs, leurs limitations et les avancées récentes, en particulier dans le cas de l'architecture à double anode conçue pour atteindre des régimes de fonctionnement très rapides. Leurs performances sont caractérisées en termes d'efficacité de détection de photons, de taux de comptage de bruit, de probabilité d'auto-déclenchement et de jitter, tout en optimisant les systèmes de contrôle électroniques et en développant une stratégie de gestion d'inductance. Les résultats de caractérisation démontrent que ces détecteurs sont particulièrement adaptés aux communications quantiques, notamment à la distribution de clés quantiques et à la génération asynchrone de photons uniques. Des perspectives d'amélioration de la vitesse et l'intégration de ces détecteurs sont également proposées.

À partir de ces avancées, une première démonstration d'un système intégré de distri-

bution de clés quantiques est réalisée. Celui-ci repose sur un protocole fondamental de communication quantique basé sur l'encodage temporel et l'utilisation d'états atténués. L'intégration sur circuits photoniques, tant du côté de l'émetteur que du côté du récepteur, permet de concevoir une plateforme compacte et performante. Les principaux défis, tels que la dispersion chromatique dans les fibres optiques et la stabilité des interféromètres, sont surmontés grâce à des choix originaux dans la conception optique et électronique. Des tests en laboratoire et sur le terrain confirment la viabilité de cette approche.

Enfin, ce travail explore la génération de nombres aléatoires quantiques, en s'appuyant sur des protocoles semi-indépendants du dispositif. Une réalisation expérimentale est proposée, reposant sur la détection homodyne équilibrée. Celle-ci intègre un amplificateur transimpédance optimisé, ainsi qu'un circuit photonique dédié comprenant modulateurs, atténuateurs et amplificateurs optiques. L'ensemble permet d'obtenir une large bande passante, une séparation nette entre le bruit quantique et le bruit classique, et une stabilité de fonctionnement adaptée aux applications pratiques.

Dans leur ensemble, ces résultats contribuent à l'avancement des technologies quantiques intégrées en améliorant la détection des photons uniques, en démontrant des prototypes de communication sécurisée et en mettant en œuvre une génération rapide et fiable de nombres aléatoires quantiques. Ils ouvrent la voie au déploiement de systèmes de communication quantique sécurisés et d'applications concrètes de l'information quantique.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The protection of information has been a central concern of societies for as long as information itself has carried value [2, 3]. From the simple substitution ciphers of ancient civilizations [4], to the mechanical encryption devices of the 20th century such as the Enigma machine [5], the ability to conceal messages has often shaped the outcomes of political, military, and economic struggles. Cryptography, in this sense, has always evolved in response to the threats of its time. As new means of communication emerged, so did the need for new methods to secure these communications.

As we advance through this digital era, vast amounts of personal, financial, and strategic information flow continuously across global networks. All these banking transactions, medical records, private conversations and even state secrets leave behind a digital footprint. As a result, more than ever, the confidentiality, integrity, and authenticity of information are paramount and a concern at the forefront of individuals and organizations alike. The rise of cybercrime and data breaches has underscored the vulnerabilities inherent in our interconnected world [6–8]. Moreover, the advancements of powerful computational technologies have raised the stakes even more, as adversaries are no longer limited to human codebreakers but can implement sophisticated algorithms, and increased processing power can compromise traditional cryptographic schemes. At the same time, mass surveillance capabilities of corporations have expanded to unprecedented levels, often operating in legal gray areas that challenge traditional notions of privacy. All of this naturally raises concerns for individuals regarding the security of their personal data and communications that they so critically depend on.

The security of most digital communications relies on public-key cryptography [9]. Public key cryptography, also known as asymmetric cryptography, relies its

safety on one-way functions and the mathematical difficulty of factoring integers. 'Difficulty' here means that the solving time increases exponentially with the amount of information. So the security of these systems is not based on mathematical principles but rather computational limitations and time constraints. In a public-key cryptosystem, two keys are used: a public key and a private key. One of the parties creates a public key that anyone who wishes to communicate with them can use to encrypt their message. This encrypted message however, can only be decrypted by whoever possesses a private key. The first practical application of asymmetric cryptosystem was in 1978 by Rivest, Shamir, and Adleman, in a protocol named after them - RSA [10]. This protocol underpins many of the secure protocols used in everyday internet activities, such as online banking and messaging. Until recently, this reliance seemed well-placed. However, Shor's algorithm [11], showed that a sufficiently powerful quantum computer can factor large integers and solve discrete logarithms efficiently (solves in polynomial time with respect to the amount of information), significantly reducing the time required to break widely used cryptographic schemes. Even though a quantum computer capable of implementing Shor's algorithm at a sufficiently large scale to comprise current cryptographic systems has not yet been built, the possibility, and to some, the certainty, of its future existence has raised concern about not only future communications and data, but also of previously intercepted communications and stored data. This has put into question the longevity of privacy since all encrypted data could be stored now and decrypted later. It is also important to note that classical algorithms capable of breaking public-key cryptography may also be discovered in the future, further exacerbating these concerns.

As a result, information-theoretic secure methods, which do not rely on computational assumptions, have garnered renewed interest. The One-Time Pad (OTP) is a theoretically unbreakable encryption technique that uses a random key that is as long as the message itself, where each binary bit of information from the plaintext is XORed with the corresponding bit from the key. The constraints of the OTP are that the key must be **truly random**, **used only once**, and **kept secret**. If these conditions are met, then the encryption cannot be broken without knowledge of the key, regardless of computational power. However, despite its theoretical appeal, OTP is not frequently used due to the logistical hurdles associated with generating, distributing, and securely storing large keys. As a result, its use is typically limited to highly specialized contexts, such as diplomatic or military communications. The aforementioned key distribution problem in OTP cannot be overcome with classical

methods. Many classical key distribution methods have been proposed (like the Diffie-Hellman key exchange [12]), but they all rely on computational assumptions that are vulnerable to quantum attacks, bringing us back to the original problem.

It is in this context that quantum information technologies have emerged. Unlike classical cryptographic schemes, quantum cryptography exploits the principles of quantum mechanics to provide theoretical secure communication. In brief, we first recall that it is impossible to create copies of an arbitrary unknown quantum state as described by the no-cloning theorem [13]. Additionally, when information is encoded in non-orthogonal quantum states, any attempt to obtain information on the communication, will result in a high probability of a disturbance of the transmission [13,14]. As a result, any attempts to furtively access quantum encoded data would be immediately identified by the legitimate users. Of the various quantum-based cryptographic protocols, using QKD with OTP stands out as a promising solution for practical and information-theoretic secure communication, by generating and distributing secret keys between two parties with security guaranteed by the laws of quantum mechanics [15].

However, QKD protocols alone do not address the need for true randomness, even though they rely on true randomness as a resource. Nonetheless, quantum mechanics provides yet another solution to this problem: QRNGs. As the inherent probabilistic nature of quantum measurements can be harnessed to generate truly random numbers, QRNGs exploit this property to produce sequences of bits that are fundamentally unpredictable. Accordingly, this thesis addresses the further development of both QKD and QRNG protocols as two fundamental components of secure communication systems.

On the practical side of implementing quantum-enabled secure communication, progress in QKD has been closely driven by the advancements in single-photon detector technologies, which have enabled significant improvements in key generation rates and transmission distances. These advancements have been made possible by the introduction of new materials, device architectures, and operational techniques that enhance the performance of single-photon detectors. On this front breakthroughs have been enabled by Superconducting Nanowire Single-Photon Detectors (SNSPDs). While the theoretical limit for point-to-point, fiber-based, repeaterless QKD (not refering to TF-QKD) is approximately 500km [16], this boundary was nearly reached in 2018 by Boaron et al. [17] who demonstrated QKD over 421km. Such performance was directly linked to the capabilities of the (then) state-of-the-art SNSPDs used, which had ultra-low dark count rates and timing jitters while provid-

ing detection efficiencies of 40-60%. Complementing these distance achievements, recent work by Grünenfelder et al. [18] addressed key bottlenecks in practical QKD implementation. By developing custom multipixel SNSPD arrays combined with fast acquisition electronics and real-time key distillation capabilities, the record-breaking secret key rate of 64 Mbps was achieved. These parallel advances in maximizing both transmission distance and key generation rate illustrate how single photon detection technology has enabled QKD systems to approach fundamental theoretical limits while simultaneously enhancing practical performance parameters for real-world applications.

While the latest SNSPDs offer superior performance metrics - including photon detection efficiencies beyond 90%, extremely low dark count rates, and timing jitter down to picosecond levels [19]- their requirement for cryogenic cooling significantly increases the system complexity, cost, power consumption and volume. This limitation presents a significant barrier to the widespread deployment of QKD technologies in real-world applications, especially where compact form factors, room-temperature and continuous operation are essential requirements. In contrast, Single-Photon Avalanche Diode (SPAD) technology presents a more practical and cheaper alternative. SPADs offer significant advantages in terms of reliability, compactness, lower operational voltage, and the ability to function at room temperature or with minimal cooling. These characteristics make SPADs particularly well-suited for field-deployed QKD systems.

Other than detector architecture, the community has also explored alternative materials to detect single photons in the Near-Infrared (NIR). A notable example is Germanium (Ge), which has a cutoff frequency (the wavelength at which absorption drops sharply) at 1.55µm. As early as the 60s, Ge-based photodiodes were demonstrated [20, 21]. However, they suffered from high dark currents and poor noise performance when compared to Silicon (Si) detectors [22]. As the interest in optical communications and consequently detections in the $1.3-1.55$µm range grew, Ge was largely overshadowed by Indium Gallium Arsenide (InGaAs) platforms, and research into Ge detectors stagnated. However, with the rise of Si photonics in the 2000s, Ge has started to gain some interest again, as it is compatible with Complementary Metal-Oxide-Semiconductor (CMOS) fabrication processes and can be epitaxially grown on Si. The interest to integrate detectors with Si photonics and the potential for low-cost manufacturing has driven renewed research into Ge-based single photon detectors for LiDAR, quantum communication, and time resolved imaging [23]. Strained and doped Ge layers have been investigated to enhance responsivity and

suppress noise, while Germanium Tin (GeSn) alloys offer the possibility of extending sensitivity further into the infrared while mitigating thermal noise. In our early collaborative work with École Polytechnique Fédérale de Lausanne (EPFL) on Ge SPADs, we saw both the potential and the persistent challenges in developing these approaches, particularly for quantum applications where performance demands are especially stringent.

However, alongside the reduced efficiency performance when compared to SNSPDs, SPADs still face a fundamental challenge that limits the performance of QKD systems at short distances, and that is afterpulsing. Afterpulsing happens when a charge carrier is trapped during an avalanche event and is subsequently released, triggering a false detection [24, 25]. A way to mitigate noise from afterpulsing is to wait for the trapped carriers to be released before re-arming the detector [26]. This waiting time is called hold-off time and can be in the order of hundreds of µs, thereby significantly limiting the key generation rates [27]. The first work described in this thesis was motivated by the need to address this bottleneck of SPADs and improve their performance for QKD applications. This, alongside the need for practical and accessible QKD systems, led to finding a solution that was small in footprint and could operate at high rates and with low noise. This involved investigating a novel commercially available detector design that claims to enhance SPADs speed and reliability while tackling the afterpulsing issue (Park et al. [28]).

While the improved detectors have been the main contributors to the key rate and distance progress in QKD, they do not directly enable practicality and scalability of QKD systems. Moving beyond proof-of-principle demonstrations to real-world applications requires overcoming other significant engineering challenges, including system integration, miniaturization, and cost reduction. In our group in 2023, Sax et al. [1] demonstrated a proof-of-principle integrated QKD system. This work showed that many of the essential components of QKD could be integrated, pointing to the possibility of a cost-effective and deployable solution.

However, this work was also far from being an deployable prototype. Several challenges became evident that prevented the system from reaching the level of miniaturization and robustness required for large scale deployment. These included the use of bulk power supplies, function generators, and laser driving electronics that contributed to a large system footprint. The system also relied on chromatic dispersion compensation fibers to mitigate pulse broadening over long distances. As we will discuss in detail in chapter 3, given the speed of the time-bin encoded scheme used, the 200ps time bins made the system particularly susceptible to chro-

matic dispersion, which severely limited the maximum distance achievable without compensation. The receiver chip was also non-tunable, with the basis selection probability fixed on-chip at 90:10. This limited the system to be used at large distances, or incredibly low mean photon numbers in order to not saturate the detectors. The non-tunability also meant that the system could not adapt to changing channel conditions, which would be essential in a real-world deployment. The system also required bulky detectors and readout electronics: The photonic integrated circuit itself required large control systems, including two FPGAs per communication unit, with one used solely as a communication bridge between the computer and the other FPGA responsible for the experimental control. Additionally, external filtering components were required to narrow the laser's linewidth and improve signal quality, further complicating the setup.

These obstacles reflected not only technical limitations, but also broader difficulties faced by the QKD community in transitioning from laboratory feasibility to commercial scalability. In this thesis we set out to address the previously mentioned challenges, with the goal of moving integrated QKD systems closer to practical deployment. In short, virtually all Printed Circuit Board (PCB) footprints were reduced and integrated into a single unit per communication party. The receiver chip was also redesigned to include a basis selection element. The external filtering and variable coupler attenuator were substituted by an on-chip ring filter and Variable Optical Attenuator (VOA), respectively, on the transmitter. The integration of compact, detectors and readout electronics further streamlined the setup, making it fit into a standard 19in 3U rack unit. Another significant improvement was the removal of the bulky dispersion compensation fibers. This was achieved by halving the system's repetition rate from 2.5GHz to 1.25GHz, thereby doubling the time bin width to 400ps and reducing the system's susceptibility to chromatic dispersion. With these improvements we hoped to collectively advance the integrated QKD system towards a practical and deployable solution, addressing key bottlenecks that had previously hindered its scalability.

As mentioned earlier in this chapter, true randomness is another critical component of secure communication systems, and key to executing a secure OTP. QRNGs provide a way to guarantee unpredictability by exploiting the inherent probabilistic nature of quantum mechanics. Traditional Pseudo-Random Number Generators (PRNGs) rely on deterministic algorithms to produce sequences of seemingly random numbers. While it can be sufficient for many applications, PRNGs fall short in scenarios where high security is paramount. In contrast, QRNGs exploit the

inherent probabilistic nature of quantum phenomena such as single-photon detection, phase noise, or vacuum fluctuations, to produce truly random outputs. Several architectures have been explored, ranging from trusted-device implementations, to Device Independent (DI) implementations where no information on the internal workings of the hardware is assumed, and Semi-Device Independent (SDI) protocols that relax/remove assumptions about the internal workings of the hardware. While DI-QRNGs are theoretically robust they are extremely challenging to implement experimentally and suffer from low generation rates. As a result, SDI-QRNGs have surged as a compromise, offering high generation rates and practical security guarantees under more realistic conditions. Our group has been working on QRNGs for several years, and the most recent contribution was the development of a self-testing, SDI QRNG [29] prototype. This SDI approach allowed us to certify the randomness without having to fully characterize the entirety of the system. The self-testing aspect allowed the user to verify in real-time that the setup is functioning correctly, guaranteeing the generation of certified random bits. Yet, just as in QKD, the gap between this proof-of-principle demonstration and a practical deployment remains significant. In order to improve QRNGs, the most obvious direction to follow would be, as with QKD, the on-chip integration of electronic and optical components. The realization of QRNGs on a Photonic Integrated Circuit (PIC) would significantly reduce their footprint and power consumption, allowing them to directly replace traditionally used random number generators within existing devices. The successful integration of QRNGs on-chip would also allow them to be embedded directly within larger quantum communication systems.

This thesis explores the intersection of single-photon detection technologies, quantum key distribution, and quantum random number generation, with a focus on addressing the practical challenges that currently limit their widespread deployment. The work presented here aims to advance the state-of-the-art in these areas by investigating novel detector designs, integrated QKD and QRNG systems, and scalable implementations. The core theme here can therefore be summarized as **bridging the gap between laboratory demonstrations and deployable technologies in quantum communication**. The body of this thesis is organized as follows:

Chapter 2 introduces the fundamentals of single-photon detection, with emphasis on fast-gated InGaAs SPADs and brief mention of Ge-based devices. It then presents the work carried out to improve the performance of Dual Anode SPADs, as well as their experimental performance characterizations.

Chapter 3 summarizes the basics of QKD and describes the integrated work car-

ried out in our group. We will discuss the main limitations encountered, and analyze the gap between proof-of-principle demonstrations and prototype readiness. The bulk of the chapter is dedicated to the improvements made to the original PIC based system, and to presenting the results obtained with the upgraded prototype.

Chapter 4 focuses on quantum random number generation, outlining the theoretical framework of self-testing approaches, reports on our work on the integrated implementation thus far, and discusses the challenges and future directions for practical QRNG systems.

Chapter 5 summarizes the main contributions of this work and places them in the context of the broader research landscape, and identifies future directions for advancing QKD, QRNGs, and detector technologies towards scalable deployment.

Chapter 2

High-Speed Single-Photon Avalanche Diodes

2.1 Introduction to Single-Photon Avalanche Diodes

Single-photon detection is a critical component in quantum photonic systems, particularly for practical implementations of QKD. In recent years, progress in QKD has been driven by the development of advanced single-photon detectors, which have enabled significant improvements in key generation rates and transmission distances. These advancements have been made possible by the introduction of new materials, device architectures, and operational techniques that enhance the performance of single-photon detectors.

Recent breakthroughs have been enabled by SNSPDs. While the theoretical limit for point to point repeaterless QKD is approximately 500km [16] this boundary was nearly reached in 2018 by Boaron et al. [30] who demonstrated QKD over 421 km. Such performance is directly linked to the capabilities of the state-of-the-art SNSPDs used, which achieved ultra-low dark count rates and jitter while providing detection efficiencies of 40-60%. Complementing these distance achievements, recent work by Grünenfelder et al. [18] addressed key bottlenecks in practical QKD implementation. By developing custom multipixel SNSPD arrays combined with fast acquisition electronics and real-time key distillation capabilities, the record-breaking secret key rate of 64 Mbps was achieved. These parallel advances in maximizing both transmission distance and key generation rate illustrate how single photon detection technology has enabled QKD systems to approach fundamental theoretical limits while simultaneously enhancing practical performance parameters for real-world applications.

While SNSPDs offer superior performance metrics - including photon detection efficiencies beyond 90%, extremely low dark count rates, and timing jitter down to picosecond levels - their requirement for cryogenic cooling significantly increases the system complexity, cost, power consumption and volume. This limitation presents a significant barrier to the widespread deployment of QKD technology in real-world applications, especially where compact form factors, room-temperature and continuous operation are essential requirements. In contrast, SPADs present a more practical and cheaper alternative. SPADs offer significant advantages in terms of reliability, compactness, lower operational voltage, and ability to function at room temperature or with minimal cooling. These characteristics make SPADs particularly well-suited for field-deployed QKD systems where considerations of size, power consumption, and operational simplicity are paramount.

However, SPADs still face a fundamental challenge that limits their application in high-speed QKD systems at short distances: afterpulsing. This phenomenon occurs when carriers trapped during an avalanche event are subsequently released, triggering false detection events that introduce errors in the quantum communication protocol. To mitigate afterpulsing effects, conventional SPAD operation requires implementing long dead times, which significantly limit the maximum achievable count rate and, consequently, the key generation rate.

This chapter will examine recent advances in high-speed SPAD technologies designed specifically to overcome afterpulsing limitations. Approaches including self-differencing techniques [31–35], sine-wave gating [**?**, 32, 36–39] methods, and innovative dual-anode SPAD architectures [28] enable efficient discrimination of weak avalanche signals while minimizing afterpulsing probabilities.

In particular, this chapter will detail my work in implementing dual-anode SPADs operating at 1GHz gating frequencies. We will present the electronic systems developed for these detectors, including customized readout circuitry, control mechanisms, and parameter optimization techniques. Our approach demonstrates how these specialized components can be integrated into a practical system that achieves the settings required for QKD applications, while addressing the critical challenges of afterpulsing and signal discrimination.

2.1.1 Operating Principle of a p-i-n Junction Diode

As one may deduce by the name, single-photon avalanche diodes are p-n or p-i-n diode junctions, consisting of p-type semiconductors with positively charged carriers (holes) and n-type semiconductors with negatively charged carriers (electrons). As

depicted in Figure 2.1 a p-n junction is formed by joining a p-type semiconductor and an n-type semiconductor together, while in the p-i-n junction an intrinsic (undoped) type semiconductor is placed between the two.

When the two materials are placed together, due to the concentration gradient, the free electrons from the n-type material diffuse to the p-type material and combine with the free holes, creating a negatively charged area on the p-side near the junction and leaving a positively charged area on the n-type material. These two regions of charged material, together, make the depletion region. The depletion region continues to grow until the negative charge of the n region, repels the diffusion of more electrons into the p region. As the depletion region continues to grow, the also growing electric field created by these charged regions, opposes the diffusion of more electrons until an equilibrium is reached and the region seases to grow, and as a result, a potential difference is created across the junction [40].

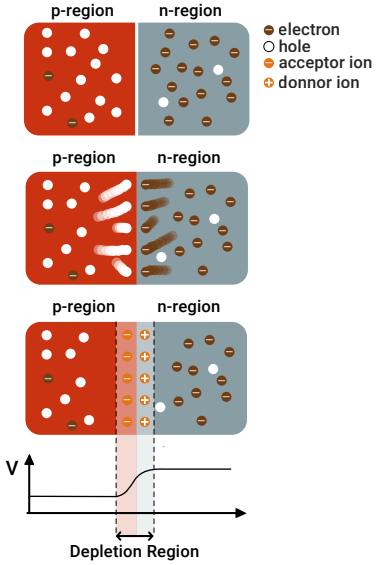


Figure 2.1: Non-biased p-n junction.

As seen in Figure 2.2, if one were now to *bias* this junction, meaning apply a DC voltage across the junction, the potential difference across the junction can be manipulated. For simplicity, from here onwards, since we are are considering a diode scenario, the p region shall be denoted *anode* and the n region *cathode*. There are two ways one can bias the junction, namely forward bias and reverse bias.

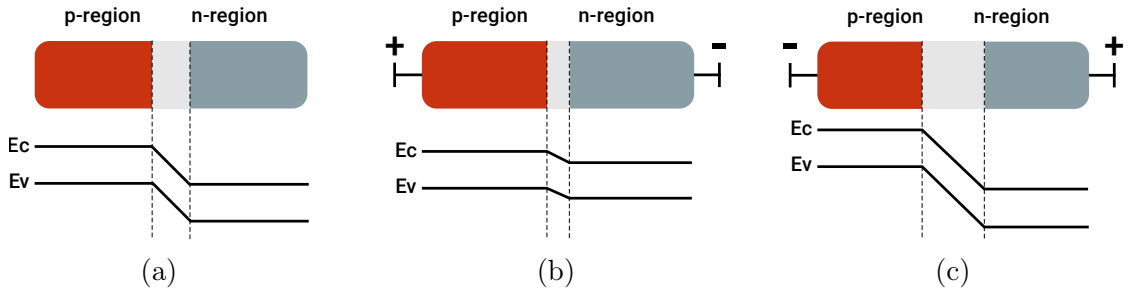


Figure 2.2: a) Non-biased p-n junction. b) Forward biased p-n junction. c) Reverse biased p-n junction.

If we first look at the forward bias scenario Fig. 2.2b, where the cathode is connected to the negative terminal of the source and the anode to the positive, the voltage across the juntion is reduced, disturbing the equilibrium state and causing

the majority carriers to drift towards the junction. As the free electrons reach the p region, they recombine with the holes in the valence band and are subsequently attracted to the electric positive terminal of the region, giving rise to a current flow. This current is proportional to the number of carriers that are injected into the depletion region, and is therefore proportional to the applied voltage.

In the reverse bias scenario, the positively charged electrode connected to the n region attracts the free electrons towards the positive electrode, whereas in the p type material, the holes are accelerated towards the negative electrode. The net result is that the depletion layer grows wider resulting in the increased potential barrier and electric field, leading to a halt in the current flow, turning the diode into an insulator. It is important to mention that even in this scenario a small amount of current (usually micro-amperes) can still be measured going from the n side to the p side due to thermally generated minority carriers (electrons in the p region and holes in the n region). This is referred to as the leakage current. A standard p-n junction is designed to conduct current easily in the forward direction but blocks current in reverse bias up to a certain threshold voltage. Even in reverse bias, only a very small reverse leakage current flows, which is stable and non-destructive. This controlled one-way conduction is what makes diodes useful as rectifiers, signal limiters, and general electronic switches. SPADs, however, are intentionally reverse biased beyond that threshold. A detailed explanation of their principle of operation follows in the next subsection.

2.1.2 Operating Principle of SPADs

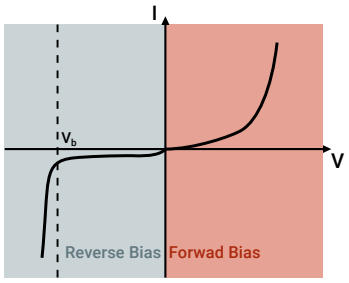


Figure 2.3: Representative IV curve of a diode. I - Current; V - Voltage; V_B - Breakdown voltage.

As shown in Fig.(2.3), if one keeps increasing the reverse voltage applied to a diode, past a certain threshold we will see the current gain rising sharply with the applied voltage. This is because the magnitude of electric field is sufficient to accelarate the minority charges in the depletion region such that their kinetic energy

is enough to ionize atoms of the semiconductor material, generating additional free carriers. These offspring carriers will then impact ionize other free carriers in an exponential manner, resulting in a high-current signal that can be detected by designated electronics. This effect is called an avalanche breakdown, and is the main mechanism behind semiconductor single photon detectors. The current through the diode is stopped as soon as the diode voltage drops below a certain value, ceasing the avalanche. This voltage is referred to as breakdown voltage (V_b) and is the minimum reverse bias voltage required to initiate the avalanche multiplication process [41]. More accurately, this process is not instantaneous and can occur within a narrow voltage interval, due to the stochastic nature of the multiplication process. The voltage at which avalanches occur also suffers from some hysteresis, meaning that the voltage at which the avalanche stops (quenching voltage) is slightly lower than the voltage at which it starts (breakdown voltage). This happens because once the avalanche is initiated, the high density of free carriers in the multiplication region alters the local electric field. Once the avalanche process is initiated, the current will continue to increase rapidly, and if it is not limited the junction can be destroyed due to thermal runaway [42], which is when the rate of heat generation becomes greater than the rate of heat removal.

More generally, for a reverse-biased photodiode, three operation modes can be defined: linear mode, avalanche mode and Geiger mode. In linear mode the didode is biased below V_b and the current is proportional to the incoming light, with Gain$\leq$1. In avalanche mode, the diode is biased slighlty above V_b, but the current is still proportional to the incoming light with a Gain$\leq$5000 (material and architecture dependent [43]). This is the region where Avalanche Photodiodes (APDs) are operated, the linear avalanche mode still offers linearity, allowing APDs to respond proportionally to incoming light intensity. This is particularly beneficial when accurately measuring signal strength is required—such as in laser rangefinding, fiber-optic communication, imaging sensors, and other precision optical systems [44]. In Geiger Mode, the current gain is "infinite" due to high reverse bias, thus allowing the detection of single photons. Each absorbed photon can trigger a self-sustaining avalanche that produces a large, digital-like pulse independent of the photon's energy. This photon counting mode is the basis of operation for SPADs. SPADs also provide information regarding the time of arrival of each photon potentially up to tens of picosecond precision, thus allowing us to perform photon arrival time statistics.

Both the fabrication, device structure and operating conditions play a key role

in determining the performance of SPADs. The main performance parameters of SPADs are the Photon Detection Efficiency (PDE), Dark Count Rate (DCR), Afterpulsing Probability (APP) and timing jitter. These parameters and how they are measured are discussed in detail in the following subsections.

2.1.2.1 Photon Detection Efficiency

Photon Detection Efficiency (PDE) is definied [45] as the product of the probability of having a photon absorbed (absorption probability density, η_{abs}), the probability of the generated free carrier being collected by the high multiplication field (collection probability, P_{coll}) and the probability of a carrier in the depletion region generating an avalanche (triggering probability, P_{trig}):

$$\text{PDE} = \int_0^{x_{end}} (1 - R) \cdot \eta_{abs}(x) \cdot P_{coll}(x) \cdot P_{trig}(x) dx \tag{2.1}$$

where the R is the reflection coefficient of the surface of the detector, integrated over the possible absorption depths x.

2.1.2.2 Dark Count Rate

The number of detections of an avalanche without an incident photon per unit time is referred to as the Dark Count Rate (DCR). DCs can be generated due to thermal effects like the Shockley-Read-Hall (SRH) mechanism and field-mediated effects such as Band-to-Band Tunneling (BTBT) as studied in [46] and [47]. At room temperature, the dominant source of DCs is the Shockley-Read-Hall (SRH) mechanism, more specifically carrier trap-assisted thermal generation of free carriers. In the work of [47], the authors show that the SRH generation rate is proportional to the trap density in the material and consequently its volume and fabrication. In a more detailed analysis, the SRH mechanism can have an enhancment factor dependent on the electric field. This will make the Dark Count Rate (DCR) dependent on both fabrication and operation conditions. Relatively independent to temperature, Band-to-Band Tunneling (BTBT) happens due to high electric fields in the depletion region, allowing the excitation of carriers from the valence band to the conduction band by tunneling effect. As temperatures lower, this becomes the dominating source of DCs. To mitigate this effect, a proper engineering of the electric field in the depletion region is required as demonstrated in [47].

2.1.2.3 Afterpulsing Probability

Afterpulsing refers to a set of phenomena that result in charge carriers being trapped in the depletion region during an avalanche event [24,25]. These traps typically have a lifetime on the order of microseconds [25], and the trapped carriers can be released either when the SPAD is reverse-biased below or above the breakdown voltage, and in the latter case a spurious signal correlated to a previous event will be registered. The probability of this happening is called the Afterpulsing Probability (APP) and is dependent on the trap density in the material, the electric field in the depletion region, deep-level capture probability, the triggering efficiency and the avalanche charge.

From the above, it is clear that both the DCR and Afterpulsing Probability (APP) are dependent on the material quality and the design of the detector, but the operating conditions also significantly influence the performance of SPADs. Minimizing the number of deep-level defects is essential to reduce the number of trapped carriers. This translates to higher quality fabrication processes and the use of high-quality wafers. To reduce trapping, it is crucial to keep the charge passing through the device during an avalanche as low as possible. This can be ensured by keeping the excess bias voltage as low as possible. However, the PDE is directly proportional to the excess bias voltage, thus creating a trade-off between having high PDE and low APP.

2.1.2.4 Jitter

The jitter of a SPAD is the time difference uncertainty between the arrival of a photon and the detection of the avalanche signal and is a crucial indicator of the timing resolution of the system. Jitter is affected by different components, all of them affected by statistical fluctuations. As discussed in the previous section, the avalanche starts in the depletion region. However, looking at Fig. 2.4, we can further divide this region into two sub-regions: the multiplication region, and the drift region [48]. The multiplication region, with the strongest electric field, is where the avalanche is initiated and the carriers are multiplied. What can be referred to as the drift region, contains usually the absorption region, and with a low electric field, accelerates the generated carriers towards the anode. Carriers generated close to the high electric region will naturally have a shorter drift time than carriers generated further away. Moreover, some free carriers may even be generated in the neutral area of the depletion region (no electric field): This scenario is responsible for the exponential tail seen in jitter measurements [49].

Another contributor to the timing response of a detector is the build-up time of the avalanche, i.e. its propagation time. This can divided into two distinct phases, namely initial current growth and lateral propagation. In the first phase, current is concentrated in a narrow filament around the photon absorption point, developing parallel to the electric field vector. During the second phase, this current begins to spread across the entire SPAD area through two main mechanisms; multiplication-assisted diffusion and hot carrier effects.

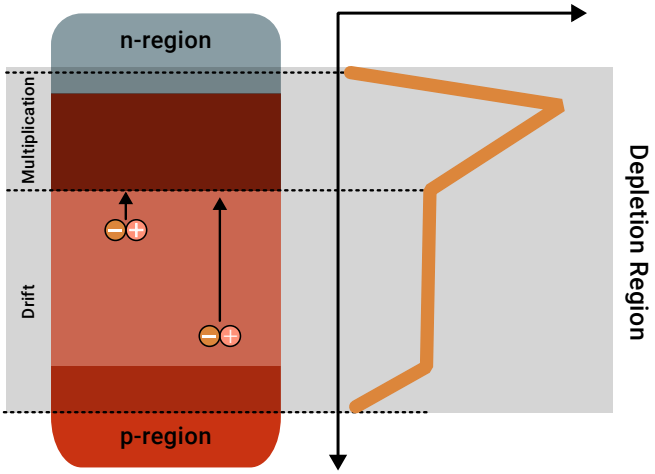


Figure 2.4: Detailed structure of a SPAD. The depletion region can be divided into two sub-regions: the multiplication region, and the drift region.

Multiplication-assisted diffusion occurs due to the sharp carrier density gradient between the initial filament and surrounding regions. Carriers gradually diffuse outward, triggering avalanches in adjacent areas. This mechanism introduces statistical fluctuations in two ways: the process itself is inherently random, involving both diffusion and impact ionization; the propagation pattern varies depending on where the initial photon is absorbed (center vs. edge of detector), which is also random [50]. Hot carriers within the filament can emit secondary photons. These photons may then be absorbed elsewhere in the detector, potentially triggering new avalanches in different regions of the same device [51].

2.1.3 SPADs for Near Infra-Red Detection

In general, Si-SPADs provide by far the best ratings among all the materials used for single photon detection [52]. Silicon's spectral response is well-matched to the 400-1000nm wavelength range, enabling many applications in quantum optics requiring single-photon detection, including free-space quantum key distribution [53], quantum imaging [54], and quantum memories [55]. This wavelength range also benefits from the high-quality single-photon sources available and low-loss integrated platforms [56–59]. However, for applications requiring telecom wavelengths (1310µm and 1550µm), such as fiber-based QKD, silicon's bandgap does not allow for efficient photon absorption, leading to very low detection efficiencies. This requires the use

of alternative materials with narrower bandgaps that can effectively absorb photons in the near-infra-red range.

As such, two main approaches have emerged for detecting infrared photons in quantum applications. The first is frequency conversion, while the second involves direct infrared detection through materials such as InGaAs and Ge. Frequency conversion involves converting the infrared photons to visible wavelengths, which can then be detected by Si-SPADs. This approach has been used in free-space QKD experiments [60–62]. However, frequency conversion has its drawbacks as it is a complex process, can introduce additional noise, and reduces the overall efficiency of the system. In contrast, the direct detection approach using infrared SPADs, while not yet matching the performance of superconducting detectors, offers a key practical advantage: the ability to operate near room temperature. This makes them an increasingly viable option for quantum photonics applications. In the next section, we will discuss the most common material used for infrared SPADs: InGaAs.

2.1.3.1 InGaAs-SPADs

The material most commonly used for SPADs in quantum communications is InGaAs due to its bandgap, $E_g = 0.75eV$, at room temperature, which corresponds to a cut-off wavelength of around 1600nm, making it ideal for detections at telecom wavelengths. Research on InGaAs-based SPADs began as early as the mid 1990s, when commercially available APDs were explored for Geiger-mode operation [63,64]. In the late 2000s, the first detectors designed for Geiger-mode operation were developed [65].

The most widely used design for InGaAs single-photon detectors is the separate absorption, charge, and multiplication (SACM) InGaAs/Indium Phosphide (InP) structure [66, 67] (see Figure 2.5). This structure was developed to overcome the high noise and inefficient avalanche breakdown typical of InGaAs detectors at the time. InP is particularly well suited for this role because its wide bandgap (1.35 eV) makes it transparent to telecom light (limiting the absorption to the absorption layer (InGaAs)) and is also effective at suppressing thermally generated carriers while still supporting the high electric fields needed in the multiplication layer to achieve strong avalanche gain. Additionally, InP is lattice-matched to InGaAs [68, 69], allowing high-quality epitaxial growth with minimal defects. These combined properties make the SACM InGaAs/InP structure highly effective for reliable single-photon detection.

Referring back to Figure 2.5, the Separate Absorption Charge and Multiplication

(SACM) structure consists of three main layers: the absorption layer, the charge layer, and the multiplication layer. The absorption layer is made of InGaAs, which has a high absorption coefficient at telecom wavelengths, allowing efficient photon absorption. The thickness of this layer is optimized to balance absorption efficiency and carrier transit time. The charge layer, typically made of InP, serves to control the electric field distribution across the device. The correct engineering of the doping concentration and thickness of this layer is crucial to create a high electric field in the multiplication region while maintaining a lower field in the absorption region to minimize premature breakdown and consequently reducing the DCR. Finally, the multiplication layer, also made of InP, is where avalanche multiplication occurs. This layer is engineered to have a high electric field that enables impact ionization, leading to a rapid increase in carrier density and a high avalanche current when a photon is absorbed, making it easier to detect by discriminating electronics. Modern SPADs also feature anti-reflective coatings to reduce surface reflections and enhance photon absorption, further improving the overall detection efficiency.

The performance of SPADs is not only dependent on fabrication, but is also closely linked to their operating conditions. Because avalanche events in SPADs are self-sustaining, proper quenching (the process of stopping the avalanche) is essential both to protect the device from breakage and to restore the detector to a ready state for subsequent photon arrivals. The performance of InGaAsInP SPADs therefore evolved hand-in-hand with quenching strategies. Three main quenching approaches are used: **passive quenching**, **active quenching**, and **gated mode operation** [70].

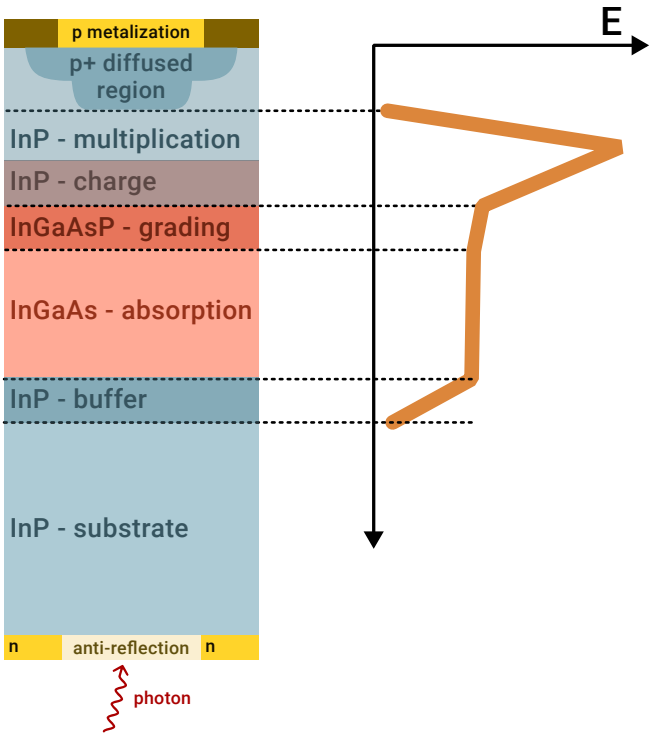


Figure 2.5: Detailed SACM structure of a SPAD and corresponding electric field profile.

Passive Quenching: In these circuits, the avalanche current quenches itself by

developing a voltage drop across a high value resistance load. This resistor, typically in the order of $10^4\Omega$, is chosen to be high enough to lower the voltage below the breakdown voltage, but low enough to allow for a fast recovery time. Since the detector and resisitor form an RC circuit, the recovery time is dependent on the product of the resistance and the capacitance of the diode. In practice, this trade-off between effective quenching and fast recovery constrains the performance of passive circuits. This is why active circuits are more appealing to high-speed applications [71–73].

Active Quenching: To avoid a slow recovery time from avalanches, a new approach was implemented, that is active quenching. In this approach, the avalanche current is sensed by a fast comparator, which then rapidly switches the bias voltage source to below breakdown voltage. After a designated hold-off time, the bias voltage is restored to the operating level. The advantages offered by this approach are the fast recovery time and avalanche detection, as well as the short and well-defined durations of the avalanche current and of the dead time [26, 74–76].

Gated Mode Operation: Another common strategy used for quenching SPADs is to use Gated Quenching Circuits. With these circuits, the SPAD is biased below the breakdown voltage. A gate signal (square or sine wave) is then applied across the diode. This raises the bias voltage above breakdown only during the time the gate signal is high, allowing for the avalanche to be triggered. The gate signal is at the same time responsible for quenching the avalanche by lowering the bias voltage below breakdown at the end of the gate width. This method allows for high-speed operation and can be synchronized with other system components.

Hybrid Quenching: Hybrid quenching schemes exploit the strengths of both passive and active quenching methods. These schemes can be active-quenching with passive-reset, passive-quenching with active-reset and even mixed passive-active quenching [77, 78].

2.1.3.2 Ge-SPADs

An alternative material that has been looked into for single photon detection in the infra-red is germanium. With an indirect bandgap of 0.66eV (1.88 µm) and a direct bandgap of 0.80eV (1.55 µm), Ge is sensitive to the entire Short-Wave Infrared (SWIR) range, making it suitable for telecom applications [20]. Among possible

combinations, Ge-on-Si structures are particularly interesting due to their potential for monolithic integration with silicon photonics, which could in turn greatly reduce fabrication costs [21, 22, 79]. This architecture uses germanium as the absorber and silicon as the multiplication layer, exploiting the strong absorption of Ge in the SWIR range and the favorable avalanche properties of Si. Silicon has a superior epitaxial quality compared to InP, which is commonly used in InGaAs-SPADs, leading to lower defect densities and improved device performance [22]. Additionally, the type-I band alignment between Ge and Si allows efficient carrier transport without the need for a grading layer, simplifying device design [80].

However, the development of Ge-on-Si SPADs has faced significant challenges. The substantial Ge/Si lattice mismatch (4.2%) causes relaxation defects. These defects come in the form of recombination centers leading to high DCRs and dislocations that worsen afterpulsing [22]. Another limitation comes from Ge's band structure. While it has a direct bandgap at 0.8eV, efficient absorption at these wavelengths requires room temperature operation [22]. However, to suppress DCR, cryogenic cooling is typically needed. This trade-off between absorption efficiency and noise performance has been a major hurdle in realizing practical Ge-on-Si SPADs.

The development of Ge-on-Si SPADs nevertheless progressed considerably over the past decade. The first demonstration came in 2011, when Lu et al. [81] reported Ge-on-Si devices operating in Geiger mode, though the reported performance was later found to be mischaracterized. In 2013, Warburton et al. [82] improved the design by thickening the Si multiplication layer and employing a mesa structure, which achieved modest photon detection efficiencies but suffered from high DCR due to surface trap states at the etched mesa walls. A breakthrough came in 2019 with the work of Vines et al. [22] who introduced a planar architecture. By distancing the active region from the mesa edges, they reduced surface electric fields and thus lowered the DCR. These devices achieved up to 38% PDE at 1.3µm, demonstrated LiDAR operation, and exhibited improved afterpulsing performance when compared to InGaAs/InP SPADs. Building on this progress, Llin et al. [80] in 2020 demonstrated devices with smaller active areas, reducing trap densities and achieving PDE up to 25% with DCRs as high as 1MHz at 165K.

Although encouraging results have been reported, Ge-on-Si SPADs still lag behind InGaAs/InP devices in key performance metrics. Their performance is strongly limited by dark counts and the need for cryogenic cooling, which hinders their practical use and prevents compatibility with compact thermoelectric cooling. Looking forward, one promising route to overcome these issues is alloying Ge with tin (GeSn),

which can enhance absorption in the SWIR range while maintaining CMOS compatibility. GeSn-on-Si avalanche photodiodes have already demonstrated superior responsivity at 1.55µm compared to pure Ge devices, but experimental demonstrations of GeSn-based SPADs remain purely theoretical due to significant fabrication challenges.

2.1.3.3 Work on germanium SPADs

In collaboration with Prof. Fontcuberta at EPFL and her (then) doctoral student Dr. Giunto, we worked on the characterization of GeSn-on-Si SPADs. Their work encompassed the first steps toward the realization of SPADs based on GeSn-on-Si structures. The motivation behind it stemmed from the limitations of conventional InGaAs/InP SPADs used at 1.55µm for LiDAR and optical communication. InGaAs/InP devices are costly, rely on indium supplies (a scarce resource), and are not compatible with large-scale silicon integration. In contrast, GeSn offers the possibility of using abundant materials while enabling CMOS monolithic integration, making it an attractive candidate for scalable and cost-effective SWIR single-photon detection. The proposed device used a SACM architecture, where GeSn served as the absorber, grown on a Ge buffer on a Si substrate. Inspired by the planar SPAD design of Vines et al. [22], the architecture carefully distanced the active region from the mesa walls to mitigate electric-field hot spots and surface-related dark carriers. However, the devices suffered from metallic contamination and mechanical issues during wire bonding, which damaged anti-reflection coatings and led to poor reproducibility.

On our end, we focused on designing and building the electronic system necessary to operate the SPADs in Geiger mode. This work involved developing the biasing and gating circuits, as well as the readout electronics. In addition, we carried out IV characterizations across different temperatures and devices, measured the intrinsic capacitance of the detectors, and investigated their behavior under gated operation. The results of Dr. Giunto's work are presented in **Chapter 6** of his thesis [23].

Although still at an early exploratory stage, GeSn-on-Si SPADs could provide a CMOS-compatible and sustainable path to SWIR single-photon detection if these material and design challenges are addressed.

2.1.4 SPADs in Quantum Communications

Since the late 1990s, InGaAs-SPADs have become foundational components in quantum communication, powering early demonstrations of QKD [83–87]. Their sensi-

tivity to near-infrared photons aligns perfectly with telecommunications infrastructures, making them instrumental for real-world implementations. Improvements in fast-gated InGaAs SPADs have allowed for the realization of QKD with increased distances and Secret Key Rates (SKRs) [88,89]. These advances were further pushed by low-noise free-running SPADs, which pushed the limits of QKD to exceed 300km of fiber [90].

Beyond QKD, InGaAs-SPADs have been used in a wide array of other quantum communication protocols namely, quantum teleportation experiments [91] and Quantum Secret Sharing (QSS) [92, 93]. In terms of quantum resources, QRNG systems have relied on SPADs for high-quality, high-speed randomness generation [94]. SPADs have also played roles in experimental quantum repeaters and quantum memory systems—and continue to be vital for characterizing single-photon sources [95, 96].

In addition to these quantum-specific applications, InGaAs-SPADs are key in several non-quantum domains. Due to high atmospheric transmittance and reduced solar background noise, and eye safety of the near-infrared band, InGaAs-SPADs have also been utilized in remote sensing, aerosol monitoring LiDAR, and time-of-flight ranging [97–99]. SPADs have also been used in optical time-domain reflectometry (OTDR) [100], confocal fluorescence lifetime imaging and confocal-based fluorescence fluctuation spectroscopy [101, 102].

InGaAs-SPADs's combination of telecom-wavelength sensitivity, high detection efficiency, room-temperature (or Peltier-cooled) operation, and compatibility with array-based integration ensures their ongoing relevance in current and future quantum networks, LiDAR systems, and advanced photonic instrumentation. While superconducting detectors may offer superior performance in certain metrics, InGaAs-SPADs strike a balance between cost, complexity, and ease of deployability, making them the workhorse technology of choice across both research and industry.

2.1.4.1 Free Running SPADs

For asynchronous photon detection operations, InGaAs/InP SPADs are commonly operated in free-running mode. Ever since their introduction, the primary approach to implementing free-running SPADs have been through Negative Feedback Avalanche Diode (NFAD)s. The goal of NFADs is to monolithically integrate a high-value quenching resistor directly on-chip, in order to minimize the capacitive load [103, 104].

As mentioned, in a typical passive quenching circuit, the quenching resistor is

placed in series with the detector, in a control PCB for example, in order to quench the avalanche by the voltage drop is causes across the resistor. This large quenching resistance, alongside the parasistic capacitance of the circuit, lead to a large RC time constant, which in turn leads to a slow recovery time. This slow recovery time is detrimental to the performance of the detector, as it limits the maximum count rate achievable. To reduce the effect of stray capacitances provenient from the PCB traces, bondings and packaging, the idea to integrate the quenching resistor directly on-chip was proposed, and the concept of the NFAD was born [103].

When designing the integrated resistor of an NFAD, a trade-off has to be made between the quenching time and the reset time. The resistor causes a voltage drop proportional to both the avalanche current and the resistance value, causing the voltage in the diode to drop below breakdown, quenching the avalanche. Once the current flowing through the diode is stopped, the diode re-charges through the same resistor, until the bias voltage is restored to its initial value. The recharge time is dependent on the RC time constant of the integrated resistor and the diode's capacitance, as a consequence, the resistor's value must be high enough to allow proper and swift quenching to reduce APP, but low enough to allow for a fast reset time to not compromise the maximum achievable count rate.

In practice, NFADs improve in performance with the addition of an external active quenching circuit as these types of circuits allow for a precise control of the hold-off time and a faster quenching of the avalanche, which is crucial to reduce APP. Such hybrid approach has been demonstrated by Lunghi et al. [71] who showed how these detectors can benefit from both the low noise of NFADs and the fast quenching time of active quenching circuits.

In contrast to gated SPADs, where the device is biased above breakdown during narrow time windows synchronized with the photon arrival, NFADs operate in a continuous regime. In gated SPADs, the DCR is strongly influenced by the gate signal's duty cycle. Because the detector is biased above breakdown for a shorter amount of time, the chance of a thermally generated carrier initiating an avalanche is also reduced, leading to lower DCR values. In free-running SPADs, however, the detector is always biased above breakdown, making it susceptible to thermally generated carriers at all times.

This means that NFADs typically report higher DCRs when compared to gated SPADs operated under similar conditions. The DCR of NFADs can be significantly improved by cooling the device, as for low enough temperature, the thermally generated dark counts can be completely suppressed. This leaves only the field-assisted

tunneling effect as the main contributor to the DCR. As demonstrated by Korzh et al. in [90], DCRs as low as 1 cps can be achieved at 10% PDE when the device is cooled to -110^{o}C. For this reason, NFADs are often cooled to temperatures close to -100^{o}C. This in turn adds complexity and cost to the system, as a fridge is required, since this temperature is too low to be achieved with a simple Peltier cooler.

2.1.4.2 Gated SPADs

When the electronic signals of the gates are coupled to a SPAD, parasitic signals that superimpose with the avalanche signals are generated due to the inherent capacitance of the SPAD. Any rapid change in voltage across the diode causes a large transient spike proportional to both the capacitance and the rate of voltage change. These parasitic signals,commonly reffered to as Capacitance Response (CR), can be much larger than the avalanche signals from single photon events. Suppressing these derivative signals is crucial to effectively extract avalanche signals, and is a key task in the gated SPAD electronics. The presence of large CRs makes it challenging to detect avalanche signals which in turn affects in the overall performance of the detector. High CRs typically require higher excess bias voltages in order be able to discriminate single photon events, which negatively impacts both DCR and APP. The amplitude of the CR depends on the rise (and fall) time of the applied gates, the gate amplitudes as well as the quenching circuits. As the gate frequency increases, the CR becomes more pronounced, as both the avalanche signals get smaller, due to the reduced gate width, and the capacitive response gets larger, due to the increased rate of change of voltage. This means that the Signal-to-Noise Ratio (SNR) of the avalanche signals is significantly reduced at high frequencies. Nevertheless, increasing the gating frequency is vital for QKD and related applications, in order to increase the key generation rate. To overcome these challenges, two main techniques have been developed to allow for high frequency operation of SPADs: **Sine Wave Gating** and **Self-Differencing**.

Sine Wave Gating: This technique uses sine-wave gates instead of square pulses. The capacitive impedance of the diode $Z_C = \frac{1}{2\pi fC}$ decreases with increasing frequency, meaning that high-frequency components in the gating signal can couple more effectively through the junction capacitance, resulting in higher CR responses. This is the main advantage of using sine waves to gate the detectors, as unlike square pulses, which contain high-frequency harmonics, sine waves have minimal spectral content outside their fundamental frequency. Additionally, the smooth transitions of sine waves generate less voltage spikes

compared to the sharp edges of square pulses which can further contribute to the CR.

Given the simple frequency spectrum of sine waves, filtering techniques can be effectively employed to suppress the CR, as the parasitic signals can in principle be filtered out with band-stop/notch filters, leaving only the avalanches detectable [32,36,105]. With GHz-level frequencies, gating windows as short as $\sim$ 200ps are possible, significantly suppressing afterpulsing and raising count rates to hundreds of MHz [33, 106].

Self-Differencing: This technique splits and delays the output signals of the detector by exactly one gate period, then subtracts them to the current incoming output to cancel out the parasitic responses, isolating the avalanches. This approach creates two "copies" of the detector output signal, one direct path and one with a calibrated time delay, typically set to the gating period f_g. The principle behind this is that the CR is a periodic gating artifact and is equal between at every output, while avalanche events occur randomly and independently. When the delayed signal is subtracted from the direct signal, the periodic components subtract, filtering out the CR. A downside to this method is that it limits the maximum count rate to $\sim \frac{f}{2}$ [31, 107].

Approaches that combine Sine Wave Gating (SWG) and self-differencing [32,108] have also been demonstrated, leveraging the advantages of both techniques and reducing the filtering demands. SWG/hybrid approaches have become the most widely adopted technique for high-speed InGaAs/InP SPADs, due to its relative simplicity and effectiveness. To date, high-speed SWG SPADs have been key components in numerous practical QKD systems, including a 4600km space-to-ground quantum communication network [109, 110]. Nevertheless, pushing the performance of SWG SPADs remains an active area of research, with ongoing efforts to enhance detection efficiency, reduce dark counts, and increase gating frequencies. For PDE and DCR optimization, shortened gate widths have been found improve performance [34,37,39]. If one is working with a fixed gating frequency, when using SWG, a shorter gate can be achieved by increasing the peak to peak amplitude (V_{pp})of the gate signal. Losev et al. [39] however, demonstrated that the DCR increased significantly when the gate-off voltage was too low, limiting the increase in the gate amplitude. Exploiting the the V_{pp} tunability of SWG, Tada et al. [38] demonstrated an SWG Single Photon Detector (SPD) with an amplitude of 50 V_{pp} at 1.27GHz, and achieved a PDE of 53.4% and a DCR of 3.5×10^{-4}/gate. Another interesting

observation arising from the SWG studies is that the maximum achievable PDEs are higher at higher temperatures [34, 37]. The authors explained that this can be due narrowing of the absorption band at higher temperatures, increasing the detectors absorption efficiency.

To this day, much work is still being done to push the gating frequency of SWG SPADs even higher. Though gating frequencies have reached up to 2.5GHz [111], pushing the gating frequency even higher is still an active area of research. In the work of Liang et al. [112], it was reported that the afterpulse sharply increases when the gating frequency reaches 2.5GHz , showing that more work is needed to understand the limits of high frequency gating, on both the fabrication and the signal processing level.

2.2 Dual-Anode SPADs for High-Speed Quantum Applications

To address the performance bottlenecks of conventional single-anode SPADs, Park *et al.* [28] introduced a dual-anode architecture to enable improved speed and reliability in single-photon detection. These devices, manufactured and sold by Wooriro Co. Ltd., provide a practical and accessible platform for high-speed quantum experiments. While the dual-anode design is conceptually appealing, during our investigations the detectors showed certain performance constraints, particularly in imperfect capacitance matching, resulting in residual high capacitance responses. Nevertheless, their commercial availability provided a valuable testbed for exploring novel approaches to high-speed single-photon detection, and the work presented in this section was based on these detectors.

2.2.1 Operating Principle of DA-SPADs

The detectors used, refered to as Dual Anode Single-Photon Avalanche Diodes (DA-SPADs), were each comprised of two on-chip diodes separated by an isolation wall sharing a common cathode. Both diodes had similar architectures; however, the *dummy*, was engineered to have a higher breakdown voltage compared to its sibling. This was achieved by adjusting the relative thicknesses α (Figure 2.6a) of the multiplication layers of the diodes. Since a thicker multiplication layer requires a higher reverse bias voltage to achieve the same electric field strength needed for avalanche breakdown, this shifts the breakdown voltage in the *dummy* diode to

higher values. This diffence however is enough to reflect in the capacitance value of the diodes, which proves to be crucial in high-speed applications. To compensate for this difference, in fabrication, the active area diameter of the *dummy* diode was 1µm larger than that of the main diode. This design tries to make the capacitance of both diodes as similar as possible, while still allowing for a significant difference in breakdown voltage [28], to keep the *dummy* diode from detecting photons. The goal was then to use the *dummy* diode to create a parasitic signal that can be easily subtracted from the main avalanche signal in subsequent electronics, thus improving the SNR, which will in turn improve the APP and DCR, as smaller avalanches and consequently, smaller V_{bias} can be used to achive the same PDE.

The dual-anode architecture adopts a principle similar to that of self-differencing gating and the two detector method of Scarcella et al. [106]. However, it achieves suppression of the capacitance response using comparatively simpler electronics, as self-differencing techniques require signal splitting, precision-matched differential delay lines, complex impedance matching, and high-speed differential amplifiers to cancel out the capacitive response [113]. The DA-SPAD, on the other hand, generates the differential signal at the device level, which simplifies the electronic requirements and reduces the complexity of the readout system.

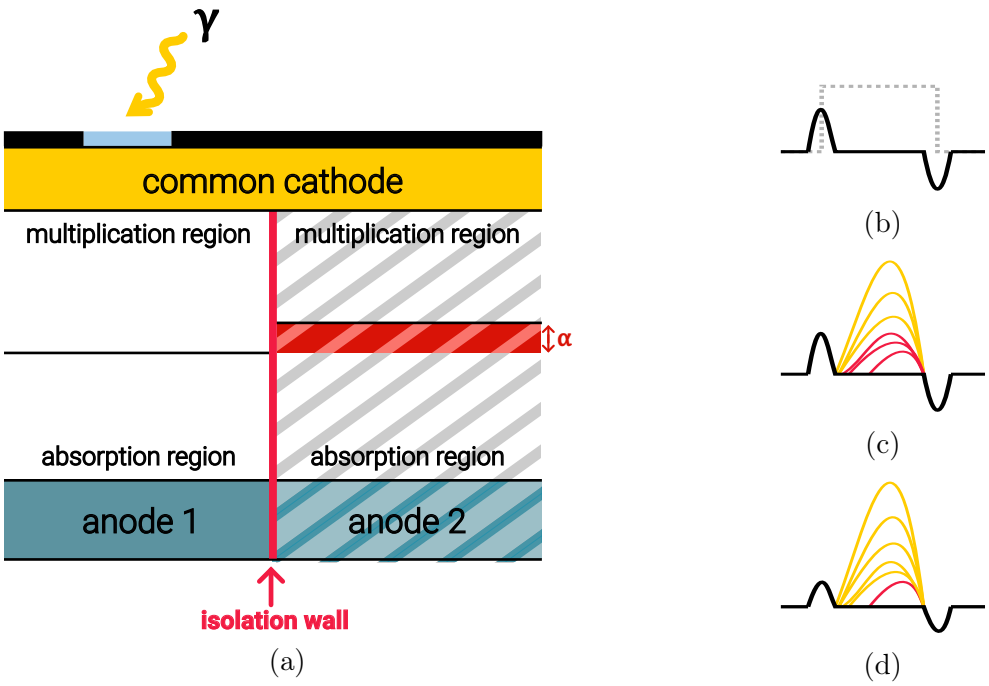
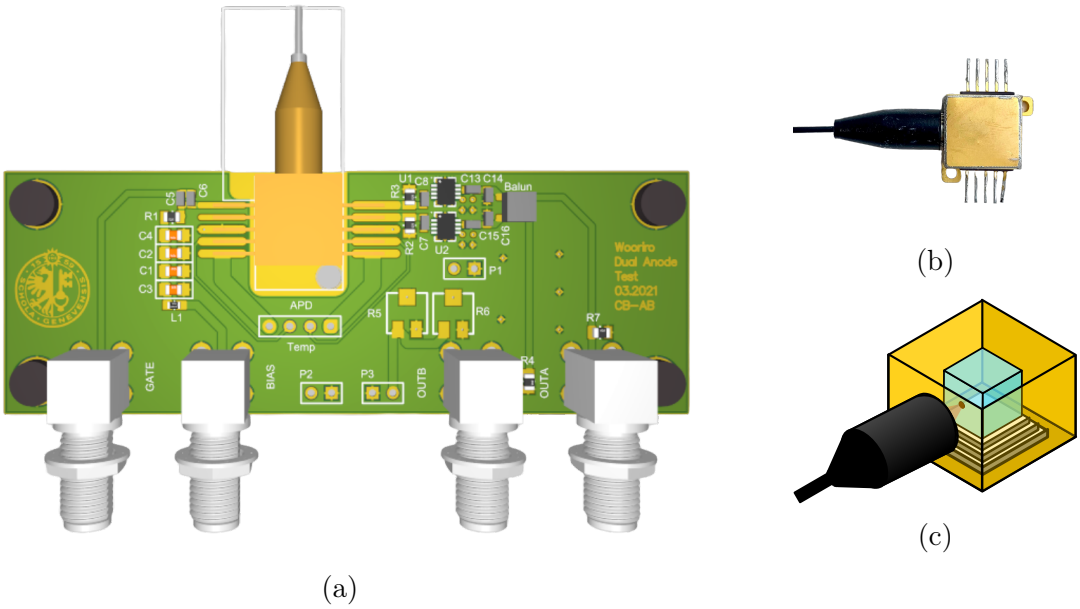


Figure 2.6: a) Schematic of the DA-SPAD detector. b) Capacitance Response mapped to applied gate. c) Avalanche signal superimposed with subtracted parasitic gate signal. d) Avalanche signal superimposed with parasitic gate signal.

2.2.2 Detector Control Electronics



(a)

(b)

(c)

Figure 2.7: a) In-house designed PCB to controll electronics. b) Detector package. c) inside view of detector package.

We designed a PCB in-house to control the detector (Figure 2.7a). To the common cathode, we applied voltage from an external source meter via an SubMiniature version A (SMA) connector. For the safety of the detector, the supplied signal was filtered by a typical Radio Frequency (RF) bias injection network design: Various capacitors in parallel to ground and a ferrite bead in series with the bias line. The capacitors were used for Direct Current (DC) blocking, RF bypassing, and decoupling. The values of the capacitors were chosen to allow for a low impedance path for RF signals while blocking DC signals. On the other hand, the ferrite bead was used for RF isolation - acting as a high-impedance element for RF frequencies while allowing DC bias current to pass through - preventing RF leakage and the bias line from acting as an antenna or creating unwanted coupling. The gating signal was also connected to the cathode via an SMA connector. In this line we implemented parallel capacitors bypassing for RF applications.

In practice, capacitors generally have some parasitic inductance that creates self-resonance. At frequencies above self-resonance, they behave more like inductors, so using two different capacitors helps maintain low impedance across a wider frequency range, as when one capacitor becomes inductive, the other can still provide capacitive bypassing. The larger capacitor handles the "bulk" bypassing for frequencies roughly

up to several hundred MHz, while the smaller one takes over at GHz range.

Each detector's anode was connected to a non-inverting amplifier with tunable gain (AD8353ACPZ, Analog Devices) to account for amplitude mismatches in the CR responses. The output of the two anodes was then combined at a 1:1 transmission line balun (ETC1-1-13, *MACOM*) before leaving the PCB.

Intuitively, one might expect the two outputs of the balun to be identical and symmetrical. However, this is often not the case. This can be due to manufacturing tolerances, where the balun's windings have small variations in inductance, coupling coefficient, and resistance. This can also be an issue from the PCB manufacturing, as trace width variations can affect impedance matching. Discontinuities in the routing to the Ground (GND) plane and coupling to neighbouring components may also differ between the two paths. It is also important to mention that the balun's performance varies with frequency as both phase and amplitude matching degrade at frequencies away from the designed center. For reference, a good balun can achieve an amplitude matching within 0.5dB and phase matching within 5 degrees.

During our preliminary tests, the DA-SPADs exhibited anomalous readings in counts that we correlated with the operation of nearby instruments. Hence, in order to mitigate unwanted electromagnetic interference picked up by the detectors, a Faraday cage was constructed. This aluminium shielding was enough to prevent spurious signals in the detection system.

When it came to gating the detectors, there were several options available. From the outset, for ease of implementation and increased performance, we had already decided to operate in gated mode and chosen sine-wave gating. Compared to square gating signals, a sinusoidal wave contains only a single fundamental frequency component. Since the capacitance response is proportional to the derivative of the applied bias, higher-frequency components in the gate waveform translate into sharper transients and higher CR. A pure sine gate therefore minimizes CR resulting in cleaner transients that are easier to suppress or subtract in subsequent electronics. At this stage, the two most critical parameters to consider are the amplitude and frequency of the gate signal. The gate amplitude is a critical parameter in the operation of SPADs, especially in high-speed applications as it directly influences several key aspects of detector performance:

Avalanche Probability: The gate amplitude is directly related to how much reverse bias voltage (excess bias V_{ex}) can be applied to the detector. The higher V_{ex}, the stronger the electric field across the SPAD. This increased electric field enhances the probability that a single photon-generated electron will trigger

an avalanche. Consequently, this ties directly to the detector's PDE [114].

Signal Strength: Following up in the previous item, the increase in V_{ex} is directly related to an increase in the avalanche gain. A stronger signal can be more easily distinguished from noise, improving the overall SNR of the system. However, here we face the first trade-off: Stronger avalanche signals also lead to more trapped carriers, increasing the APP, as well as a higher likelihood of spurious avalanches from thermal generation, increasing the DCR [114].

Dark Count Rate: Following up on the last topic, an increased electric field across the SPAD also increases the thermal generation of free carriers that can trigger an avalanche signal [114].

Timing Resolution: Another consequence of the increase in applied V_{ex} are its effects on avalanche build-up time. Higher voltages can lead to more rapid avalanche development, improving timing resolution and jitter [115].

Signal to Noise Ratio: Larger applied gate amplitudes also lead to larger CRs in the detector output. This is due to the capacitive coupling mechanism, where the gate signal capacitively couples to the output through the device's intrinsic capacitance. This capacitive coupling follows a linear relationship $I = C \times \frac{dV}{dt}$, where I is the coupled current. For larger gate signal amplitudes (higher V), one would get proportionally larger coupled currents and thus larger capacitive response signals at the output [116]. This increases the noise floor.

Since the CR at $20V_{pp}$ gating was only about 1.4 times higher than at $10V_{pp}$, we chose to operate the detectors at $20V_{pp}$.

This choice is justified by the non-linear relationship between the electric field strength and the avalanche triggering probability (P_{trig}). By increasing the excess bias, we enhance the impact ionization coefficient, ensuring that a higher fraction of photo-generated carriers contribute to a detectable signal. While DCR typically scales with V_{ex} due to field-enhanced tunneling, our characterization confirms that at 20 V_{pp}, the signal-to-noise improvement gained from the increased PDE and reduced timing jitter outweighs the marginal rise in thermal transitions. And although counterintuive, higher fields reduce the carrier residence time in the junction, potentially decreasing the probability of carriers being captured by deep-level traps in the lattice [117], leading to an improvement in the APP that we observe using a gate at 20 V_{pp}.

The second critical parameter was the gate frequency as it is directly related to the maximum speed at which the detector can operate. The higher the frequency, the shorter the gate duration, which in turn also reduces the time available for afterpulsing and dark counts to occur. This was beneficial as it allows for faster detection of single photons and reduces the dead time of the detector. However, it also required faster electronics to read out the avalanche signal before it decays below the noise floor. Another aspect to take into consideration is smaller available build up time that the avalanche has, as the the frequency increases. This requires, higher values of applied excess bias, V_{ex} to promote faster build-ups. The gate frequency was pushed to the maximum possible value of 1GHz, which was limited by the detector's inherent capacitance, capacitance mismatch between *dummy* and diode, the speed of the electronics used to read out the avalanche signal.

2.2.2.1 Capacitance Response Optimization

The CR was first looked into for each channel individually in a support PCB without the balun. The PCB contained only the detector control electronics. The analysis shown in Figure 2.9a revealed slight differences between channels, with the CR from the signal diode being larger than that from the dummy diode. This was expected, hence the amplifiers positioned before the balun were employed to provide compensation. The amplifiers must later be tuned to match the CR of both channels depending on the operation frequency.

In order to reduce the contribution of higher order harmonics from the sine wave generator to the CR, the gate signal was filtered by a low pass filter (LFCG-1200+ from Minicircuits) with a cut-off frequency at 1.2GHz, which achieved over $30dB$ suppression of higher order harmonics (see Fig. 2.8). This was enough to, at 1GHz gating, reduce the CR from 980mV to 840mV.

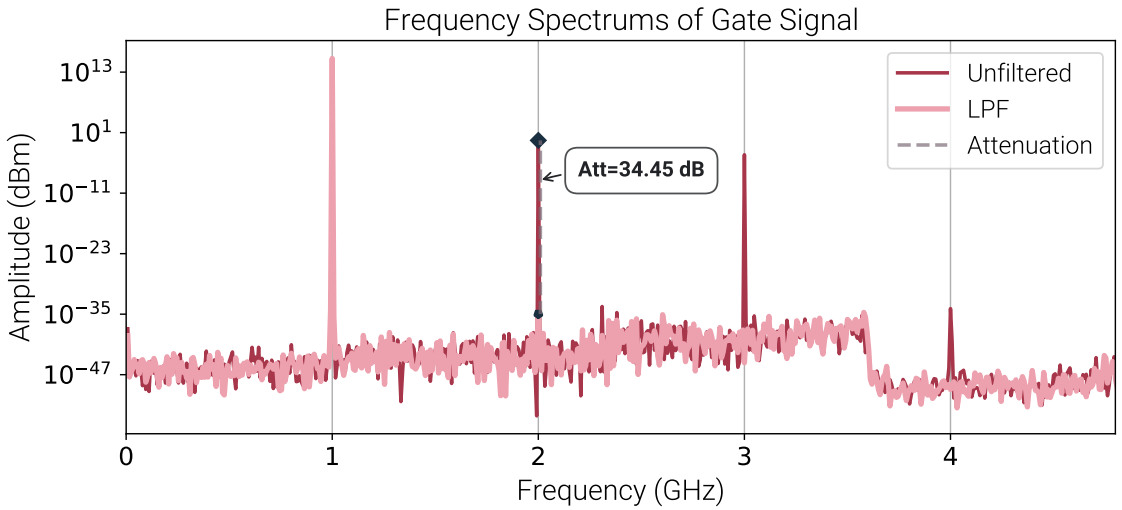
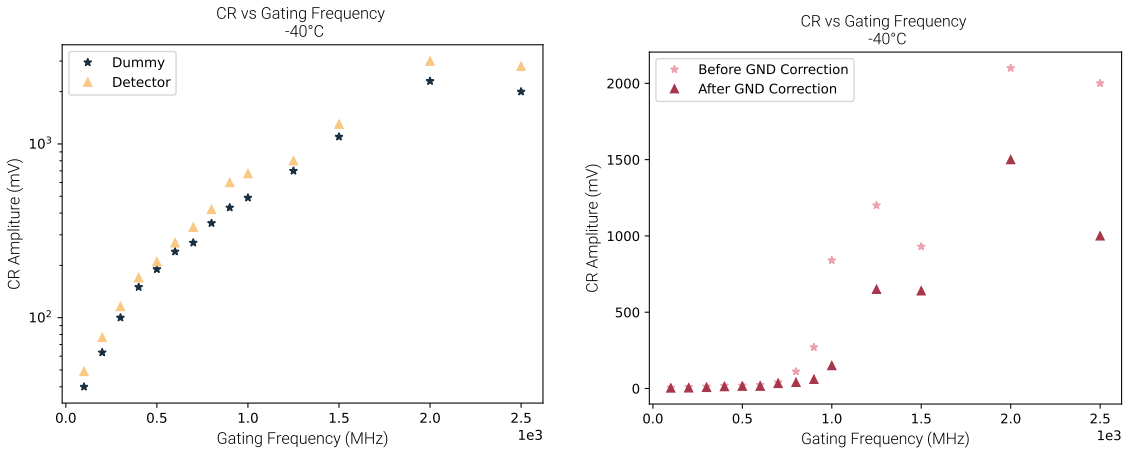


Figure 2.8: Spectrum of the gate signal before and after filtering. The detectors are gated with a sine wave generator (Agilent E4433B) at 1 GHz. The low pass filter (LFCG-1200+) suppresses higher order harmonics by over $30dB$ at 2GHz. Measured using a spectrum analyzer (Rohde & Schwarz FSW26).



(a) (b)

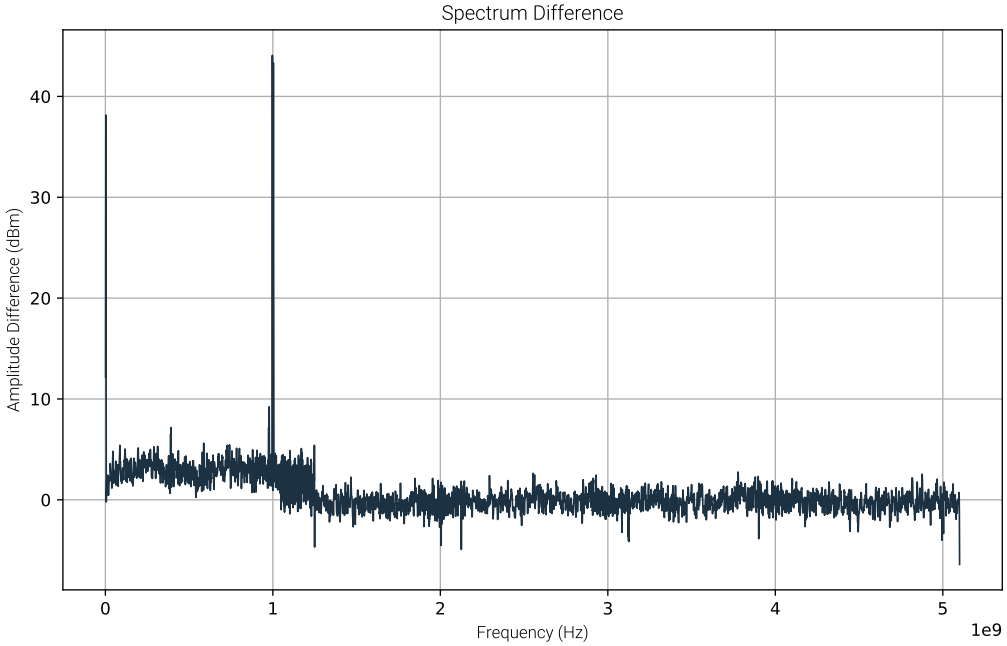
Figure 2.9: Capacitance Response (CR) of the DA-SPAD with a filtered gate signal. a) CR of the detector and dummy diode measured individually, before balun subtraction. b) Combined CR of the detector and dummy diode after balun subtraction. The two curves show the difference in CR with proper and improper grounding techniques.

In Figure 2.9b, the CR of the detector and dummy diode is shown after subtraction in the balun. Both curves show significant degradation at frequencies above 800MHz, where the CR spikes above 100mV_{pp}. A significant source of this noise was identified as interference from the power supply. To mitigate this issue, the

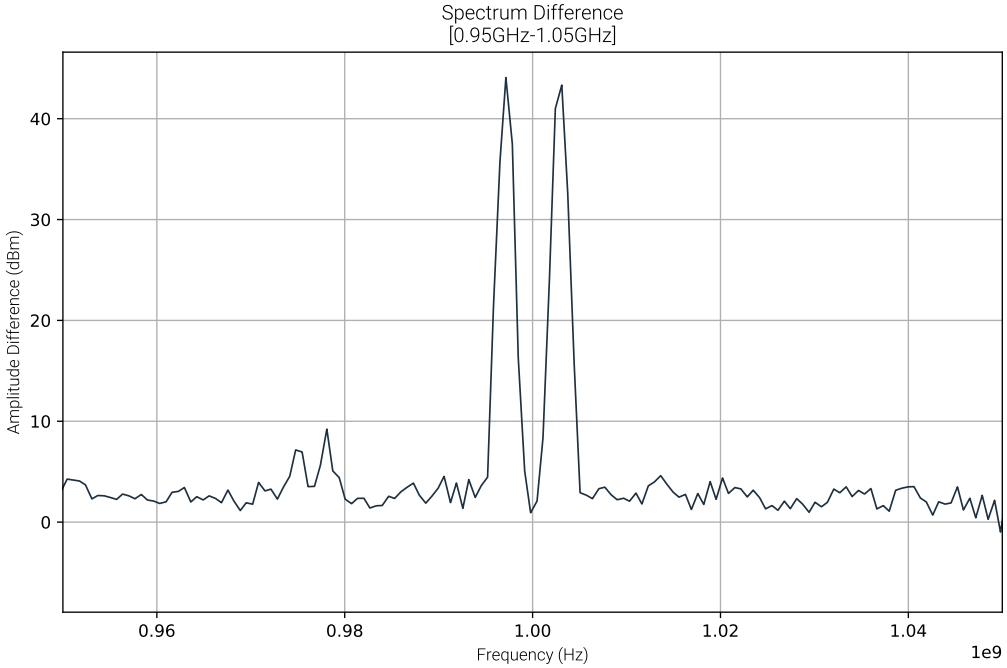
filtering capacitors had to be tuned and a cable ferrite bead was also added to the power supply's power supply line. These modifications resulted in a substantial 15dB attenuation of the CR at 1GHz.

Still unsatisfied with the CR response, we further investigated the cause of the poor performance at high frequencies. A critical discovery was the phase mismatch between the two balun outputs, which proved to be the main contributor to poor CR response at high frequencies. This was addressed by incorporating a tunable capacitor in series with the balun output, allowing for coarse phase adjustment and slight performance improvement. The combined optimizations successfully reduced the CR to levels sufficient for detecting avalanche events at 1 GHz frequencies.

To assess whether filtering could further improve the CR, we measured the frequency spectrum of the CR at 1GHz gating, above and below the avalanche threshold. The result of the subtraction of both frequency responses is shown in Figure 2.10. From this data it is evident that most of the avalanche signal is concentrated at the fundamental frequency of 1GHz and spread out through lower frequencies. This means that a notch filter at 1GHz would not be effective, as no notch filter is narrow enough to only filter the 1GHz component without also affecting the avalanche signal. The sudden drop in the spectrum at 1.2GHz is likely due to the attenuation caused by the limited bandwith of the amplifiers used pre-balun, as well as the limited bandwidth of the balun itself.



(a) Difference in amplitude between the two spectrums of the electrical signal generated by the DA-SPAD when biased below and above breakdown.



(b) Difference in amplitude between the two spectrums of the electrical signal generated by the DA-SPAD when biased below and above breakdown, with focus on the first order harmonic.

Figure 2.10: Difference in amplitudes between the two spectrums produced at the output of the DA-SPAD when biased below and above breakdown at 1 GHz gating frequency. Measured using a spectrum analyzer (Rohde & Schwarz FSW26).

The described system architecture demonstrated a maximum operational frequency of 1GHz. At 1.25GHz, the observed avalanche signals were notably smaller. This reduction in amplitude may be attributed not only to the shorter gate widths inherent at higher frequencies but also to the frequency-dependent attenuation of higher-order harmonics by the existing electronic components. The bandwidth of the current balun was limited to 3GHz, while the tunable amplifiers had a bandwidth of 2.7GHz. These likely contributed to the diminished signal at elevated frequencies.

In retrospect, the best solution moving forward would be to fully redesign the substraction circuitry given the considerable mismatches in both amplitude and phase of the two diodes. A more detailed analysis of this can be found in Section 2.3.1.

2.2.2.2 Avalanche Rise Time Measurement using Hot Carrier Luminescence

In order to study the maximum speed of the DA-SPADs, we used a technique based on hot carrier luminescence. This method allows to measure the spatial distribution of avalanche breakdown in the device. By measuring the rise time of the avalanche signal, one can estimate the maximum detection speed of the detection process.

Hot carrier luminescence is a phenomenon that happens typically in semiconductor devices, such as Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFET)s and SPADs, when high-energy carriers recombine and emit photons. This process was particularly relevant in our context of SPADs, where avalanche breakdowns create a large number of energetic carriers.

When a material absorbs photons, electrons can be excited to high energy levels in the conduction band, leaving behind holes in the valence band. These electrons and holes have excess kinetic energy and are called hot carriers. Normally, these hot carriers quickly lose their excess energy through phonon interactions (vibrations of the crystal lattice), a process known as cooling. However, in hot-carrier luminescence, these carriers recombine and emit photons before they have fully cooled down, resulting in a unique light emission called luminescence. The emiting spectrum can range from near-infrared to visible wavelengths, depending both on the semiconductor material, temperature and the applied electric field. This process is very fast, typically occurring on the order of picoseconds. This is particularly useful for studying the dynamics of avalanche breakdown in SPADs, as it allows for real-time observation of the avalanche process without disturbing it [118, 119].

More specifically, the luminescence arises from several physical processes involv-

ing the energetic carriers [118, 120]:

Direct Processes ($\Delta\vec{k}$=0): Which includes radiative transitions within conduction-to-conduction (c-c) bands; radiative transitions within valence-to-valence (v-v) bands; and conduction-to-valence (c-v) band recombinations.

Indirect Processes ($\Delta\vec{k} \neq 0$): These processes require momentum conservation via impurity assistance or phonon interaction. They can be further divided into radiative transitions within c-c bands; radiative transitions within v-v bands and c-v band recombinations. The intraband transitions can be direct or indirect, with the indirect transitions either phonon assisted, Phonon Assisted (PA) or Ionised Impurity Assisted (IA).

Bremsstrahlung Radiation: When hot carriers are decelerated by collisions with impurities or lattice defects, they emit electromagnetic radiation. The energy spectrum depends on the carrier velocities and scattering mechanisms.

To do hot carrier luminescence measurements, one must take into account several technical challenges. Firstly, hot carrier luminescence produces weak optical signals that require the use of SPADs or other single photon counting technologies. Another important factor is the temporal resolution of the detection systems. As mentioned, the luminesnce occurs on picosecond timescales, so the both the dector used and subsequence discriminating and time tagging electronics, must be able to capture these fast events. The measurement must also differentiate the faint hot carrier emission from various background sources including ambient light, thermal radiation, and other luminescence mechanisms within the device, this sets the requirement for appropriate filtering and shielding techniques.

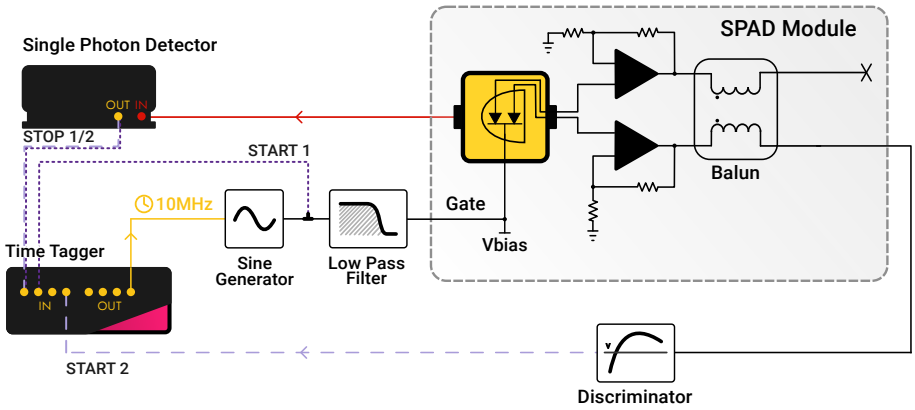


Figure 2.11: Experimental set-up for hot carrier luminescence measurements.

In Figure 2.11 is the set-up used to measure the hot carrier luminescence. The DA-SPAD was placed inside an insulating package to shield it from external electromagnetic interference. The light emitted by the DA-SPAD was collected by a standard optical fiber and connected to a comercial free running single-photon detector (ID220, ID Quantique) with 20% PDE and 10µs dead time.

The electrical part of the setup was comprised of a sine generator that provided the gate signal. After being filtered, the signal was sent to the detector. Outside the detector's control PCB, the generated signal was discriminated. The resulting Transistor-Transistor Logic (TTL) signal was then sent to a time-tagger (ID900, ID Quantique). The time-tagger also serves as the master clock, providing a 10MHz clock signal to the sine generator.

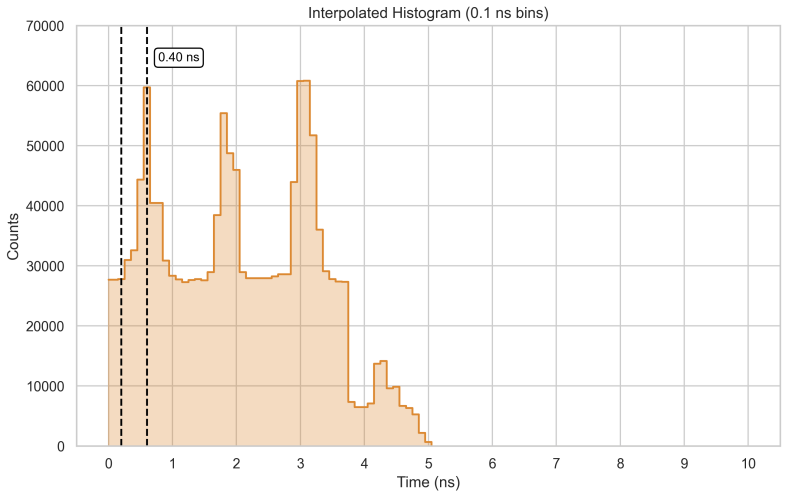


Figure 2.12: Measured time-resolved hot-carrier luminescence from a SPAD operated with a 1 GHz gate frequency. The histogram shows the temporal distribution of emitted photons corresponding to avalanche initiation and quenching cycles.

Two different methods were then employed to obtain the histogram. In scenario 1, the gate signal was used to trigger the start of the histogram, while in scenario 2, the output signal from the avalanche detection on the DA-SPAD was used to trigger the start of the histogram. In both cases, the stop signal was given by the hot carrier luminescence detection in the ID220. Both methods yielded similar results, as shown in Figure 2.12. Each peak corresponds to luminescence bursts generated during the avalanche buildup and quenching phases of the gating cycle. The periodic structure confirms synchronization with the gate frequency and provides insight into the temporal dynamics of avalanche initiation and extinction. Although the spacing between the avalanches is 1.2ns and 1.3ns for the first and second, and second and

third respectively, the measurement is still consistent with our set-up as this delay can be explained at the semiconductor level by an avalanche latency caused by the stochastic nature of the avalanche effects, or electronically by the electronic jitter and dead-time of the free-running detector (30µs), whose clock is not synchronized with the remaining of the system. The rise time of the avalanche signal was measured to be around 400ps. If we assume ideal quenching (disregard the fall time of the avalanche signal), this suggests that the maximum speed of these detectors is around 2.5GHz. While this is definitely an optimistic estimation, it does indicate that there is still room for improvement in the electronics to optimize the gating performance of these detectors.

2.2.3 Characterization of DA-SPADs

2.2.3.1 Quantifying Dark Count Rate, Photon Detection Efficiency, and Afterpulsing Probability

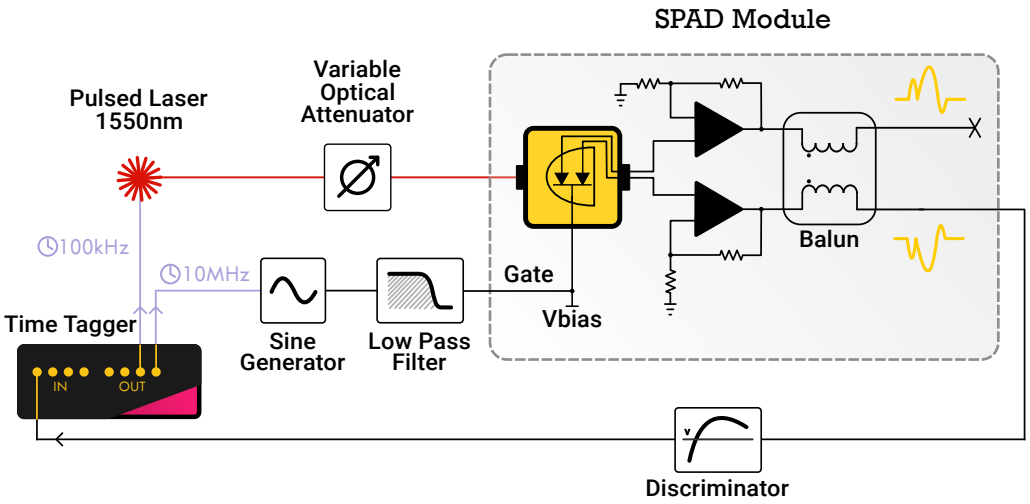


Figure 2.13: Schematic of the setup used to characterize SPADs performance.

Figure 2.13 is a schematic representation of the experimental setup used for characterizing the DA-SPADs. The detectors are cooled to the lowest possible temperature to minimize thermal noise, $-40°C$. The optical part of the set was very simple, and consisted of a pulsed 1550nm laser (PicoQuant - LDH-P-FA-1530/XL6) that was attenuated to sub-single photon level using a tunable optical attenuator from EXFO (FVA-60dB). This optical signal was subsequently sent to the detector to be used for characterisation.

The electrical part of the experimental setup was composed of a sine generator

(ESG-D Series Signal Generator, Agilent) that provided the gate signal at the desired frequency. As mentioned, this signal was filtered by a RF low pass filter to remove higher harmonics and then connected to the detector's PCB. The output signals of the detector were subtracted in a balun and sent to a in-house designed discriminator. The output TTL signal of this discriminator is sent to the time-tagger (ID900, *ID Quantique*), which again acted as a master clock for the entire characterization setup. It provided a 10MHz clock signal to the sine generator and a periodic trigger signal to the pulsed laser. The trigger delay has to be adjusted such that the laser pulses (<40ps Full Width at Half Maximum (FWHM)) occur entirely within the detector's active gate window ($\leq$ 300ps).

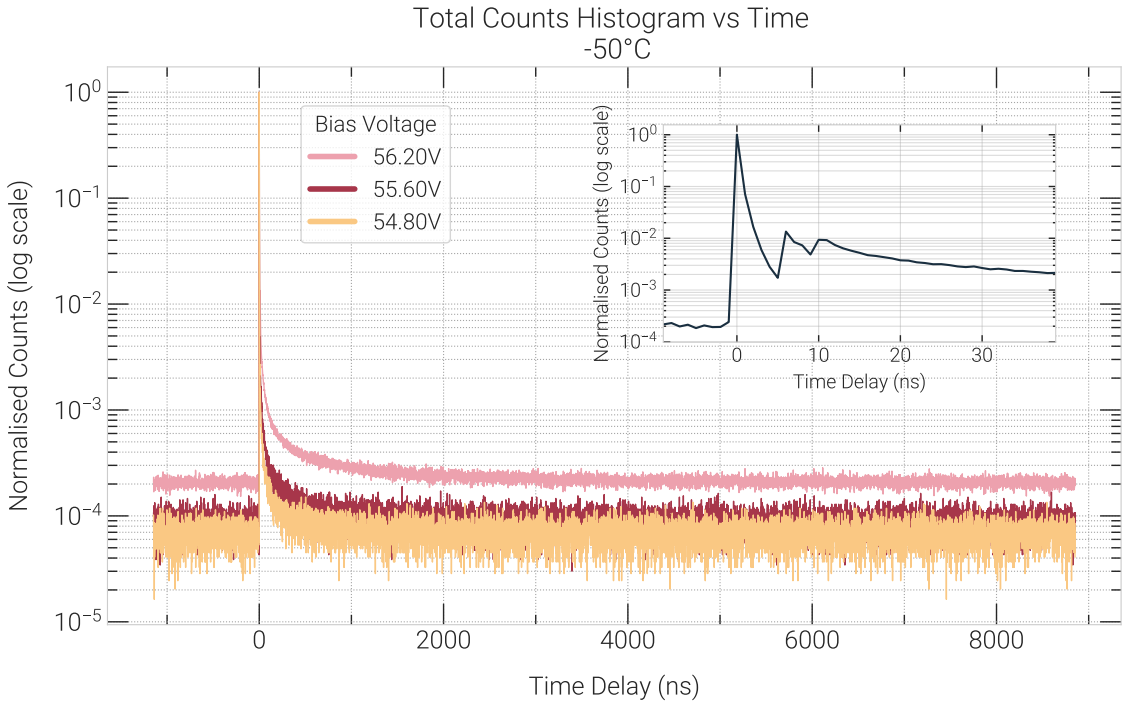


Figure 2.14: Histogram of the counts collected by the time tagger.

To measure the PDE as accurately as possible, we set the laser's repetition rate (f_{L}) to 100kHz. This low repetition rate provided enough time between laser pulses to mitigate afterpulsing effects. Without this delay, trapped carriers can trigger additional avalanches, leading to an overestimation of the PDE measurement.

The PDE, APP and DCR can be characterized by collecting a single histogram of time differences between time-tags, cyclically mapped to a domain corresponding to the pulsed laser's period ($\frac{1}{f_L}$), with a bin size of one gate period $\frac{1}{f_g}$. The PDE was retrieved from the number of counts in the bin that corresponds to the arrival of the laser's photons C_{L}. Since the laser was correctly synchronized and aligned

with the detector's gates, this bin was easily identifiable (see Figure 2.14). For the PDE calculation we also take into account the effects of the DCR in the counts of C_{L}. The PDE was then calculated according to 2.2, where Δt is the histogram's bin width, and $N_{\mathrm{L}} = \frac{1}{f_{\mathrm{L}}} \times \Delta T$ (where ΔT refers to the integration time in seconds).

$$PDE = \frac{C_{\mathrm{L}} - DCR \times \Delta t}{\mu' \times N_{\mathrm{L}}}. \tag{2.2}$$

For the PDE calculation, we only take into consideration the events where at least one photon was emitted in the laser pulse. Therefore, μ' given by 2.3, the corrected mean photon number, which takes into account the Poissonian emission statistics of the laser is considered for equation 2.2:

$$\mu' = 1 - \left.\frac{\mu^x e^{-\mu}}{x!}\right|_{x=0} = 1 - e^{-\mu}. \tag{2.3}$$

Two methods were attempted to calculate the DCR: by collecting a histogram in dark conditions (with the laser OFF) or by using the same histogram used for PDE calculation, in which case the laser ON. In the first scenario, the counts in the entire histogram were considered for the calculation of the DCR ($DCR = \frac{C_T}{\Delta T}$). Alternatively, for the second method, only the counts in the last 200 bins (200ns) of the PDE histogram were considered. We confirmed that both methods yielded the same results, which also indicated that we chose a sufficient low laser repetition rate to avoid afterpulsing effects.

The APP was calculated for various dead-time settings through post-processing analysis, rather than by actively quenching the detectors for a chosen hold-off time. We opted to do this for simplicity, since active and passive deadtime yielded the same results and have no differencing effect on the trapped carriers [114]. Two ways of implementing this in post processing could be either to: 1. Apply a time window after each detected photon, where we set all counts to zero for N microseconds to simulate the dead-time; 2. Tune the discrimination circuit to set the monostable's pulse length to N microseconds. For simplicity, since the specifications of the time-tagger allowed to acquire a long enough histogram ($1/f_L$) with sufficient precision ($1/f_g$), the first method was chosen. This approach achieved the same result as physically implementing different hold-off times in the discriminator circuit but offers two key advantages: it requires only one data set for all analyses and eliminates the need to modify or resolder the discriminator hardware. The APP was quantified using the relationship defined in equation 2.4, where C_{T} refers to the total counts collected over time t, and C_{L} is once again the total counts in the laser bin:

$$P_{\mathrm{ap}} = \frac{C_{\mathrm{T}} - C_{\mathrm{L}} - DCR \times \Delta T}{C_{\mathrm{L}} - DCR \times \Delta t} \tag{2.4}$$

In Fig. 2.14 is shown the histogram of detected photons, which reveals the temporal distribution of afterpulsing events. When photons were detected, we observed a sharp decrease in counts in subsequent 10ns. This pattern emerged from the discriminator's inherent dead-time, which is caused by the bandwidth limitations of its discrimination circuit (see Fig. 2.15). Each discrimination channel was made up of a comparator followed by a monostable that was responsible for producing a TTL signal compatible with most laboratory instruments. The monostable also had the added benefit of allowing for the implementation of a digital dead-time. By varying $\mathrm{C_t}$ and $\mathrm{R_t}$, the time constant can be tuned and consequently the dead time *t*. However, the monostable chip used allowed for a minimum pulse width of 10ns.

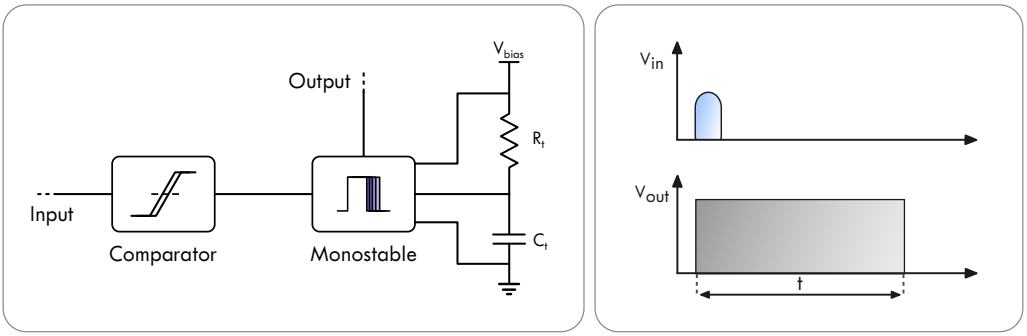


Figure 2.15: Schematic representation of discriminating circuit.

In the histogram one can also clearly see the effect of increased applied bias voltage on dark counts and afterpulsing events. As the bias voltage is increased, the dark count rate rises, leading to a higher baseline in the histogram. Simultaneously, the afterpulsing events become more pronounced, indicating a greater likelihood of these events occurring at higher bias voltages. And the exponential nature of the afterpulsing events is apparent. It is also worth mentioning that the count drop after the laser bin (Figure 2.14) is not a sharp cutoff but rather a gradual decrease with bumps. This happens because of how the monostable circuit works. If a trigger occurs while the RC capacitor is still recovering, it can sometime still produce a pulse, but with a reduced width.

2.2.3.2 Jitter Measurement

The jitter of the detectors was measured using the same experimental setup and tagging technique as in Fig.2.13. For the time tagging, we used a *ID1000*(*ID Quan-*

tique), which has a resolution of 1ps *high resolution mode*. A histogram was again acquired over 60s and analyzed to extract the jitter information.

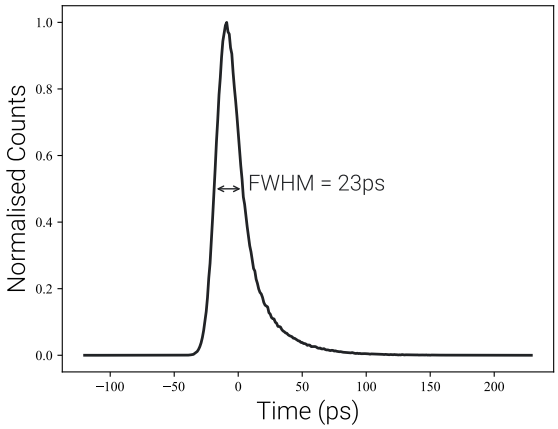


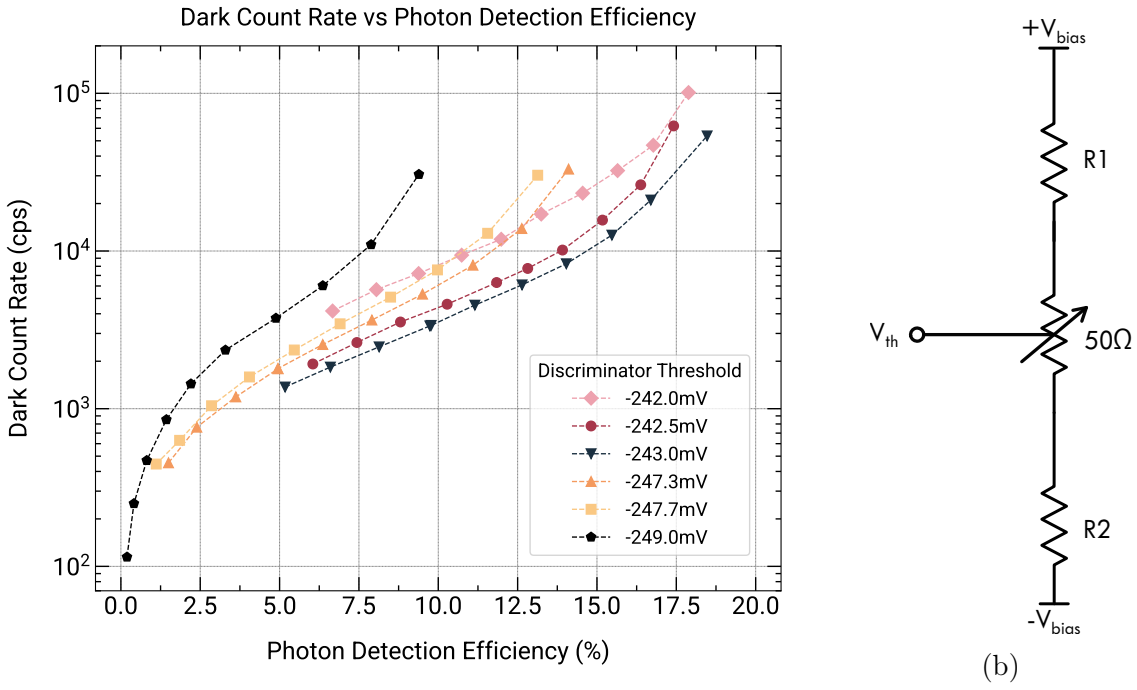
Figure 2.16: Jitter measurement results.

Fig. 2.16 shows the jitter measurement of the detector, including the contribution of the discriminator. The detector jitter is upper bound by the FWHM of the observed histogram peak as the measurement is a convolution of the detector response and laser pulse width ($\leq$30 ps). The measured jitter of 23ps at FWHM (Fig. 2.16) demonstrates that the operating conditions were well optimized. With a very narrow gate signal ($\leq$300ps) precisely aligned with the laser pulses, the photons are detected at the peak excess bias of the gate. This minimized the statistical spread in avalanche build-up times, leading to the reduced jitter. In general, a wider sine gate allows photons to arrive over a range of instantaneous biases (lower near the gate edges), which results in a spread of avalanche build-up times.

Additionally, cooling the detector to $-40°C$, alongside the dead-time applied, helps to reduce thermal noise and trap-related effects and mitigate dark counts and afterpulsing effects, which can contribute to the jitter as both form a uniform background that broadens the timing peak and raises jitter. Afterpulsing events in particular are the main contributors to the long tail present in jitter measurements [121].

2.2.3.3 External Discriminator Optimization

As briefly mentioned in the previous section, due to limitations in the minimum measurable pulse width of the time-tagger used, an external discriminator had to be used to allow the discrimination of the small avalanche signals output by the detector. The discriminator used was developed in-house, powered by 12V, and had



(a) (b)

Figure 2.17: a) Plot of dark count rates as a function of the photon detection efficiency as a function of the discriminator threshold voltage. b) Schematic representation of the discriminator circuit.

four falling edge discrimination channels. Each channel had an independent and tunable discrimination level controlled by a potentiometer (see Fig. 2.17b).

To allow greater precision in setting the discrimination voltage, the range of voltages could be arbitrarily tuned to any value, as long as it fell within the limiting values of [−5V,5V], as these are limited by the available power supplies in the discriminator board. This tuning range was determined by the circuit components. By shorting one of the potentiometer terminals to ground, the adjustment resolution of the discriminator voltage was further improved, since the output range was limited to [−5V,0V]. The resolution was determined by the potentiometer's mechanical range (number of turns), and can be further refined by adjusting the values of two resistors, R_1 and R_2, which set the maximum voltage through a voltage divider. We may solve for R_1 and R_2 via the following system of equations:

$$\begin{cases} V_{min} = R_1 \times I \\ V_{max} = (R_1 + 500\Omega) \times I \\ -V_0 = (R_1 + R_2 + 500\Omega) \times I \end{cases} . \tag{2.5}$$

where V_{min} and V_{max} are the minimum and maximum voltages of the desired range, V_0 is the voltage of the power supply, and I is the current flowing through the resistors.

The discrimination threshold is another important contributor to the performance of the detector. Naturally, we want to set it as low as possible to catch all small avalanches to improve the PDE. However, due to imperfections in the detector and associated electronic components, the measured signal is distorted in the gates following and prior to avalanche signal. This distortion manifests itself as a shift in the offset of the detector signal, as well as the appearance of additional peaks outside the gate window. This well illustrated in Fig.2.17b where the decrese in discrimination threshold leads to a progressive increase in DCR before hitting the CR wall, which would correspond to a $DCR = 1$Gcps.

The choice of discriminator threshold thus represents a delicate trade-off between maximizing sensitivity and suppressing spurious detections. Setting V_{th} too low increases the probability of registering noise-related avalanches, which artificially raises the DCR. On the other hand, a threshold that is too high risks rejecting genuine photon events, especially those generating weaker avalanches, due to statistical fluctuations in carrier multiplication or time of arrival in the gate, therefore reducing the effective PDE. In addition, this signal deformation can alter the apparent timing response of the detector, as distortions in the recovery baseline may shift the discriminator crossing point contributing to the jitter. Depending on the application, different optimal working points can be found from Fig. 2.17a, allowing the user to select a threshold that best balances detection efficiency and noise for their specific requirements. It is also important to note that the optimal threshold may depend on external factors such as temperature and electronic bandwidth, meaning that careful calibration is often required for each operating condition to guarantee reliable performance.

2.2.4 Applications

2.2.4.1 Heralded Single-Photon Sources

High-frequency gating SPADs can be adapted for asynchronous photon detection as demonstrated by Tosi et al. [122]. In this work, they proposed what they coined a gate-free running approach, that accomplishes an equivalent to free-running operations, while benefiting from short active times (< 500ps) that significantly limit APP. While in a standard gating approach the gate signal is synchronized with the

incoming optical signal, in the gate-free approach this synchronization is removed, allowing the SPADs to sample the incoming photons asynchronously. Over time this results in a uniform free-running like response. This is particularly useful in applications where the timing of photon arrival is unpredictable or when the source of photons does not emit at regular intervals. Count rates as high as 100Mcps, were reported even though such was achieved at the cost of a worse PDE performance (3% reported in the paper). The key advantage of this approach is that it can achieve extremely high count rates, making it suitable for applications requiring rapid photon detection, such as time-resolved spectroscopy and high-speed imaging. The short active time due to the gate signal ensures lower dark count rates and afterpulsing effects when compared to free-running SPADs [35, 108, 123, 124]. As mentioned, APP is the limiting factor for maximum count rates in SPADs, as it requires long hold-off times to be suppressed. Gate-free SPDs have been demonstrated across diverse applications, including 1.5GHz operation in 3D imaging systems and chirped amplitude modulation LiDAR. In 2023, Liang et al. reported a 1GHz device supporting count rates up to 500 Mcps, while Hagihara et al. applied the approach to compressive single-pixel imaging.

Using our SPAD device, we applied this fast-gating asynchronous detection principle to a Heralded Single Photon Source (HSPS). A HSPS is a device that probabilistically generates (ideally indistinguishable) single photons as a quantum resource. It usually relies on a nonlinear optical process, such as spontaneous parametric down-conversion or spontaneous four-wave mixing. These photons are highly time-correlated, and as a result, when one of them is detected (the heralding photon), it signals the presence of its partner photon. This allows the user to know that a single photon is available in the time frequency of interest [125]. Details on the theory of HSPS photon statistics and the methodology of our work may be found in the pre-print [126] (also attached at the end of this thesis).

Based on the data in Figure 2.17a, we found that a reasonable discrimination threshold of -243mV, which gave the best PDE performance for the same DCR. This yielded a workpoint that gave a PDE of 15.5% and a DCR of 12.6kcps. The detector was then integrated into the heralding arm of a HSPS operating at 1550nm. The heralded photons were detected by a SNSPD with a PDE of around 90% and a DCR of less than 20cps. The full experimental details can be found in [126].

The off-the-shelf, commercial nature of all HSPS components in this work as well as the DA-SPAD heralding detector demonstrates the potential for future, improved HSPSs in field-deployed quantum networks.

2.3 Discussion on Future Implementations

In principle, the dummy anode in a DA-SPAD should replicate only the parasitic capacitive behavior of the active junction, enabling clean subtraction of non-avalanche contributions. However, in practice, imperfect capacitance matching between the active and dummy diodes leads to residual capacitive artifacts in the output signal. These high-capacitance responses obscure the avalanche signal, degrade the achievable signal-to-noise ratio, and limit the effectiveness of the dual-anode concept. As a result, while the architecture holds promise for cleaner and faster single-photon detection, the current implementations still exhibit performance bottlenecks tied directly to capacitance mismatches.

2.3.1 Increasing the speed of DA-SPADs

A promising approach to overcome the capacitance mismatch and further increase the speed of DA-SPADs is to independently gate the two anodes with separate, tunable signals, as demonstrated by Scarcella et al. [106]. In this scheme, each anode would receive its own gate signal. The goal would be to pre-compensate the mismatches in the DA-SPADs capacitance by precisely adjusting the amplitude and phase of each gate signal. By carefully tuning these parameters, since the two diodes are already very similar, it would be possible to compensate for the intrinsic differences between the detector and *dummy* diodes, even frequencies above 1GHz.

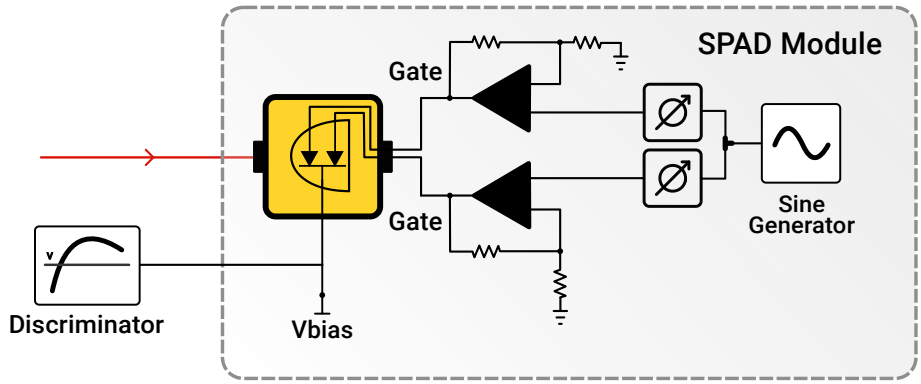


Figure 2.18: Possible implementation of the dual-gate scheme for DA-SPADs.

This implementation, shown in Fig. 2.18, involves generating two synchronized sine gate signals or splitting a common signal. Either a tunable amplifier of a RF variable attenuator should be placed in each path to allow for an independent and precise control of the amplitude. Following the amplitude matching stage, a phase tuning stage is required. This can be achieved using a RF phase shifter or a tunable

delay line. This order is important because most phase shifters are rated for lower power. The optimal settings would then be found by monitoring the detector's output in an oscilloscope or spectrum analyzer and trying to minimize the CR.

This dual-gate technique would enable a much higher suppression of the CR, even in the presence of significant device mismatch. As a result, the avalanche signal would become more distinguishable, allowing for operation at better performances at 1GHz and enabling work at higher gating frequencies, allowing for the integration of these detectors into our group's time-bin encoded QKD systems at 1.25GHz and 2.5GHz, that require gating frequencies at 2.5GHz and 5GHz respectively (discussed in the next chapter).

An additional advantage of this approach is that it aligns well with the capabilities of modern FPGAs (RFSoC ZCU216) that are now multi-channel with tunable high-speed outputs, which can provide both the amplitude and phase control. These devices would allow an easy integration of the gating electronics with the control and processing stages of a QKD system, without the need for the sine generator, amplifiers and phase shifters. In particular, this implementation could drive multiple detectors simultaneously at the multi-GHz gating frequencies demanded by time-bin encoded QKD. This would provide a scalable and compact solution for real-world deployments.

Chapter 3

Prototyping Integrated Quantum Key Distribution System

3.1 Introduction to Quantum Key Distribution

In an era where digital communication underpins nearly every facet of modern life, ensuring the confidentiality and integrity of transmitted information has become paramount. Traditional cryptographic methods, while effective, often rely on computational assumptions that may be vulnerable to future advancements in algorithms or the advent of quantum computing. In this context, QKD emerges as a revolutionary approach to secure communication, leveraging the principles of quantum mechanics to provide information theoretical security.

In brief, QKD is a process where distant parties exploit the physics of quantum measurements to establish a shared (symmetric) secret key for encrypting messages exchanged between them [127]. This is in contrast with conventional cryptographic schemes, such as the asymmetric Rivest-Shamir-Adleman (RSA) key distribution protocol, where the encryption security relies on the computational difficulty of certain mathematical problems (e.g. large prime factorization for RSA), such that the secret key needed for decryption cannot be easily derived from a publicly-known key used for encryption [10]. While these classical protocols have proven effective for most security applications thus far, they face the ever-growing threat of large-scale quantum computers that can efficiently solve classically difficult mathematical problems [9].

On the other hand, by also using quantum mechanics, QKD offers information-theoretic security: Using the OTP scheme, as long as the shared secret key is kept secure by each party after distribution, and used for encrypting messages only once,

the encrypted messages cannot be deciphered by eavesdroppers, regardless of their computational power [127]. Additionally, due to the sensitivity of quantum states to acts of measurement, any eavesdropping attempt during key distribution stage inevitably disturbs the in-transit quantum information and can therefore be detected with some probability [15]. Since the first protocol proposed in 1984 by Bennett and Brassard [15], numerous experimental demonstrations have validated the feasibility of various QKD protocols in both optical fibers and free-space channels, including metropolitan testbeds and satellite-based implementations.

Despite this progress, most QKD systems have so far relied on bulk optical components, which limit compactness, scalability, and long-term stability. Integrated photonic technology offers a path to overcoming these limitations by enabling the co-integration of quantum light sources, modulators, interferometers, and detectors on a single chip. This integration not only reduces footprint and cost but also enhances robustness and compatibility with existing telecommunication infrastructure.

In this chapter, we first introduce the BB84 QKD protocol and some of its practical simplifications (Section 3.2), before presenting the design and realization of a prototype integrated QKD system (Section 3.3). We conclude with a discussion on performance and perspectives for real-world deployment (Section 3.4).

3.2 The BB84 Protocol

Prepare and Measure (P&M) protocols represent one of the most widely implemented approaches to Quantum Key Distribution QKD. These protocols operate on a straightforward principle: the transmitter (traditionally named Alice) prepares quantum states and sends them over a quantum channel to the receiver (Bob), who then performs measurements on these states.

A standard P&M QKD protocol consists of the following steps [127]:

1. **State Preparation:** Alice prepares quantum states. Different encoding methods can be used in a P&M scenario, such as polarization, phase and time bin encoding.

2. **State Transmission:** The quantum states are transmitted over a Quantum Channel (QC), which can be optical fibers or free space.

3. **State Measurement:** Bob measures each received quantum state using randomly chosen measurement bases/settings.

4. **Basis Reconciliation:** After transmission, Alice and Bob communicate over an authenticated classical channel to reveal which preparation and measurement bases they used for each bit (but not the actual bit values). They discard all bits where their bases did not match, keeping only the results where they used compatible bases. This forms their *sifted key*.

5. **Parameter Estimation:** Alice and Bob select a small random subset of their sifted key to estimate potential eavesdropping by calculating the error rate (in quantum protocols denoted as the Quantum Bit Error Rate (QBER)). The QBER quantifies the proportion of key bits where Alice's and Bob's values differ. It reflects both the noise in the quantum channel and any potential eavesdropper interference. If the QBER exceeds a predetermined threshold, the protocol is aborted due to the suspected presence of an eavesdropper (traditionally called Eve).

6. **Error Correction:** When proceeding, they use error correction algorithms to fix any discrepancies in their exchanged keys, in an attempt to ensure both parties have identical keys.

7. **Privacy Amplification:** To ensure the final key is secure, Alice and Bob apply privacy amplification techniques to reduce the amount of information Eve might have gained during the exchange. This is typically formalized by the quantum leftover-hash lemma (Tomamichel et al. [128]). It states that if the reconciled key X has sufficient smooth min-entropy against Eve's side information, then applying a random two-universal hash function h to obtain a shorter key $K = h(X)$ will yield a key that is close to uniform and independent of Eve. The length of the final key must satisfy:

$$l \leq H_{\min}^{\epsilon'}(X|E) - 2\log\frac{1}{\epsilon - 2\epsilon'} \tag{3.1}$$

   ensuring that the joint state of the key and Eve's system is within ϵ of an ideal uniform key.

8. **Key Authentication:** Since the classical communication channel is assumed to be authenticated from the outset, the final step before usage is to use a portion of the already established and privacy-amplified key (K) to authenticate future classical messages(key refreshment).

9. **Key Usage:** The resulting secure key can then be used for encrypted com-

munication using conventional encryption methods like the One-Time Pad.

As mentioned, the above steps imply the use of an authentication scheme as Alice and Bob must ensure they are communicating with each other and not an impostor. Without authentication, the protocol is vulnerable to Man-in-the-Middle (MitM) attacks, where an adversary could intercept and impersonate both parties.

To maintain the information-theoretic security guaranteed by quantum mechanics, the authentication scheme itself must be information-theoretically secure. This is typically achieved using Wegman-Carter message authentication codes (MACs) based on universal hashing [129,130]. While this method requires a pre-shared secret key, it allows QKD to function as a key expansion protocol: the generated secret key is significantly longer than the key consumed for authentication.

Alternatively, classical digital signatures (e.g., RSA, ECDSA) could be used. However, these rely on computational hardness assumptions. Using them would downgrade the overall security of the QKD system from "unconditional" to merely "computational," potentially leaving the authentication vulnerable to future quantum adversaries [131].

The cornerstone of P&M protocols is the BB84 protocol. As mentioned, this pioneering protocol established the foundation for practical quantum cryptography and remains the most prominent implementation of QKD technology in use today.

In its first version, the BB84 protocol proposal encoded the quantum states using polarization degrees of freedom within single photons. It used four non-orthogonal states, which were represented by two bases: the horizontal/vertical **Z** basis ($|H\rangle$ and $|V\rangle$) and the diagonal/antidiagonal **X** basis ($|+\rangle$ and $|-\rangle$), where $|+\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|H\rangle - |V\rangle)$. The scheme worked as follows:

1. Alice generates a random bit string (using a QRNG for example) - this is going to be the basis for the secret key she will end up sharing with Bob.

2. Another random string, the same length as the key, is generated to assign a preparation basis (**Z** or **X**) to each bit.

3. Alice prepares the photon polarisation states according to the generated strings - where $|H\rangle$ and $|+\rangle$ correspond to bit 0 and $|V\rangle$ and $|-\rangle$ to bit 1 - and sends them to Bob over a quantum channel.

4. Bob randomly chooses a measurement basis (**Z** or **X**) for each received state and measures the quantum states.

5. Alice and Bob disclose over classical channel their respective preparation and measurement bases, discarding events where they have used different bases. At this stage, Bob will have to disclose the detection events as well.

6. To ensure no eavesdropper intercepted the key, Alice and Bob compare a subset of the generated keys over the public channel. They calculate the QBER, and if it is below a specific threshold (theoretically 11% in symmetric cases), they proceed. If the QBER is above the calculated threshold, they abort the protocol, assuming the channel is compromised.

7. Alice and Bob use classical error correction algorithms to correct any remaining errors in the key so in the end, they share identical bit strings.

8. Finally, they apply a hashing function to compress the key, reducing its length to eliminate any partial information an eavesdropper might have gained. The result of this process is the final, secure shared secret key.

3.2.1 Simplifications to the BB84 Protocol

3.2.1.1 Weak Coherent Pulses

A natural simplification of the BB84 protocol is the use of weak coherent light instead of single photons. This is an effective way to sidestep the shortcomings of current Single Photon Sources (SPSs), which include relatively low repetition rates and relatively high cost. In particular, SPSs are often not deterministic, meaning that they rely on a probabilistic signal that heralds the presence of a single photon in the mode of interest [132].

Using Weak Coherent Pulses (WCPs), the number of photons in each pulse follows the Poisson probability distribution. The probability of finding n photons in a pulse with a mean photon number μ is given by:

$$P(n) = \frac{\mu^n e^{-\mu}}{n!} \tag{3.2}$$

For a sufficiently low μ, most pulses will contain either 0 or 1 photon. These are called vacuum and single-photon pulses, respectively. While these pulses maintain the modeled security of a SPS, multiphoton pulses are vulnerable to Photon Number Splitting (PNS) attacks. In a PNS attack, Eve performs a photon number measurement on each pulse transmitted by Alice. For the multiphoton pulses, she extracts and stores one photon in a quantum memory while allowing the remaining

photons to continue to Bob. After Alice and Bob perform their basis reconciliation, Eve measures her stored photon using the correct basis. Critically, since the photons that reach Bob remain undisturbed, this attack generates no detectable errors while giving Eve full information on the measured bits.

A robust method for securing P&M protocols that use WCPs against PNS attacks is implementing decoy states. In this approach, Alice sends phase randomized WCPs with different intensities: μ_s (signal state), μ_1, and μ_2 (two decoy states). Bob then measures and records the detection rates corresponding to each intensity level, that can be accessed after parameter estimation/sifting. By analyzing and comparing these detection rates across the different intensity levels, Alice and Bob can accurately estimate the number of single-photon detections in their exchange. Previous works, such as by Wang et. al. [133], have demonstrated that this decoy state method provides sufficiently tight security bounds for QKD implementations, even in highly lossy channels, offering strong protection against PNS attacks.

3.2.1.2 1-decoy State Protocols

Also proposed in [133] was the 1-decoy state protocol as a simplified version of the decoy state method. In this approach, Alice sends only one signal state and one decoy state, which significantly reduces the complexity of the protocol while maintaining security against PNS attacks. Early attempts at analysing the protocol performance suggested its SKR was slightly below the 2-decoy protocol. However, these early comparisons did not account for statistical corrections required by finite-key implementations. In the work of Rusca et. al. [27], the authors demonstrated that the 1-decoy protocol could achieve a SKR comparable to the 2-decoy protocol when considering finite-key effects within shorter acquisition times.

While the performance difference between the two protocols might appear modest in theoretical terms, the practical implementation benefits can translate to significant experimental and economic advantages. The 1-decoy approach greatly simplifies the experimental setup and reduces complexity without compromising security, making it the preferred choice for many real-world QKD deployments. This finding is particularly valuable for experimental implementations where resources are limited and acquisition times are a critical consideration.

A further simplification of the BB84 protocol was proposed by Fung and Lo in [134]. This 3-state BB84 used only three states where the first two states $|0\rangle$ and $|1\rangle$ were used for the key exchange, and the third state $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ was used for security estimation.

This modification offers several practical advantages, including reduced hardware complexity (removes the need for a phase modulator as all states can be prepared solely using an Intensity Modulator (IM)), simplified state detection (allows for a fully passive Bob), and an adapted security analysis. On the other hand, as highlighted by the authors [134], this simplification comes at the cost of a slight reduction in the key generation rate, as the absence of the $|+\rangle$ state limits the amount of information that can be extracted for security estimation, which in turn makes the error estimation less accurate, therefore forcing more conservative bounds in the finite-key analysis.

Base	bit	State	μ_1	μ_2	μ_2
Z	0	$\|\varphi_0\rangle = \|\alpha\rangle_E + \|0\rangle_L$			
	1	$\|\varphi_1\rangle = \|0\rangle_E + \|\alpha\rangle_L$			
X	0	$\|\varphi_+\rangle = \frac{1}{\sqrt{2}}(\|\varphi_1\rangle + \|\varphi_0\rangle)$			
	1	$\|\varphi_-\rangle = \frac{1}{\sqrt{2}}(\|\varphi_1\rangle - \|\varphi_0\rangle)$			

(a) The standard BB84 protocol implementation with four states in two bases (**Z** and **X**) and two decoy intensity levels (μ_2 and μ_3).

Base	bit	State	μ_1	μ_2
Z	0	$\|\varphi_0\rangle = \|\alpha\rangle_E + \|0\rangle_L$		
	1	$\|\varphi_1\rangle = \|0\rangle_E + \|\alpha\rangle_L$		
X		$\|\varphi_+\rangle = \frac{1}{\sqrt{2}}(\|\varphi_1\rangle + \|\varphi_0\rangle)$		

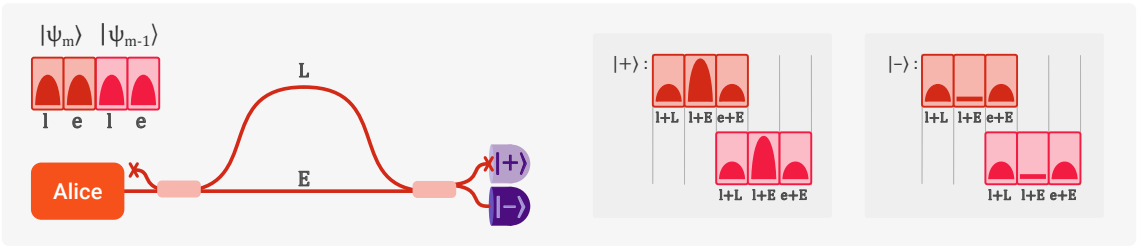
(b) The simplified 3-state BB84 protocol with a single decoy state (μ_2), showing the reduction from four to three states.

Figure 3.1: Comparison of state encoding schemes in BB84 protocol variants. Each state is represented by its quantum description and corresponding pulse intensity pattern.

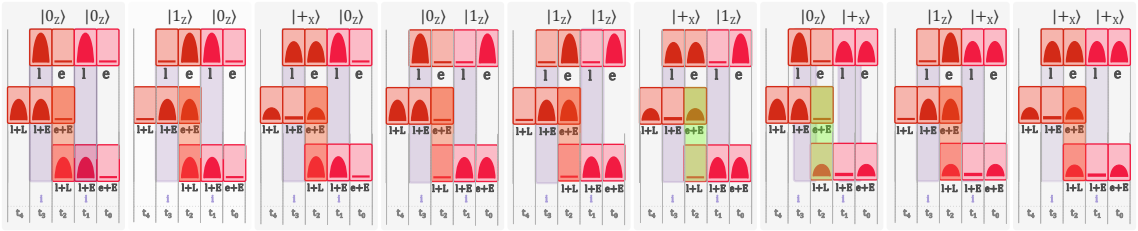
Figure 3.1 shows the differences in the state preparation between the standard and simplified implementation of the BB84 protocol mentioned in this subsection. The main distinction lies in the number of quantum states used for encoding information and the number of intensity levels that need to be encoded. Both have big implications in the experimental setup complexity, while maintaining a similar level of security.

3.2.1.3 2-detector 3-state BB84 Protocol

In previous works of our group, under Rusca et. al. [135] and Boaron et. al. [30], the authors proposed further simplifications to the 3-state BB84 protocol, by requiring Bob to measure only one output of the **X** basis interferometer. In this scenario, Alice sends only one monitoring state $|+_x\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and Bob measures only the state orthogonal to it $|-_x\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$.



(a) Interferometric measurement scheme showing how the incoming $|+\rangle$ state spreads across three time-bins, with constructive and destructive interference patterns at the output ports. Only the port with destructive interference is monitored.



(b) Comprehensive analysis of all possible two-qubit sequence combinations at the receiver, illustrating the critical temporal overlap between adjacent qubits. Green-highlighted sequences indicate the $|+\rangle$ states properly isolated by vacuum states, which are essential for accurate phase error rate estimation without the overestimation effects from overlapping time-bins.

Figure 3.2: Details of the **X** measurement basis in the 2-detector 3-state BB84 protocol implementation.

Just as in the time-bin approach to BB84, incoming states are randomly split into two arms for the base measurements. In the signal arm (**Z** basis) the states will go to a single-photon detector to measure the time of arrival of the photons. The states that are sent to the monitoring arm (refer to Figure 3.2a) and measured in the **X** basis will go through a second unbalanced Mach-Zehnder Interferometer (MZI) with the same delay as the one on Alice's side, spreading over a total of three time bins. The phase difference between Alice's and Bob's interferometers are fixed such that a detection in the central interfering time-bin corresponds to the state $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. However, unlike the 3-state BB84 proposed by [136] the photons projected onto the state $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ exit the interferometer via the port that in not connected to a detector.

With access to both $|-\rangle$ and $|+\rangle$ states, the phase error rate ϕ_z can be easily estimated:

$$\phi_z = \frac{errors}{total\ counts} \tag{3.3}$$

Where *errors* are the number of detections in the $|-\rangle$ state and *total counts*

are the total number of detections in both $|-\rangle$ and $|+\rangle$ states. ϕ_z quantifies the information gained by an eavesdropper on the key bits during the transmission. It's correct estimation is crucial for assessing the security of the key, and for determining the amount of privacy amplification required to ensure the highest possible SKR.

In the scenario of one **X** detector, the information obtained from monitoring $|-\rangle$ must then be used to reconstruct the events of $|+\rangle$. For a detailed analysis of the security of this protocol, we refer to the work of [137].

It is worth noting that since the interferometer delay is exactly corresponding to one time bin, the *early*(e) time bin of the state $|\psi_m\rangle$, overlaps with the *late*(l) time bin of the state $|\psi_{m-1}\rangle$ as seen in 3.2b. This overlap means that the central time bin of the interferometer's output will contain contributions from two consecutive states, $|\psi_{m-1}\rangle$ and $|\psi_m\rangle$. This temporal overlap can lead to interference effects that complicate the accurate estimation of the phase error rate ϕ_Z. Because of this, for the calculation of ϕ_z we must only consider the sent $|+\rangle$ states that are preceded and followed by a vacuum state, highlighted in green in Figure3.2b.

Now, as Bob only obtains partial information about the **X** basis statistics, this reduces the precision in the phase error rate estimation, since the missing outcome $|+_x\rangle$ must be inferred statistically. The achievable SKR is, as a consequence, slightly lower than when using two **X** basis detectors, with this reduction becoming more apparent in higher loss links [30, 135]. However, the practical advantages of this simplification, which include reduced system complexity, lower costs, and improved reliability, greatly outweigh the performance trade-off, making it a preferred choice for a real-world QKD implementation.

3.2.2 Time-bin Encoded 3-State BB84 with 1-Decoy State

The protocol used in our work was thus the natural culmination of all simplifications previously discussed. Using 1)weak coherent pulses, 2)1-decoy state approach, 3)the 3-state BB84 protocol, and 4)a 2-detector configuration, this time-bin encoded scheme emerged as an elegant solution that balances security and practical implementation concerns. This protocol was previously described in the works of [1,17,18,138]. Furthermore, all pulses were configured with random relative phase, effectively neutralizing coherent attack strategies. This was achieved naturally by using a gain-switched laser [139].

In brief, as depicted in Figure 3.3, in Alice's setup, a gain switched laser was used as a light source. These pulses then travelled through an unbalanced Michelson interferometer, where the time-bin quantum bits (qubits) are generated, with a

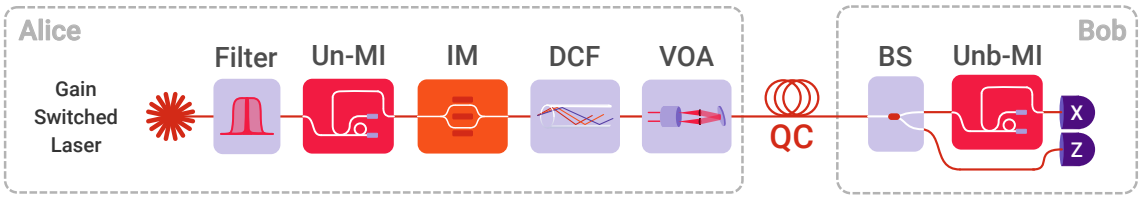


Figure 3.3: Experimental setup for a time-bin encoded BB84 protocol. Un-MI - unbalanced Michelson interferometer, IM - intensity modulator, VOA - variable optical attenuator, DCF - dispersion compensating fiber, BS - beamsplitter, X - monitor basis single-photon detector, Z - data basis single-photon detector.

$\frac{1}{2}\frac{1}{f_L}$ps delay - f_L is the pulsed laser's frequency. The interferometer features a phase shifter in one arm for phase adjustment.

The qubit state encoding was accomplished through an intensity modulator operating at $2f_L$. This modulator generated the four distinct intensity levels required to encode both the three quantum states and their corresponding decoy states (refer to 3.1b). The selection of qubit states and decoy levels occurred randomly, determined by external random number generators. Prior to transmission, the generated states passed through a variable optical attenuator, which precisely adjusted the mean photon number μ for both signal and decoy states. When performing QKD over long distances, the qubit states were transmitted through a dispersion compensating fiber spool, which has an opposite width dispersion response to normal telecom fibers, before leaving Alice. This critical component ensured that all states arrive at Bob's setup with identical temporal widths, effectively minimizing both the QBER and the phase error rate ϕ_Z

On Bob's side, the measurement basis selection occurred passively via a beamsplitter. As mentioned, for the **Z** basis measurements, used to generate the raw key, the qubits directly enter a single-photon detector that records the time of arrival. For the **X** basis measurements, used to estimate potential eavesdropper information, an unbalanced interferometer matching Alice's delay, measures the coherence between consecutive pulses.

The security proof of the protocol used was based on the work of [137] and the formula used for the SKR performance (after privacy amplification) was based on the security analysis of the 1-decoy state protocol [27].

$$\mathrm{SKR} = \frac{1}{t}\left[s_0 + s_1\left(1 - h\left(\phi_z\right)\right) - \lambda - 6\log_2\left(19/\epsilon_{\mathrm{sec}}\right) - \log_2\left(2/\epsilon_{\mathrm{corr}}\right)\right] \tag{3.4}$$

where t is the block acquisition time, s_0 and s_1 are the lower bounds on the num-

ber of vacuum events single-photon events in the Z basis, respectively, $h(\cdot)$ is the binary entropy function $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$, ϕ_z is the upper bound on the phase error rate, λ is the leakage of the bits during the error correction process, and ϵ_{sec} and ϵ_{corr} are the security (10^{-9}) and error correction (10^{-15}) parameters, respectively. This equation expresses the asymptotic and finite-key contributions to the SKR. $(s_0 + s_1\,(1 - h\,(\phi_z)))$ represents the extractable secret bits from the raw key, reduced by the entropy associated with the phase error rate. The subsequent terms subtract the information leaked during error correction (λ) and apply finite size corrections $(6\log_2\,(19/\epsilon_{\text{sec}}\,) - \log_2\,(2/\epsilon_{\text{corr}}\,))$.

Sifting

The sifting was done in real time by the Field-Programmable Gate Arrays (FPGAs) of Alice and Bob using the service channel. It could handle detection rates of up to 200 Mcps. The way the sifting was implemented required Alice to store her sent states in local memory until she receives the detection information from Bob. The internal memory of the FPGAs allowed us to store up to 2^{18} sent bits.

Error Correction

Since this implementation featured relatively low data rates, we used the *Cascade* error correction algorithm [140]. Cascade is an information reconciliation protocol specifically for key agreement. The aim of Cascade is to correct the errors that may exist between the two correlated sequences of bits generated by Alice and Bob after the quantum transmission phase. The protocol is round-based, where in each pass Alice and Bob partition their keys into blocks and exchange information on the block parity to detect errors in these blocks. The block size is calculated as a function of the QBER. The two parties perform a dichotomic search to locate and correct as many errors as possible while disclosing the minimum amount of information. A dichotomic search, similar to a binary search, works by first dividing the block into two halves. The parity of the first half is computed and exchanged. If the parities differ, it indicates that there is an odd number of errors in that half, prompting further division and parity checks until the erroneous bit is identified and corrected. If the parities match, it indicates an even number of errors (including zero), and no further action is taken on that half during that pass. In each following pass, the block size is doubled and the process is repeated. After each pass, the keys are shuffled to ensure that errors that may have been in the same block in one pass are likely to be in different blocks in the next pass. This shuffling increases the likelihood of detecting and correcting errors in subsequent passes. The process continues for a predetermined number of passes or until no errors are detected. Because errors

corrections in later passes (due to the shuffling) may flip bits that had been part of earlier-block partitions, one can change the parity of blocks from earlier passes, enabling further corrections in those earlier passes. This Cascade effect can then significantly improve the error-correction reach.

The inherent need of the protocol for multiple rounds of communication between Alice and Bob can introduce latency, especially for long-distance links. To mitigate this, one can send bulk parity requests (many blocks at once) and process blocks (and sub-blocks) in parallel, so that while waiting for a reply for one block, the protocol proceeds on others. However, this parallelization is still limited, deeming Cascade less suitable for very high-speed QKD systems which have higher data rates. The efficiency of the Cascade protocol is highly dependent on the initial QBER and the number of passes. More passes can improve error correction but also increase the amount of information disclosed to Eve, which can reduce the final key length after privacy amplification. Therefore, a balance must be struck between achieving a low QBER and minimizing information leakage.

3.3 Integrated Quantum Key Distribution System

3.3.1 Proof-of-Principle Prototype

In the work of Sax et. al. [1], the authors presented a proof of principle integrated QKD system based on the 2-detector 3-state BB84 protocol. Alice's prototype was fabricated using standard silicon photonic technology, allowing for easy integration with existing optical networks. Following previous works, Bob's setup was designed to be fully passive, relying on a beamsplitter for basis selection. This made silica (SiO_2) a prime choice for the receiver's optical platform, as it provides a low-loss and low-cost solution for the implementation of passive optical components. The system was tested over a 150km fiber link, demonstrating its feasibility for real-world applications.

As depicted in Figure 3.4, while considerable effort was made to integrate components into the PICs, several stand-alone external optical components were still necessary. These included passive optical components like a spectral filter, a Dispersion Compensating Fiber (DCF) spool, and a variable optical attenuator, alongside devices such as pulse generators, power supplies and custom control PCBs. These elements, while useful in the prototyping stage, limit portability as they are neither scalable nor practical for field deployment. Building upon the foundational work in [1], the next-generation QKD system presented in this thesis introduces several

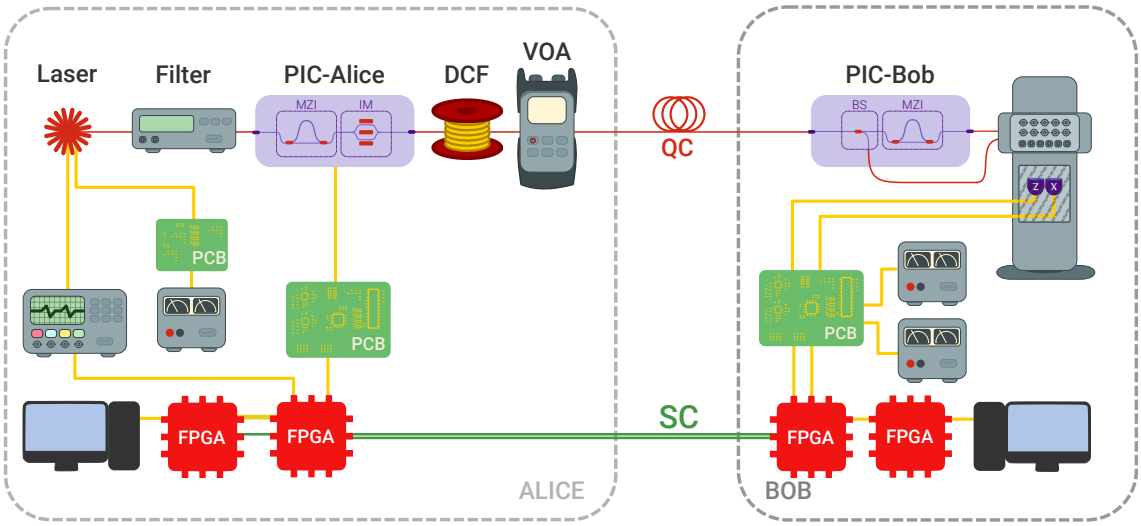


Figure 3.4: Set up used in the work of Sax et al. [1].The single-photon detectors used are SNSPDs and NFADs, both requiring either cryogenic cooling or a Stirling cooler. PIC - photonic integrated circuit, FPGA - field-programmable gate array, PCB - printed circuit board, DCF - dispersion compensating fiber, VOA - variable optical attenuator.

significant improvements aimed at enhancing practicality and performance.

3.3.2 Chromatic Dispersion in Fibers

In particular, an important component in the earlier QKD system was the DCF spool that mitigated the effects of chromatic dispersion. Chromatic dispersion refers to the phenomenon in optical fibers where different wavelengths of light travel at different speeds through a medium. This occurs because the refractive index of a material depends on the wavelength of the propagating light. If the refractive index decreases with increasing wavelength, the medium is said to induce *normal* dispersion. Conversely, *anomalous* dispersion occurs when the refractive index increases with increasing wavelength [141].

Hence, as laser pulses travel through an optical fiber, the wavelengths comprising the wave packet experience varying delays. This causes the pulse to broaden in time, potentially by a significant extent over long fiber distances as depicted in Figure 3.5a.

In the context of our QKD protocol, this temporal broadening effect can lead to adjacent laser pulses overlapping in time, as shown in Figure 3.5b. Naturally, such behavior is particularly problematic for time-bin encoded QKD systems. Given the example in Figure3.5b, if a state **Z,0** is sent, the temporal broadening can cause the *early* time bin to overlap with the *late*. This overlap can cause a trigger in the

late time bin as well which leads to the receiver interpreting it as **X** state instead of **Z,0**. Overlaps can also happen between neighboring qubits, not only within the same qubit, since the bin width is 400ps for each time bin and the repetition rate is 1.25GHz (800ps between consecutive qubits). Chromatic dispersion is therefore highly detrimental to the QBER of the QKD system.

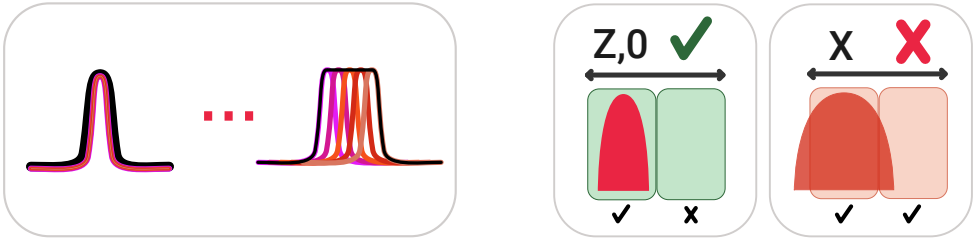
In [1], a large DCF spool was necessary to cancel out the dispersion within the QC and thus maintain the integrity of the time-bin encoded states. DCFs work by introducing large anomalous dispersion to counteract the normal dispersion of standard single-mode fibers. The key to selecting an appropriate DCF is to match its dispersion parameter to the accumulated dispersion of the transmission fiber, ensuring that the total dispersion over the entire link is minimized. In most applications, the dispersive effect may be suitably characterized by the *second-order dispersion* parameter D, defined as:

$$D = -\frac{\lambda}{c}\frac{\mathrm{d}^2 n(\lambda)}{\mathrm{d}\lambda^2} \tag{3.5}$$

where λ is the wavelength, c is the speed of light in vacuum, and $n(\lambda)$ is the wavelength-dependent refractive index of the fiber [142]. For a fourier limited pulse, the dispersion-induced temporal broadening can then be approximated as an addition in quadrature of the initial pulse width T_0 and the broadened width $T_b(z)$ after traveling a distance z along the fiber [143]:

$$T_1(z) \approx \sqrt{T_0^2 + T_b(z)^2} = T_0 \left[1 + \left(\frac{z}{L_D}\right)^2\right]^{\frac{1}{2}} \tag{3.6}$$

where T_1 is the final pulse FWHM and L_D is the *dispersion length*, defined using T_0 and D as



(a) (b)

Figure 3.5: Effects of chromatic dispersion on pulse propagation in optical fibers. (a) Temporal overlap of adjacent pulses due to chromatic dispersion, leading to increased error rates in time-bin encoded QKD systems. (b) Illustration of chromatic dispersion effects on a pulse traveling through an optical fiber.

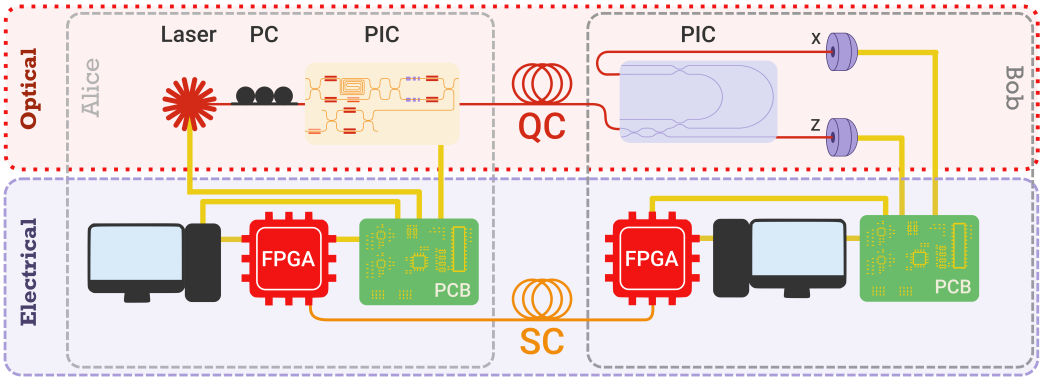


Figure 3.6: Schematic of the next-generation QKD system. The transmitter, Alice, and the Receiver, Bob, are delineated by gray dashed lines. There is a further separation between the optical (dashed red) and electrical (dashed purple) parts of the system. The optical components of Alice and Bob are connected via the quantum channel (QC), which in this case is fiber based. The FPGAs of Alice and Bob are connected via a classical communication channel, also denoted as service channel (SC). PC - polarization controller, PIC - photonic integrated circuit, FPGA - field-programmable gate array, PCB - printed circuit board.

$$L_D = \frac{T_0^2}{D}\frac{c}{\lambda_0^2} \tag{3.7}$$

With this mathematical model in mind, a well-calibrated DCF spool in the QKD transmitter can be highly effective in improving system performance. However, as was in the case of [1], this spool adds significant volume to the system and significantly reduces the system's versatility, as the spool length must be carefully adjusted based on the specific fiber link used. This is inconvenient for real world QKD deployments, where network infrastructures are heterogenous and may change over time. Any change in the network will necessitate a costly reconfiguration of the QKD system and make scaling challenging. We addressed this issue by making design changes that eliminated the need for a DCF spool, as detailed in the next section.

3.3.3 Overview of the Next-Generation QKD System

With our next-generation design, several upgrades were made with a focus on deployability and scalability of the system. The changes can be broadly categorized as:

Physical Size: Alice and Bob's setups now fit within a standard 19in 1U and 19in 3U rack respectively. The former includes a gain-switched laser, a custom PCB, the transmitter integrated chip, a computer and an FPGA. The latter features the receiver integrated chip, two NFADs, control PCBs, a computer and an FPGA.

Optics: Alice's addition to the setup include an on-chip VOA and a narrow-band spectral filter. Bob's setup now has a tunable beamsplitter for optimized basis selection.

Electronics: Both Alice and Bob's setups are now supported by custom-configured electronic boards, computers and FPGAs.

Alice's PCB can be segmented into five main subsystems: the power supply, the laser control, the PIC, the microcontroller, and the FPGA interface:

- The power distribution to the entire system is managed through a comprehensive multi-rail power supply system that converts an initial 12V input into the various voltages required by all PCB components, the PIC, the laser and the microcontroller.
- The laser control subsystem incorporates a 7-pin SMA packaged laser with a tunable bias supply voltage, and a standard Thermoelectric Cooler (TEC) controller (Thorlabs MTD415T) used to regulate the laser's temperature. The RF driving signal coming from the FPGA(Kintex UltraScale) is amplified before being sent to the laser. This amplifier is mounted on a custom-designed *mezzanine* PCB that is placed atop the *host* PCB. The mezzanine features one input and one output SMA connector for the RF signal and a variable gain amplifier. The tunable amplifier allows for the optimization of the optical pulse's amplitude and width characteristics.
- The microcontroller stage serves as the central control hub, supplying current to the PIC heaters while interfacing with the EIC to provide the control signals needed for the correct parameter settings in the intensity modulator. Additionally, the microcontroller handles the monitoring of all PIC components, ensuring optimal system operation and performance feedback.
- The FPGA interface stage converts the differential signals from the FPGA to single-ended SMA output. This generates both a laser clock and an

auxiliary clock. The laser clock features a programmable delay of up to 100ps in steps of 3ps. This allows us to tune the alignment between the laser pulses and the IM modulation pattern.

Bob's custom PCB is relatively simpler and was specifically designed to serve as a tunable and stable current source to control the receiver beamsplitter ratio. The detector control PCB used is a system from ID Quantique (IDQ) that provides the necessary bias voltage temperature control and detection readout. This is connected to a computer via USB and controlled using a custom software interface. It includes a multi-rail power supply system that converts a 12V and 5V input (provided by computer) into the various voltages required by all PCB components, and the detectors.

Detectors: Bob now features Peltier-cooled NFADs, which have a substantially smaller footprint compared to the previously used SNSPDs and fridge-cooled NFADs. Another alternative could have been to update the stirling cooler to a more compact version, however, the Peltier-cooled NFADs were chosen due to them being commercially available, further demonstrating the practicality of the system.

The evolution of the system design can be seen by comparing Figure 3.4 to Figure 3.6. The simplified representation in Figure 3.6 shows the deliberate design choices to reduce the system size, complexity, and power requirements, with a focus on integration and miniaturization, thereby enabling deployment outside of laboratory environments.

3.3.4 Laser Driver and Pulse Generation

The laser used in this system was a high bandwidth distributed feedback laser from the Gooch & Housego's AA0701 series. The laser operated at a wavelength of 1550.12nm at 25°C. It was mounted on the host PCB, which provided the necessary power and temperature control signals.

To drive the laser in a gain-switched mode, a RF signal was required to modulate the laser's output. As previously described, the FPGA interfacing PCB subsystem generated this signal, giving us a repetition rate of 1.25GHz at an amplitude of 530mV. This subsystem also included a variable gain amplifier, which allowed for the adjustment of the pulse amplitude before it was sent to the laser. This amplifier was crucial for controlling the width of the generated pulses, essential for the performance of the time-bin encoding BB84 protocol.

However, as the gain provided by the Operational Amplifier (OP-AMP) was not sufficient to drive the laser directly, a constant current needed to be supplied to bias the laser. The bias voltage was set to -0.950V. The RF signal was then amplified using an RF amplifier and sent to the laser via a 5cm coaxial cable. After optimization of the cable length, signal voltage and gain, the laser generated pulses with a FWHM width of about 40ps.

In general, it is important to minimize the coaxial cable length to mitigate the effects of the cable's parasitic capacitance and inductance. Longer cables can introduce signal attenuation and distort the sharp rising and falling edges needed for clean gain switching due to dispersion effects in coaxial cable. Also non-negligible are the effects of reflections and ringing caused by impedance mismatches along the cable path. This can cause unwanted secondary pulses or poor extinction ratios in the laser output due to ringing and overshoot generated in the driving signal. We observed a further reduction in the laser pulse width of about 4ps just by reducing the SMA length from 15cm to 5cm. This temporal profile was captured using the Hewlett Packard 54750A Digitizing Oscilloscope with Hewlett Packard 83482A Optical/Electrical Module.

Lastly, the laser's pigtailed output fiber was connected to the PIC via an Multi-fiber Termination Push-on (MTP)-16Angled Physical Contact (APC) harness. The MTP is a high-density fiber optic connector designed to terminate multiple optical fibers in a single connector housing, commonly used in high-density network applications. All 16 fibers in the harness can be used as either transmit or receive fibers, allowing for flexible configurations. In our context, the first twelve fibers were not connected. The following three fibers were used to couple light into the chip: One bypassed the MZI, which allowed us to use the Coherent One-Way (COW) protocol, and the other two coupled light in through the MZI with one of them going through the ring filter as well. The last fiber in the harness was used to couple light out of the chip.

3.3.4.1 Mitigating Chromatic Dispersion Without DCF

As mentioned in 3.3.2, we implemented certain design changes to avoid using DCF spools. For example, the added ring filter narrowed the laser's output spectrum down to 0.170nm from 0.207nm. However, the most important change was in fact halving the laser's repetition rate by modifying the FPGA bitstream to generate a 1.25GHz clock and a random state sequence at 2.5GHz. The PIC also had to be redesigned to operate at this new frequency. We achieved this by doubling the delay

of the unbalanced interferometer, responsible for generating the qubits, from 200ps to 400ps.

By doing so, the time between consecutive time bins was increased from 200ps to 400ps, while the laser pulse width remained unchanged at 40ps. If correctly centered in the 400ps modulation time window, this would give the laser pulses more room to spread out before they start to overlap with the neighbouring time bins, as illustrated in Figure 3.5a. If we consider the FWHM width, we may allow for 180ps for the pulse to spread out in one direction before it starts to overlap with the neighbouring time bin, compared to the previous 80ps. The temporal profiles were captured using the Hewlett Packard 54750A Digitizing Oscilloscope with Hewlett Packard 83482A Optical/Electrical Module, and the spectral widths were measured using an Anritsu MS9740A Optical Spectrum Analyzer.

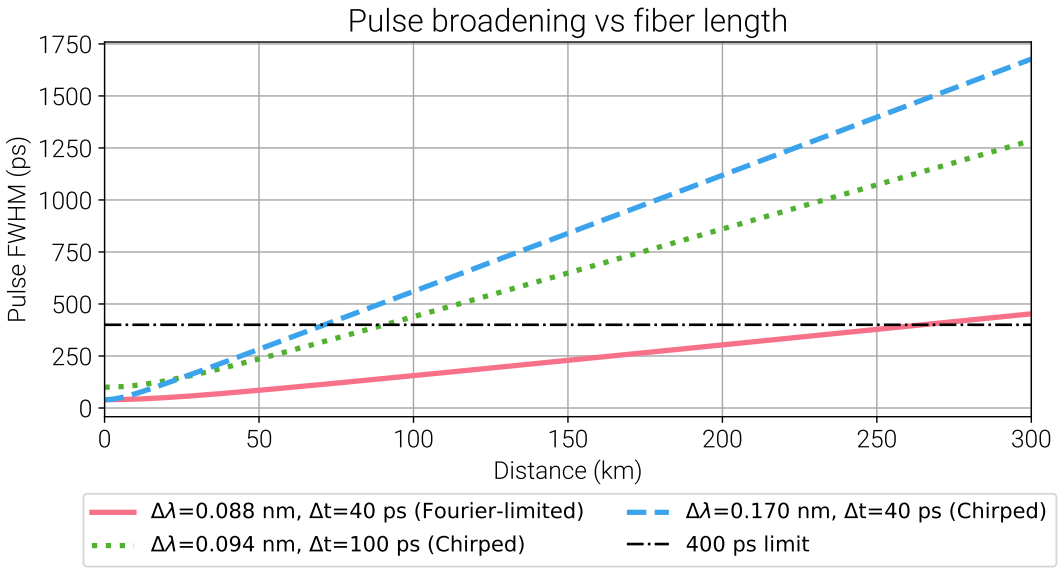


Figure 3.7: Simulation of pulse broadening due to chromatic dispersion over various fiber lengths, assuming different initial pulse widths and spectral widths. The horizontal black line indicates the 400ps threshold, beyond which overlap between adjacent time bins occurs, leading to increased error rates in time-bin encoded QKD systems. The labels () indicate the function used for each curve, with Gaussian pulses using Equation 3.6 and chirped Gaussian pulses using Equation 3.8.

Running a quick simulation using Equation 3.6, assuming a Fourier-limited pulse of 40ps FWHM and taking the dispersion parameter as that of standard single-mode fiber at 17.0 ps/(nm·km) [144], results in the solid line of Figure 3.7. We see that at 50km the pulse would have spread in time to about 85ps at FWHM, 150ps at 100km, and 300ps at 200km, all still within the 400ps time bin. This adjustment

cuts the transmission rate by half, but at the same time also mitigates the relative impact of chromatic dispersion, since the pulses are given twice as much temporal spacing to broaden without overlapping neighboring bins.

However, in practice, it is generally difficult to achieve short Fourier-limited pulses, and indeed we observed that our laser system introduced unwanted frequency chirp modulation that caused the measured spectral width to be larger than the Fourier-limited value for a given pulse width. We can account for this by modifying Equation 3.6 as

$$T_1(z) = T_0 \left[\left(1 + \left(\frac{C \cdot z}{L_D} \right) \right)^2 + \left(\frac{z}{L_D} \right)^2 \right]^{\frac{1}{2}} \tag{3.8}$$

where we introduce a chirp parameter C that can be determined from the measured spectral and temporal widths via [143]

$$2\pi \frac{c}{{\lambda_0}^2} \Delta\lambda = \left(1 + C^2\right)^{\frac{1}{2}} T_0 \tag{3.9}$$

with $\Delta\lambda$ being the wavelength FWHM of the pulse. By using Equation 3.8 with the parameter C for our actual laser parameters (40ps pulse width and 0.170nm spectral width), we obtain the dashed line in Figure 3.7. We now see that at 100km the system performance will be significantly hindered by the increased pulse width, as the pulse will now be considerably wider than the time bin itself (Figure 3.5a). In such a situation, the high QBER values may cause key generation to be unfeasible.

In order to extend the maximum working distance of the system given the actual characteristics of the laser, we decided to further narrow the laser's spectrum by increasing its time width. This was achieved by tuning the laser's bias current and the RF driving voltage. By doing so, we were able to reduce the spectral width to 0.098nm, at the cost of increasing the pulse width to 100ps at FWHM (dotted line in Figure 3.7). This improvement made working distances of up to 100km without the need for a DCF spool more attainable as shown in later sections.

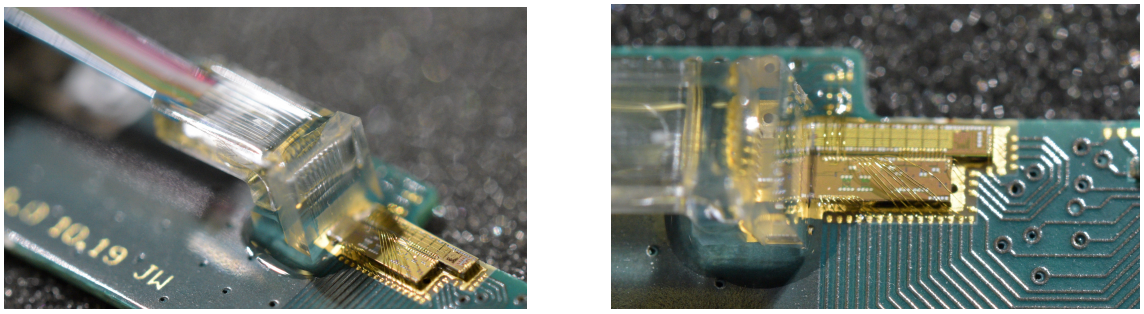
3.3.5 Integrated Transmitter (Alice)

The integrated transmitter was based on the design presented in [1]. Both chips, the PIC and the Electric Integrated Circuit (EIC), were fabricated using standard silicon photonic technology by Sicoya GmbH.

The sizes of the PIC and the EIC were 4.50mm × 1.10mm and 4.50mm × 0.75mm. Such a small footprint was achieved by the use of silicon as the main material for the PIC, as silicon has a high refractive index which allows for the fabrication of

compact waveguides and photonic components. Si also benefits from mature fabrication processes enabling high-precision and high-density integration of photonic components on a single chip, integration with electronic components, and mass production capabilities. The EIC was designed to interface with the PIC and provide the necessary control signals for the transmitter, including power supplies, amplifiers and Digital-to-Analog Converters (DACs).

As shown in Figure 3.8b, the PIC and EIC were mounted on a custom PCB, designed in-house, that acted as an interposer for the electrical connections and provided support for the 90-degree fiber array scheme used for optical coupling (Figure 3.8a).



(a) (b)

Figure 3.8: Pictures of the integrated transmitter components. (a) shows the 90-degree fiber array used for optical coupling into and out of the PIC. (b) shows the PIC and EIC mounted on the interposer PCB, which provides electrical connections and mechanical support for the fiber array.

This interposer PCB was glued, and all electrical connections were bonded to the larger mezzanine PCB as mentioned in Section 3.3.3. The mezzanine interfaces with the controlling electronics via a PCIe X4 connector.

Refering to Figure 3.9, the PIC was composed of a grating coupler, a ring filter,an unbalanced Mach-Zehnder interferometer, an intensity modulator and a variable attenuator. The grating coupler was used to couple light from the fiber into the chip, while the MZI was used to generate the time-bin qubits. The IM was responsible for the state encoding and the VOA attenuates the light sent out of Alice to the desired mean photon number.

3.3.5.1 Coupling to the Photonic Integrated Circuit

PICs require efficient fiber-to-chip coupling to enable hybrid integrated architectures as well as to minimize off-chip component power requirements for example a non-integrated laser like in our scenario. The core size mismatch between optical fibers

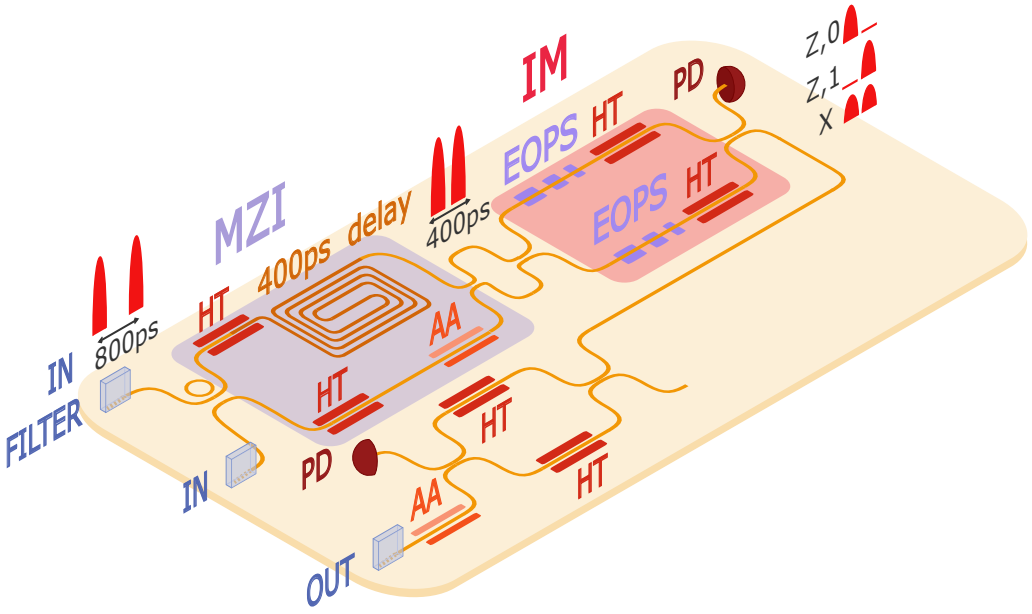


Figure 3.9: Schematic of the photonic integrated circuit showing the optical signal path. An input signal with a repetition rate of 1.25 GHz is processed through a Mach-Zehnder interferometer (MZI), which doubles the repetition rate to 2.5 GHz. Coherent neighboring pulses in the output stream encode a qubit through their temporal relationship. These qubits are then modulated by the intensity modulator (IM) to generate the three measurement basis states Z_0, Z_1, and X. HT - Heaters; EOPS - electro-optic phase shifters; MZI - Mach-Zehnder interferometer; IM - intensity modulator; AA - absorption attenuator, PD - photodetectors.

(8µm) and Si waveguides (sub-micron) presents significant coupling challenges. At the moment, two primary coupling strategies exist: edge couplers, which guide light in-plane to the chip facet; and grating couplers, which diffract light out-of-plane from the chip surface. Edge couplers offer broad bandwidth and polarization insensitivity but require lensed or Ultra-high Numerical Aperture (UHNA) fibers. On the other hand, while grating couplers are polarization-sensitive, they enable denser integration, wafer-scale testing due to their ability to provide out-of-plane access to PICs. Grating coupler are also compatible with standard single-mode fibers [145].

In this work, we use a grating coupler with a 90-degree fiber array to couple light into and out of the PIC. The coupler consists of a periodic pattern etched into the surface of the waveguide, which varies the waveguide's refractive index profile. This refractive index variation can either be periodic or apodized (often referred to "chirped" as well). Focusing on the periodic structure, as shown in Figure 3.10, the periodic refractive index variation is created through the etching of the Si waveguide. These trenches have a depth of e, length L_E and unetched length L_O. Λ refers to the period of the grating, which is the distance between two consecutive trenches $\Lambda = L_E + L_O$. The effective refractive index of the grating is given by the equation [146]:

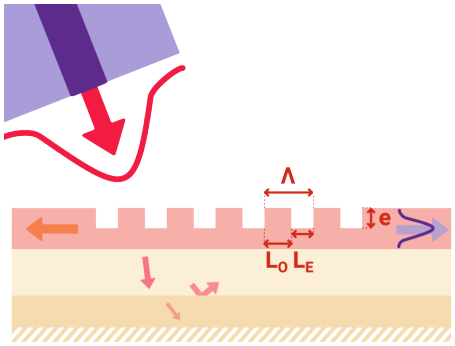


Figure 3.10: Schematic of the grating coupler used in the PIC. The grating coupler consists of a periodic pattern etched into the surface of the waveguide, which varies the waveguide's refractive index profile.

$$n_{\text{eff}} = \text{FF} \times n_O + (1 - \text{FF}) \times n_E \tag{3.10}$$

where FF is the fill factor, defined as the ratio of the etched area to the total area of the grating $\text{FF} = \frac{L_O}{\Lambda}$, n_O is the refractive index of the unetched region, and n_E is the refractive index of the etched region. The fill factor can be adjusted by changing the etching depth and width of the trenches. If we describe the behaviour of the grating in terms of the Bragg condition, the effective index of the grating can be approximated by the equation:

$$n_{\text{eff}} - n_{Si} \times \sin(\theta_1) = \frac{\lambda}{2\Lambda} \tag{3.11}$$

From equation 3.11 we can see that for a given wavelength λ, the incoming light is diffracted at an angle θ_1 that depends on geometrical parameters of the grating coupler, such as the period Λ and FF. It is also worth noting that, although not explicitly stated in Equation 3.10, the etch depth e also affect the effective index of the grating, as it has a direct impact on the refractive index of the etched region n_E.

Silicon-On-Insulator (SOI) waveguides exhibit high birefringence, making it difficult to efficiently couple both orthogonal polarization states from a single-mode fiber. Consequently, grating couplers are typically optimized for single polarization operation. As a result, to couple into our PIC we adjusted the polarization of the laser light with manual polarization control paddles to maximize the coupling efficiency.

3.3.5.2 Multimode Interference Beamsplitters

The chip employed Multimode Interference (MMI)-based couplers to divide the optical signal. This multimode interference phenomenon has become increasingly popular in recent years for applications including power splitting, polarization beamsplitting, and Wavelength Division Multiplexing. MMI allows for multi input and output configurations, making it a versatile choice for integrated photonic circuits. The advantages of MMI include lower insertion loss, wider bandwidth, and lower sensitivity to fabrication errors compared to other splitting methods like directional couplers [147].

MMI-based Beam Splitters (BSs) are composed of a narrow input waveguide, a wider central region where multimode interference occurs (MMI region) and multiple output waveguides placed at locations where the self-images form.

The dimensions of the MMI can be tuned based on the light's wavelength and the refractive index contrast between the waveguide and the medium, to achieve the desired number of output ports. One can generate a single image (mirror of the input), two identical copies (for a 1×2 splitter), or more, if designed accordingly.

A coherent light wave, typically one mode from a single-mode waveguide, is fed into the wider multimode section, which supports multiple spatial modes simultaneously due of its larger width. The input mode excites several transverse modes in the multimode region. Each of these modes propagates with a different effective refractive index, and accumulate different phase shifts over the same length. As the modes propagate, they interfere constructively and destructively with one another. At specific propagation lengths - beat lengths - the interference pattern recreates

the input field or a symmetric version of it. This phenomenon is called self-imaging. At the self-imaging locations, output waveguides are placed.

These types of BSs are easy to make polarization-independent, as they rely on the interference of multiple modes rather than the polarization state of the light. This is particularly advantageous in integrated photonic circuits, where maintaining polarization stability can be challenging due to fabrication imperfections and environmental factors.

3.3.5.3 Ring Resonator Filter

A ring resonator on chip, in a filtering configuration, is a wavelength-selective optical cavity coupled to two straight bus waveguides, an input bus and an output *drop* port. These configuration is referred to as *Add-Drop*. The light that propagates in the input bus couples evanescently into the ring only when the round-trip phase shift is an integer multiple of 2π. This resonance condition is satisfied at specific resonant wavelengths (λ_{res}):

$$m\lambda_{res} = n_{eff}L \quad (m \in \mathbb{Z})$$

where n_{eff} is the effective refractive index of the waveguide, L is the ring circumference, and m is the mode number.

The ring dimensions are engineered such that the resonance condition is met, in our case, at 1549.7nm. The optical energy of the pulse interferes constructively within the ring, building up high intracavity intensity, and is subsequently coupled out to the *drop* port, which is connected to the following components of the chip. The incoming signals at off-resonant wavelengths experience destructive interference within the cavity and bypass the ring, remaining in the input bus.

Physically, the device is realized in SOI technology. The high index contrast of silicon allows for sharp bends (radii down to 3µm), which in turn, increases the Free Spectral range (FSR) of the resonator. The FSR is defined as the wavelength spacing between adjacent resonances and is given by:

$$FSR = \frac{\lambda^2}{n_g L} \tag{3.12}$$

where n_g is the group index of the waveguide mode, L is the ring circumference, and λ is the wavelength of light. n_g can be derived from the effective refractive index n_{eff} as $n_g = n_{eff} - \lambda\frac{dn_{eff}}{d\lambda}$. A larger FSR allows for better isolation of individual wavelength channels, a feature that is is particularly useful in Dense Wave Division

Multiplexing (DWDM) systems. [148, 149].

3.3.5.4 Unbalanced Mach-Zehnder Interferometer and Thermo-Optic Phase Shifters

The aforementioned beamsplitters were fundamental components of the unbalanced MZI used to generate the time-bin qubits. The unbalanced MZI consisted of an input 1×2 BS, two arms - which are simply two waveguides - one longer than the other by 400ps, this introduced a phase difference between the two arms. The two arms were then recombined at an output 2×1BS.

To control the phase of the unbalanced MZI, we used Thermo-Optic Phase Shifters (TOPS) in both arms of the interferometer. TOPSs relied on the thermo-optic effect to modulate the refractive index of the waveguide, thereby controlling the phase of the light propagating through it. The relationship between the phase shift and the temperature change is given by the equation [150]:

$$\Delta\varphi = \frac{2\pi}{\lambda}\left(\frac{dn}{dT}\right)\Delta T L \tag{3.13}$$

where λ is the wavelength of the propagating light, L is the length of the TOPS, and ΔT is the change in temperature, and $\frac{dn}{dT}$ is the thermo-optic coefficient. The thermo-optic coefficient represents the change in refractive index per change in temperature. In silicon, this coefficient is approximately 1.87×10^4 at a wavelength of 1550nm [151].

Since the temperature rise happens in a resistive heater due to power dissipation, we can write it as the following:

$$\Delta T = \frac{\eta \mathrm{P}}{\mathrm{C}_p \rho \mathrm{LS}} \tag{3.14}$$

here η is the efficiency of the heater, P is the power consumed by the heater, C_p is the heat capacity of silicon, ρ is the density of silicon, and S is the cross section of the heated region. The product $\mathrm{C}_p\rho$LS therefore stands for the thermal capacity of the heated region.

To compensate for losses in the long arm of the interferometer and to improve its visibility, a VOA was placed in the short arm of the MZI. Carrier injection VOAs, also known as Electronically Variable Optical Attenuatorss (EVOAs) exploit the free-carrier absorption effect in silicon to provide electrically controllable optical attenuation.

The optical properties of silicon are strongly perturbed by this carrier injection,

which modifies the material's refractive index through free-carrier interactions - an electro-optic effect that, while typically performance-degrading in photodetectors, is deliberately harnessed in EVOAs for attenuation. In these devices, a p-i-n junction is integrated across the silicon waveguide, where forward bias injection introduces free carriers into the undoped intrinsic region - the waveguide.

Since the free-carrier effect strength is approximately proportional to the square of wavelength, these attenuators become increasingly efficient at longer wavelengths, enabling reduced device footprints and lower power consumption compared to shorter wavelength operation. [152]

3.3.5.5 Intensity Modulator and High-Speed Phase Shifters

In order to achieve high-speed power modulation, the typical workaround is to use phase modulators in the two arms of a MZI. By introducing tunable phase shifts in both arms, the recombination of light produces an interference pattern that enables amplitude modulation.

The thermo-optic phase shifters previously mentioned were reported to only achieve modulation speeds up to 357kHz [153]. This made the technology unsuitable for the GHz range amplitude modulation we required.

The phase shifters for intensity modulation in this work were thus based on the plasma dispersion effect. In our particular scenario, this electro-refractive effect was achieved by injecting free carriers into the waveguide via a P-N junction, formed across the waveguide. This enables the injection of free carriers when the junction is placed under forward bias. These carriers modify the refractive index of the waveguide, thereby modulating the phase of the propagating light.

The high diffusion capacitance of the free-carriers (10pF) allows for high-speed modulation, allowing these modulators to achieve speeds of up to 70GHz as reported in [154].

This capacitance effect may seem counterintuive at first, as high capacitance is typically associated with slower electronic devices due to RC-limited bandwidth. However, the diffusion capacitance reflects how many carriers are injected in a P-N juntion during forward bias. Therefore, a higher diffusion capacitance translates to a stronger plasma dispersion effect, leading to greater phase modulation. Consequently, this allowed for shorter device lengths and lower driving voltages [155].

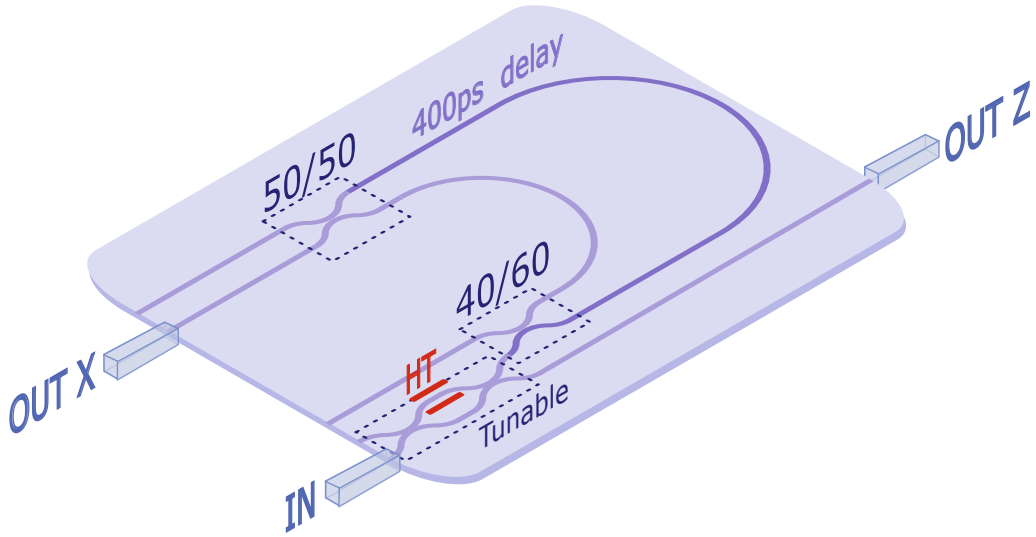


Figure 3.11: Schematic of the receiver's photonic integrated circuit showing the optical signal path. The incoming quantum states are split by a polarization-independent tunable beamsplitter (BS) into two arms, one for the **Z** basis and one for the **X** basis. The **Z** basis arm is connected to a single-photon detector, while the **X** basis arm is connected first to a fully passive unbalanced MZI. HT - heater.

3.3.6 Integrated Receiver (Bob)

As previously mentioned, it is crucial to have a low loss receiver in order to achieve high key generation rates and not sacrifice the maximum transmission distance. With this loss budget in mind, similarly to the work of [1], the receiver was designed in silica in collaboration with the group of Prof. Roberto Osellame in Politecnico di Milano. A silica platform was chosen for its low propagation loss (<0.05 dBcm) and coupling loss (<0.1 dB). Our new chip design not only had a different MZI imbalance to match our transmitter, but also features a tunable beamsplitter to optimize the basis selection based on the transmission distance. This tunable beamsplitter was a key innovation that allowed for better performance over varying distances, enhancing the overall efficiency of the QKD system, making one system suitable for different practical applications.

Another important requirement for the receiver PIC is to be polarization-independent. This polarization independence requirement must hold for both the MZI - meaning the visibility should be, ideally, 100% for any polarization state of the incoming light - and the basis selection beamsplitter, which should not introduce any polarization bias in the basis selection.

In Figure 3.11 we can see the schematic of the receiver's PIC. The PIC is composed of a polarization-independent tunable coupler and an unbalanced MZI. The

tunable coupler is used to divide the optical signal into two arms, one for the **Z** basis and one for the **X** basis. The unbalanced MZI is used to measure the coherence between consecutive pulses in the **X** basis arm. The light is butt coupled in and out of the chip.

The receiver PIC was fabricated using a Femtosecond Laser Micromachining (FLM) technique described in [156]. FLM is a well-established technique for the fabrication of integrated photonic devices, such as waveguides, in transparent materials like silica. The femtosecond laser is focused onto the bulk of the transparent material to induce nonlinear absorption processes, resulting in a localized refractive index change around the focal point that creates a waveguide structure. In this direct-writing process, waveguides are formed by translating the substrate at constant speed relative to the laser focus along the desired three-dimensional paths, enabling highly precise fabrication of complex geometries with sub-micrometer resolution. The optical performance of the resulting waveguides depends significantly on irradiation parameters and substrate properties. Typical waveguides fabricated by FLM demonstrate refractive index changes of 10^2 to 10^3 between core and cladding, providing good connectivity with standard optical fibers [157]. The substrate of our receiver PIC was an aluminum borosilicate glass (EAGLE XG, from Corning Inc.). The PIC's waveguides had a loss of 0.2 dB/cm and a birefringence of $< 3 \cdot 10^{-5}$ at 1550nm.

3.3.6.1 Polarization-Independent Couplers

The directional couplers were fabricated using the methodology described in the work of Corrieli et al. [158], namely multiscan waveguide inscription followed by thermal annealing. Inscribing waveguides next to each other to form a coupler inadvertently changes the optical properties of each waveguide. Therefore certain fabrication parameters could be optimized to produce the desired performance in the coupler. For example, the dependence of the coupling ratio on the input polarization may be minimized by controlling the separation between the two waveguides in the interaction region. An auxiliary track was also inscribed beside the coupler waveguides to remove birefringence mismatch in the coupler. The annealing process is used to reduce the birefringence of the waveguides, induced during the laser writing process. This is achieved by heating the substrate to a temperature below its glass transition temperature ($750°C$), allowing the material to relax and reduce internal stresses that contribute to birefringence. The annealing process also improves the overall optical quality of the waveguides by reducing loss and improving mode

confinement.

3.3.6.2 Visibility Characterization of the Mach-Zehnder Interferometer

The birefringence of the waveguides is a critical parameter for the performance of the MZI, as it can lead to polarization-dependent phase shifts and consequently a reduced visibility [141]. Birefringence in the waveguides leads to a polarization dependent phase shift:

$$\Delta\psi = \frac{2\pi}{\lambda}\Delta n_{eff} L \tag{3.15}$$

where Δn_{eff} is the birefringence, L is the propagation length and λ is the wavelength of the light. This phase shift can lead to a reduction in the visibility of the MZI, as the interference pattern will be affected by the polarization state of the incoming light. For two linearly polarized modes that accumulate a relative phase $\Delta\psi$ due to birefringence, the visibility V of the MZI can be expressed as:

$$V = |\cos(\frac{\Delta\psi}{2})| \tag{3.16}$$

Thus, if $\Delta\psi = 0$, the visibility V is unity, and for orthogonal polarization states ($\Delta\psi = \pi$), the visibility V drops to 0. To mitigate this effect, compensation tracks were inscribed around the waveguide of the longer arm of the MZI to balance the birefringence between the two arms [158, 159]. This is optimized for the specific wavelength of operation, the length of the MZI and the temperature of operation. In Figure 3.12 we can see the characterization of the MZI's visibility at the poorest setting. This means that the least optimal polarization state (the one most affected by the birefringence) was injected into the PIC and the visibility was measured. The polarization state was adjusted using a paddle polarization controller placed before the receiver PIC. The visibility was calculated by measuring the maximum and minimum power at one output of the MZI while sweeping the temperature at which the PIC was stabilized. The temperature was swept using the TEC controller on which the PIC was mounted. The visibility was then calculated using the equation:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{3.17}$$

Unlike what was initially designed, the best visibility was not achieved at $25°C$, but rather at $14.2°C$. This was likely due to fabrication tolerances and the second round of heating the PIC was subjected to during the tunable coupler annealing process. However, the visibility remained above 99.73%. The overall best visibility,

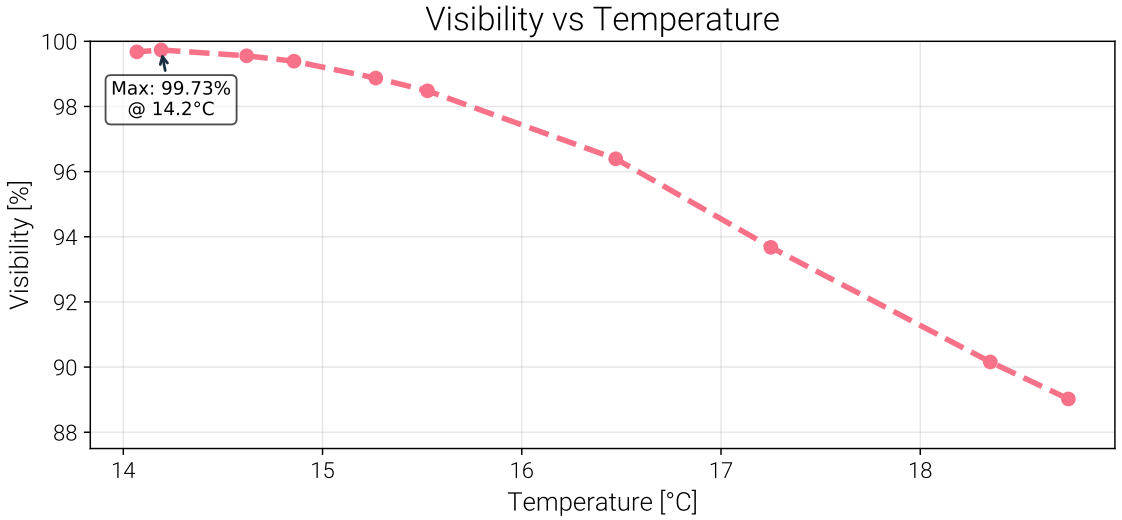


Figure 3.12: MZI visibility characterization at the worst incoming polarization state. The worst maximum visibility was measured at different temperatures to find the optimal operating temperature of the PIC. The best visibility was achieved at $14.2°C$.

achieved with the optimal incoming polarization state, was 99.92%. Decreases in the visibility manifest as errors in the **X** basis, as Bob may detect counts in the wrong interferometer output port, increasing the ϕ_z. As a result, a larger phase error bound is estimated for the privacy amplification step, reducing the final secret key rate.

3.3.6.3 Tunable Coupler Characterization

Another significant improvement relative to [1] was the use of an integrated tunable coupler for basis selection. This allowed us to optimally use the system in different quantum channel loss configurations, without the need to change the hardware. Previously, this would require the use of a completely new PIC. A tunable coupler was also useful to tune the basis selection *after* deployment in the network as the network configurations change and fibers degrade, not to mention the same blueprint can be used to fabricate all receiver PICs.

The coupler's tunability was achieved with the infusion of gold near the BS region, creating a TOPS, similarly to the phase modulators used in the transmitter. This enabled the adjustment of the BS ratio by applying a current to the gold resistor which in turn, changed the temperature of the waveguide. This translated to a change in the effective refractive index of the waveguide [160].

In Figure 3.13 is the characterization of the tunable beamsplitter. The characterization was done by coupling in light from a CW laser at 1550nm into the only

input of the PIC. The output power was then simultaneously measured at all three outputs of the PIC using power meters, one for the **Z** basis and two for the **X** basis. The splitting ratio was characterized by calculating the normalized power at the Z-basis output relative to the total detected power across all three ports ($P_Z / \sum P_i$). This approach is particularly robust as it inherently accounts for fluctuations in input coupling efficiency and propagation losses within the PIC. The current applied to the gold resistor was swept from 0mA to 70mA in steps of 2mA. The PIC was mounted on a TEC to stabilize its temperature at 25°C during the characterization, for convenience, as it is easy to stabilize at room temperature and the visibility of the MZI does not affect the splitting ratio characterization of the coupler. While the extinction ratio (visibility) of the MZI limits the achievable bounds of the tuning range (i.e., the absolute minimum and maximum power levels reaching the Z-port), it does not impact the accuracy of the characterization itself.

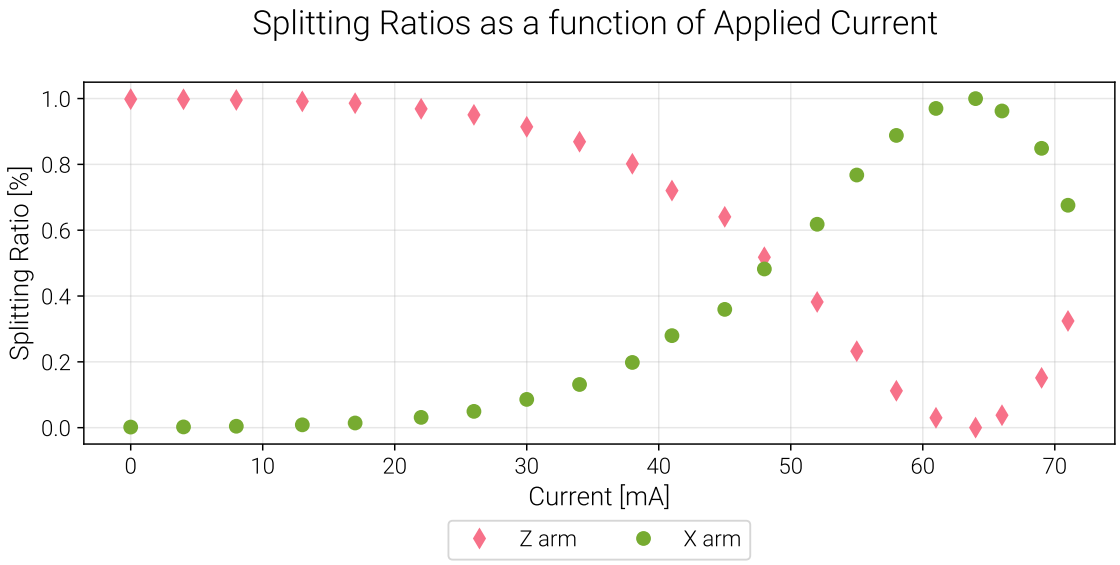


Figure 3.13: Characterization of the tunable beamsplitter. The y axis shows the transmission ratio of the beamsplitter for different currents applied to the gold resistor.

As shown in Figure 3.13, the coupling ratio can be tuned from 0 to 100% by adjusting the current applied to the gold resistor. The splitting ratio follows a $\cos^2(kI^2)$ relationship, where the applied current induces a thermo-optic phase shift via Joule heating. The quadratic dependence of power dissipation ($P = I^2R$) on the input current explains the relatively flat response at low current values followed by the rapid transition toward the inflection point.

3.3.7 NFAD Detectors

Lastly, as part of the effort to minimize system size, cost, and complexity, it was important for us to seek alternative detectors to say SNSPDs or fridge-cooled NFADs. Due to the operation shortcomings of the DA-SPADs explored in the previous chapter (i.e. maximum operating frequency of 1GHz) and being limited to commercially available detectors for demonstration purposes, we opted to use Peltier-cooled NFADs as our single-photon detectors.

Specifically, we used NFADs from Wooriro Co. Ltd. which came packaged with an internal three-stage TEC that allowed cooling of up to $-50°C$. The detectors were fiber-coupled and integrated with a dedicated intermediate control PCB. This was all housed in an insulating box with a one-stage TEC to help stabilize the temperature of the detector's package.

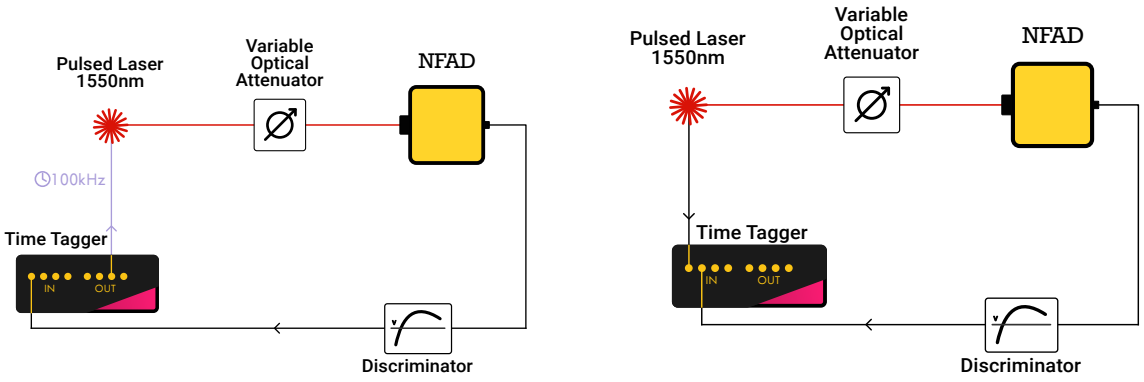


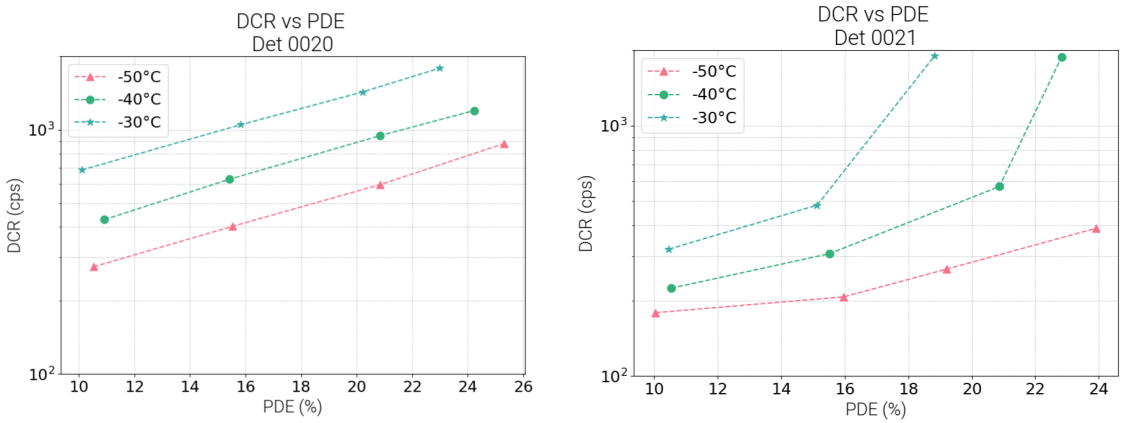
Figure 3.14: (a)Schematic of the setup used for the characterization of the NFADs. (b)Schematic of the setup used for the jitter characterization of the NFADs.

In Figure 3.14a is the schematic representation of the NFAD characterization setup, which is very similar to that described in Chapter 2. A pulsed laser (*Picoquant LDH series*) at 1550nm was attenuated to sub-single-photon level using a VOA. Optical pulses, with a repetition rate of 10kHz, attenuated to 0.1 photons per pulse were used to characterize the PDE and APP of the detectors. The detection signal was connected to a Time-to-Digital Converter (TDC) and used as the histogram's 'Stop' signal, where the laser trigger was used as the 'Start' signal. The histogram domain was set to 100µs in order to capture all the events between two consecutive laser pulses.

3.3.7.1 NFAD Detection Readout Characterization

The analysis of the measured data was also closely resembling the methods described in Chapter 2: The PDE was calculated by dividing the number of detection events in the photon arrival time-bin (with dark counts subtracted) by the total number of photons sent. Here, the signal photon arrival time-bin was defined as a 2ns coincidence window around the main peak in the histogram. The APP was calculated by counting all the events in the histogram, excluding the laser events and dark counts, and dividing them by the number of photon detection events. For the DCR characterization the same correlation measurement was performed with the laser turned off, and the DCR was calculated simply by dividing the total number of detection events by the total measurement time.

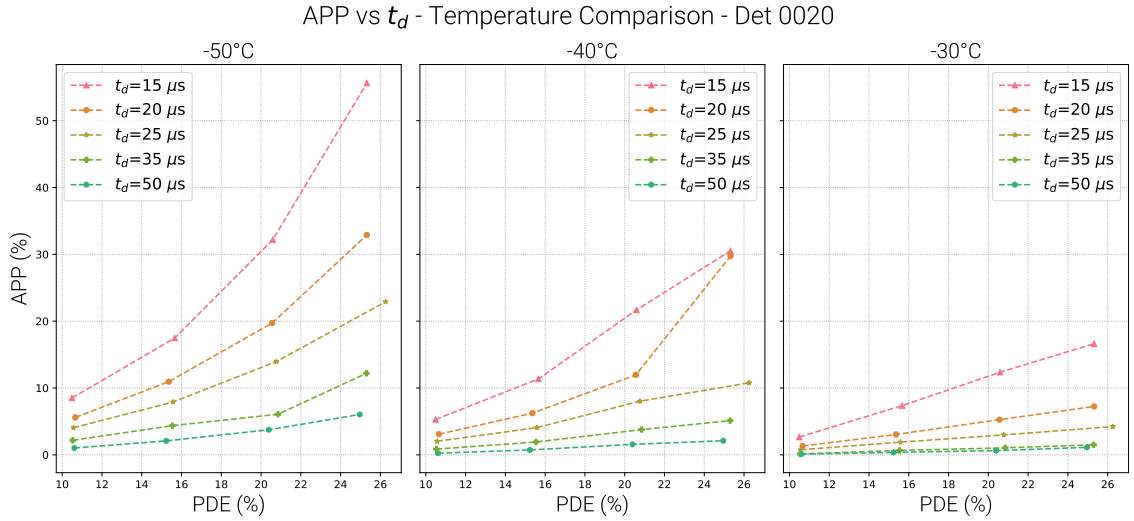
Additionally, the detectors shared a common programmable hold-off/dead time. Therefore, for the characterization, only one detector was biased above breakdown at a time, to prevent the dead time being triggered by dark count events in the other detector. This way, the performance of each detector could be measured independently. Another feature of the NFAD's controlling PCB was electronic crosstalk between the two detectors's readout channels. This could be mitigated by discriminating the incoming signal from the detector above 140mV. A threshold of 150mV was used for the characterization measurements.



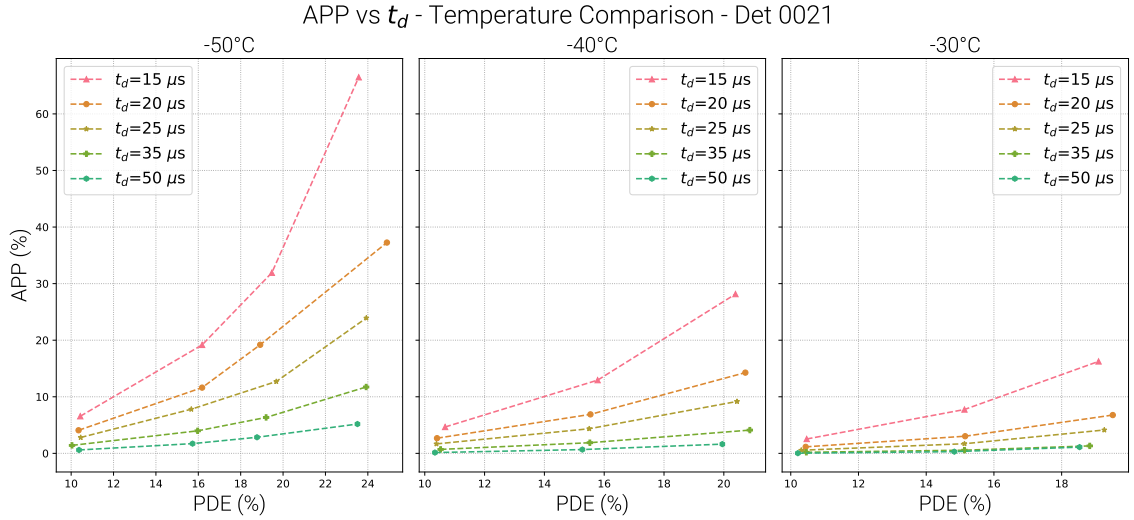
(a) (b)

Figure 3.15: (a)DCR vs PDE for detector 0020 at -50C, -40C, and -30C. (b)DCR vs PDE for detector 0021 at -50C, -40C, and -30C.

Figure 3.15 shows the DCR as a function of PDE for different temperatures, ranging from -50°C to -30°C. As expected, the DCR increased with an increase in PDE and with an increase in temperature.



(a) APP vs dead time (t_d) for -50°C, -40°C, and -30°C



(b) APP vs dead time (t_d) for -50°C, -40°C, and -30°C

Figure 3.16: Afterpulsing probability (APP) as a function of PDE for variable dead time (t_d) for different temperatures. (a) Detector 0020. (b) Detector 0021.

The increase in DCR with PDE arises because higher bias voltages create a stronger electric field in the SPAD's multiplication region. At high temperatures, this strong field enhances Shockley-Read-Hall (SRH) generation, which is the main source of dark counts. As the temperature decreases, thermal carrier generation becomes less important, so SRH contributes less to the DCR. In this low-temperature regime, the DCR is instead dominated by tunneling processes, which depend primarily on the electric field in the depletion region (A more detailed analysis on these processes can be found in Chapter 2). Both detectors showed comparable performances, with detector 0020 having slightly higher DCR for the same PDE and

temperature. At -50°C, both detectors had a DCR of a few hundred counts per second at 10% PDE, which increased to around 1kHz at PDEs above 20% at higher temperatures.

Shown in Figure 3.16 is a comparative study of APP as a function of PDE for different applied dead times and temperatures. As expected, the APP increased with an increase in PDE and a decrease in dead time and temperature. These results are consistent with the findings of earlier SPAD studies [161, 162], where it was observed that APP is higher at lower temperatures and higher PDE operations. The results also align with the widespread understanding that increasing the applied dead time is the most effective way to mitigate afterpulsing. A comparative assessment indicates that *Detector 0021* exhibits marginally higher APP levels than *Detector 0020* across equivalent operating conditions, though both devices have analogous functional dependencies.

From the results in Figure 3.15 and Figure 3.16, we see that for the purpose of this work, i.e. the field deployment over metropolitan distances (10-50km), we may choose to operate the detectors at -30°C and 10% PDE, with a dead time of 15µs. This setting provides a good compromise between low DCR and low APP, while still providing a reasonable PDE. Most importantly, this setting allows to maximize the count rate of the detectors, which is crucial for achieving high secret key rates over short distances.

3.3.7.2 NFAD Timing Jitter Characterization

The detector's jitter was calculated by measuring the time difference between the laser trigger and the detection event using a TDC. In Figure3.14b is the schematics of the setup used for the characterization. Since the detectors were free running, the laser was triggered with its internal trigger at 10kHz. The laser was used as the master clock to synchronize the TDC. The laser trigger was connected to the TDC's start channel, while the detector's output was connected to the stop channel. The TDC (ID1000, *ID Quantique*) was set to high resolution mode, allowing for a minimum bin size of 1ps. The jitter was then calculated by fitting a Gaussian to the histogram of the time differences and extracting the FWHM.

Based on the jitter characterization data shown in Figure 3.17, the NFADs exhibit distinct performances. Both detectors follow the trend of increasing jitter with a decrease in PDE. Detector 0020 demonstrates relatively stable jitter performance across the temperature ranges, having the highest jitter value of 750ps for with 10% and a minimum jitter of 230ps for 20% PDE. In contrast, detector 0021 shows lower

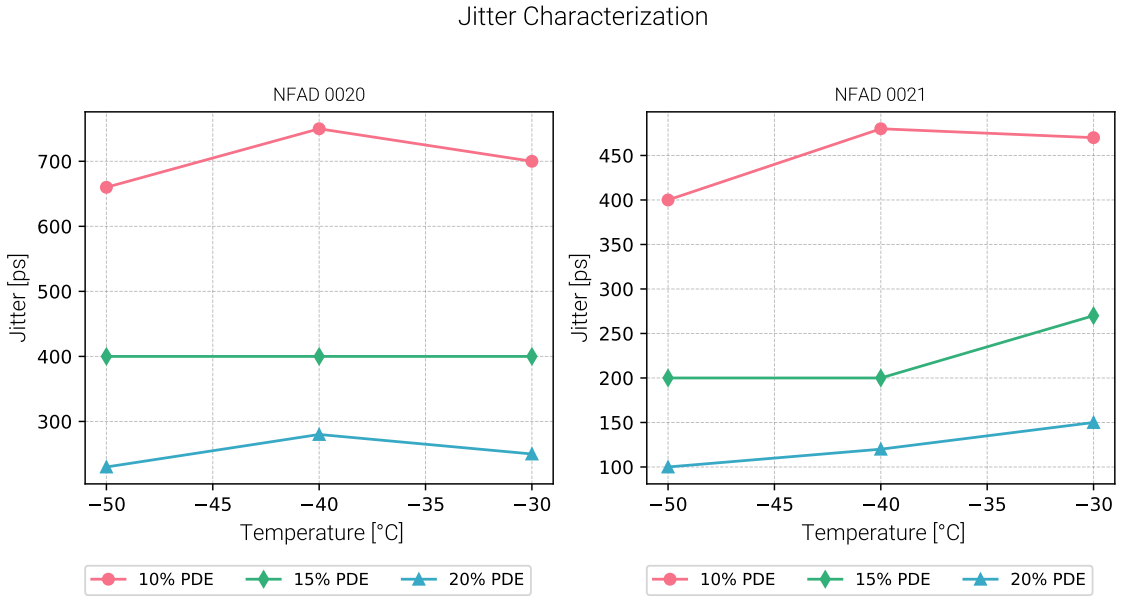


Figure 3.17: Jitter as a function of PDE for different temperatures.

absolute jitter values but exhibits more temperature-dependent behavior, with the jitter increasing with an increase in temperature. Detector 0021 has a minimum jitter of 100ps at 20% PDE, with the jitter increasing up to 470ps at 10%.

For these measurements, the excess bias was actively adjusted to achieve the desired PDE at each temperature setting. Therefore, the dominant factor affecting the changes in timing jitter for detector 0021 can be attributed to intrinsic carrier transport phenomena. Namely, the electron hole ionization coefficients, which vary with lattice temperature, modifying the stochastic multiplication dynamics and spread out avalanche initiation times [163]. In addition, photons absorbed outside the high field region can lead to delayed avalanche triggering due to carrier diffusion, which is also temperature dependent (both the diffusion constant and carrier mobility). The diffusion related tails in the timing histogram can broaden the overall jitter, particularly at lower PDE values where the average avalanche build-up time is longer, allowing more time for diffusion to impact the timing distribution [164, 165].

In order to ensure these discrepancies in performance and overall poor jitter were not due to the readout electronics, we swapped the detectors and repeated the measurements. We also tuned the discrimination threshold, the dead time, the threshold of the TDC, and the discriminated pulse shape to see if these parameters had any effect on the jitter. The results were consistent with the initial findings, indicating that the performance differences are inherent to the detectors themselves rather than the readout circuitry.

This performance variation can be attributed to difficulties encountered in the fabrication process of the NFADs. Even with strict processes and controls, semiconductor fabrication inevitably introduces variations that can significantly impact device performance. Different impurity concentrations, crystal imperfections, doping concentrations, and layer thicknesses can all lead to variations in the avalanche region geometry and electric field distribution. These small differences can dramatically affect the avalanche multiplication and consequently the timing characteristics. Another factor could be the device packaging, as parasitic capacitances and inductances from packaging and bonding can influence the device's timing response. The bad timing performance of these detectors, especially at low PDE, is also influenced by these detectors not being true NFADs, but rather SPADs bonded to a passive quenching resistor inside the temperature control package. This inherently limits the speed at which the avalanche can be quenched, leading to longer recovery times and increased timing jitter.

Unfortunately, because of the jitter performance, a PDE of 20% had to be used instead, which was not ideal for the short network distances we were targeting. As mentioned, a 10% PDE would have minimized the noise contribution of the detectors from dark counts and afterpulsing. In comparison, at 20% PDE, the dark count rate of the detectors was around 900cps, while at 10% it is around 300cps. The main problem of working at 20% PDE, however, is the extremely high APP. Therefore, the maximum allowed dead time of 100µs had to be implemented, which in turn limited the maximum count rate of the detectors to 10kcps. As a result, the detectors became more prone to saturation at low channel losses, even at low mean photon numbers. The following section will present results obtained with our system using these detectors, showing how the QKD performance was affected by these sub-optimal detector settings.

3.4 Results from System Testing and Deployment

The experimental results and settings are summarized in Table 3.1, Table 3.2, Table 3.3 and Fig. 3.19. In Tables 3.2 and 3.3, L refers to the QC length, Att to the total attenuation in the QC, η_x and τ_x refer to the efficiency and dead time, respectively, of the *data* and *monitor* detectors used. T is the detector's operating temperature, and $p_X^{(B)}$ denotes Bob's probability of measuring in the X basis. n and t_{ccp} are the raw key block size and acquisition time respectively, q_z, ϕ_z and SKR denote the $QBER_Z$, phase error rate and Secret Key Rate. For the settings used μ_0 and μ_1

refer to the mean photon number of the signal and decoy states respectively and the ratios of sending x and y states are p_x and p_y respectively. For error correction, we perform in real-time a Cascade algorithm with a block size of 8192 bits.

μ_2	μ_2	μ_1	p_x	p_y
Lab	0.50	0.27	0.33	0.67
Wooriro (deployed)	0.01	0.05	0.33	0.67
Princeton (deployed)	0.05	0.02	0.33	0.67

Table 3.1: Experimental parameters

Once the prototypes were assembled, we began testing the system in a controlled laboratory environment before proceeding to field deployment. Table 3.2 shows the results obtained using different emulated fiber distances with an external VOA in order not to introduce additional physical effects, to assess the dark count limited transmission distance. The receiver detectors used in this case were the same NFAD detectors as in [1] (*Princeton Lightwave*), cooled to $-85°C$ using a Stirling cooler, with a PDE of 20%, a DCR of 120cps, 70ps jitter, and a dead time of 7µs. Although our system was operating at half the repetition rate of the system in Sax et al. [1] (*SKR**), we achieved more than half the SKR across the same attenuation levels. This suggests that the reduction in repetition rate contributed positively to the fidelity of the state preparation, as suggested by the consistently lower QBERs observed.

Quantum Channel		Detectors						Distillation		Performance			
L (km)	Att (dB)	η_{data} (%)	τ_{data} (µs)	η_{mon} (%)	τ_{mon} (µs)	T (° C)	$p_X^{(B)}$	n	t_{ccp} (s)	q_z (%)	ϕ_z (%)	SKR (kbps)	SKR* (kbps)
0	30	21	6	20	7	-85	0.1	8×10^6	1.4×10^4	1.3	4.0	1.80	2.9
0	35	21	6	20	7	-85	0.1	8×10^6	1.7×10^4	1.3	6.0	0.90	1.3
0	40	19	27	20	2	-85	0.4	1×10^4	6.7×10^4	3.3	13.6	0.16	0.2

Table 3.2: Parameters and results of secret key exchanges for different QC attenuation. *Data from [1] that utilizes a source at 2.5GHz with the same detectors.

To validate the chromatic dispersion analysis in Section 3.3.4.1, we measured the maximum achievable transmission distance over standard Single Mode Fiber (SMF).

With optimized laser parameters (100ps pulse width, 0.098 nm spectral width), secure key exchange was achieved over 101.2km of SMF (Table 3.3, lines 1–5). Unfortunately, we were unable to verify the maximum distance in the fourier limited laser scenario (Figure 3.7) for the available narrowband filters in the lab had a width of 0.200nm.

Quantum Channel		Detectors						Distillation		Performance		
L (km)	Att (dB)	η_{data} (%)	τ_{data} (µs)	η_{mon} (%)	τ_{mon} (µs)	T (° C)	$p_X^{(B)}$	n	t_{ccp} (s)	q_z (%)	ϕ_z (%)	SKR (kbps)
12.7	2.7	19	1	19	3	-85	0.50	8×10^6	8.5×10^3	6.12	6.02	12.4
25.3	5.5	19	2	20	3	-85	0.50	8×10^6	9.4×10^3	7.28	4.05	10.1
75.5	14.6	21	3	20	6	-85	0.35	8×10^4	2.9×10^3	6.5	9.13	4.6
101.2	24	21	7	20	8	-85	0.10	8×10^4	1.02×10^3	6.5	10.7	0.7
59.8	24	20	6	20	7	-50	0.50	8×10^6	1.7×10^5	9.7	9.9	0.8

Table 3.3: Overview of the experimental parameters and performance for different fiber lengths and detector temperatures.

To assess the deployment viability, the system's performance was characterized at −50°C, a regime compatible with standard Peltier cooling in portable modules. At this temperature, DCRs increased to approximately 2 kcps, elevating the system QBERs and reducing the maximum secure transmission distance to 59.8km.

Finally, we transitioned to real-world deployment by testing our system over installed fiber network infrastructure. The transmitter was moved to a remote point in the network, allowing us to evaluate system performance under practical point-to-point deployment conditions. The link connects two university campuses in Geneva, with a total fiber length of 4.6km and an overall loss of -9.4dB (Figure 3.18). The fiber is a standard SMF and is part of a larger network that includes other users and connects other institutions. The results of the key exchange can be seen in Table 3.4. We started by again testing the system with the Stirling-cooled Princeton NFADs, cooled to -50°C. We were able to achieve an SKR of 6.7kbps with a QBER_Z of 3.3%. The

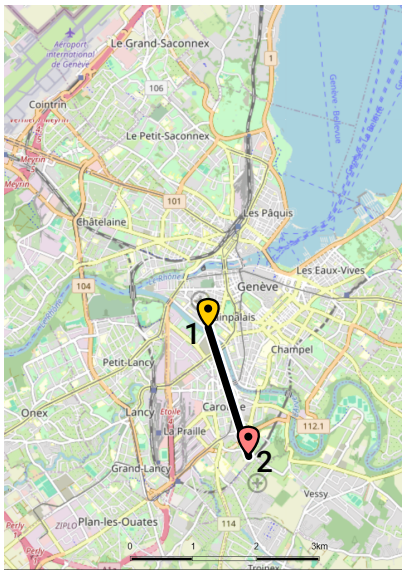


Figure 3.18: Map of the deployed fiber link in the city of Geneva. (2) - Transmitter (1) - Receiver. Map edited from https://facilmap.org

system was stable over several hours (>282h) of operation as shown in Fig. 3.19, demonstrating its viability for real-world applications.

To further understand how the system performs given certain detector specifications, we opted to swap the detector in the **X** basis back to the Princeton NFAD, while keeping the Wooriro NFAD in the **Z** basis. This allowed us to take advantage of the lower jitter and noise performance of the Princeton detector, which would help reduce the QBER_X and thus let us perform a key exchange with the Wooriro detector in the **Z** basis cooled to -30°C and with a dead time of 25µs (results in Table 3.4). This configuration was not ideal, but was the best compromise in lieu of the common dead time design of the control board for the Wooriro NFADs. This feature unfortunately prevented us from compensating for unavoidable differences in the properties of the two Wooriro detectors, which theoretically would have allowed us to further optimize the system performance using both Wooriro detectors.

In particular, the Wooriro NFADs's main bottleneck was the incredibly high jitter, which forced us to work at 20% PDE, resulting in a high DCR and APP. Consequently, we could either choose a dead time of 100µs, already in the detector's saturation regime, or increase the temperature of the detector to reduce the APP, which in turn increased the DCR. The benefit of working at a lower temperature was outweighed by the 50µs dead time we could implement at -30°C, which allowed us to work at much higher maximum count rates, leading to an increase in SKR when compared to the value obtained with the same detectors at -50°C.

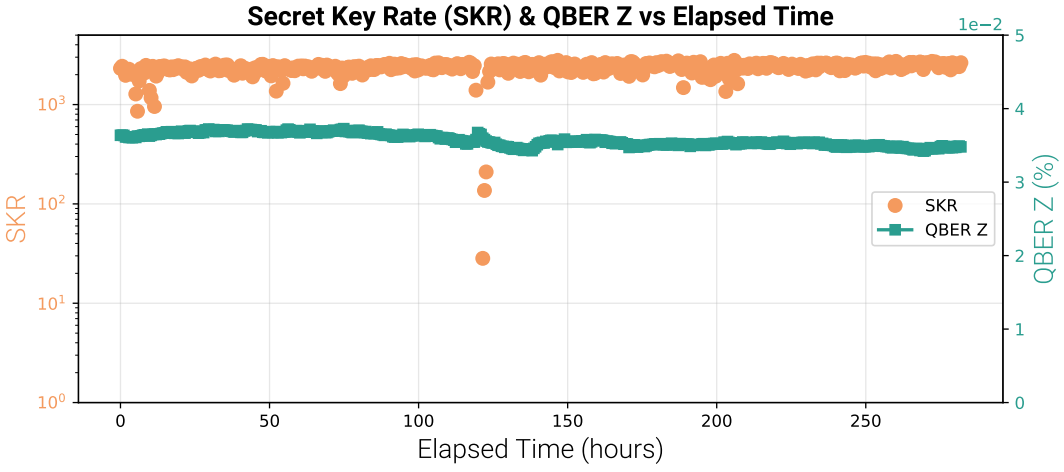


Figure 3.19: Stability of the SKR and QBER_Z over time during the point-to-point key exchange over the deployed fiber link. The system was stable over 78 hours of operation, demonstrating its viability for real-world applications.

Through this sequence of experiments, progressing from controlled laboratory

Quantum Channel		Detectors						Distillation		Performance		
L (km)	Att (dB)	η_{data} (%)	τ_{data} (µs)	η_{mon} (%)	τ_{mon} (µs)	T_{data} (° C)	$p_X^{(B)}$	n	t_{ccp} (s)	q_z (%)	ϕ_z (%)	SKR (kbps)
4.6	9.4	21	6	20	6	-50	0.5	8×10^6	7.2×10^3	3.3	2.0	6.7
4.6	9.4	20†	100†	21†	100†	-30	0.5	8×10^6	2.1×10^4	4.8	2.6	1.9
4.6	9.4	20†	25†	20	6	-30	0.5	8×10^6	8.7×10^4	6.5	2	4.2

Table 3.4: Parameters and results of key exchange of the deployed system. Alice was located at a remote point (2) and Bob at point (1). † represents the Wooriro NFADs specifications. If not specified, the Princeton NFADs were used.

conditions to field trials, we systematically linked fundamental noise limits, channel impairments, and detector properties to their real-world effects on quantum key exchange. This comprehensive approach not only validated the integrated QKD system's performance but also highlighted critical areas for future enhancement, particularly in detector technology and system integration.

3.5 Discussion on System Performance and Outlook

In this chapter, we have presented the design, fabrication, and testing of an integrated QKD system based on the time-encoded BB84 protocol with one decoy state. The system features transmitter and receiver PICs made from silicon on oxide and silica respectively. All optical components, except for the gain-switched laser and detectors, were integrated on-chip. Custom electronics were developed to support the operation of the system and compact it into standard 19in racks. The system's components were fully characterized and optimized prior to assembly. The system was first tested in the lab using an external VOA and spooled fiber to emulate different quantum channel losses and distances. Finally, the system was deployed over a 4.6km installed fiber link in the metropolitan Geneva area.

The experimental results demonstrate the system's capability to achieve secure key distribution over metropolitan distances. The results clearly demonstrated that, given sufficiently high-performance detectors, the system is fully capable of deployment in metropolitan optical networks. However, as highlighted in the introductory chapter of this work, detector performance remains the central bottleneck for practical QKD implementations [127, 166]. The quality of the detectors directly limits the achievable QBER, SKR, and consequently the feasibility of real-world deployment.

The success of our system was ultimately limited by the performance of the

available detectors. Despite extensive optimization efforts, including adjustments of the operating temperature, discrimination threshold, dead time, and bias voltage, the inherent device limitations could not be overcome. This reinforces an overarching theme in the literature: the intrinsic performance of single-photon detectors cannot be compensated for by system-level optimizations alone. Continuous advancements in detector technology, particularly in reducing dark counts, afterpulsing, and timing jitter, are critical to unlocking the full potential of quantum communication systems.

Recent advances in integrated QKD have highlighted the broader landscape in which our work is situated. Compact and stable photonic transceivers have been demonstrated, with transmitter and receiver functionalities integrated on hybrid platforms, achieving millimeter-scale footprints while maintaining robust performance [167, 168]. A major breakthrough in practical deployment has also been achieved in multiplexing QKD with classical data traffic. This enabled the simultaneous classical transmission of data signals at 33.4Tbps over 80km utilizing a single optical fiber [169]. On the free-space QKD front, the satellite QKD domain is also experiencing rapid advancements. In [170], the authors investigate the feasibility of satellite-based quantum key exchange using low-cost compact nano-satellites. This miniaturization trend, combined with decreasing launch costs, is making space-based QKD more accessible. European efforts are also advancing, with ESA planning to launch the satellite Eagle-1 with an experimental space-based QKD system, in late 2025 or 2026 [171].

Despite these advancements, several challenges remain for the widespread adoption of QKD. Many QKD deployments rely on trusted nodes for extending distance, which introduces the need to trust these intermediate points. Furthermore, channel authentication often uses classical cryptography, inheriting its vulnerabilities. Scaling to many users remains difficult and expensive, as each pair may need dedicated quantum channels or trusted nodes. Integration into existing networks with classical traffic is often non-trivial, as co-existence of classical and quantum channels can impose constraints

Another challenge that must be addressed is the appropriate management of the secret keys once generated. While QKD solves the fundamental problem of key distribution, it still leaves the challenge of key management to the classical domain. The information-theoretic guarantee provided by the quantum channel is effectively broken if the resulting keys are subsequently stored in insecure environments, where they are exposed to standard side-channel vulnerabilities. A truly secure quantum cryptosystem must encompass the entire lifecycle of the key. This raises further

questions about how to securely manage, refresh, and eventually delete old keys.

As we saw in the previous sections, virtually all QKD protocols depend on some form of trusted random number generator. Fundamentally, the security of QKD protocols are directly affected by the quality of the random numbers when generating secret keys. Hence, we now turn our attention to QRNGs, which form a crucial building block for practical implementations of QKD.

Chapter 4

Integrated Quantum Random Number Generator

4.1 Introduction to Quantum Random Number Generators

Randomness is a fundamental resource in modern science and technology. As we have seen in the previous chapter, randomness underpins secure communications, cryptographic protocols, simulations, randomized algorithms, and even foundational tests of physical theories. However, the generation of true randomness is not trivial. Classical PRNGs rely on deterministic algorithms starting with a finite seed to produce long sequences of numbers that resemble randomness. While these are computationally efficient, they are in fact predictable once the seed is known. Commonly also referred to as True Random Number Generator (TRNG), Hardware Random Number Generators (HRNGs) are based on measuring the entropy from unpredictable physical phenomena, such as thermal noise, chaotic systems, or metastable circuits. The unpredictability of these systems is rooted in their complexity and sensitivity to initial conditions, or in other words our ignorance of the system environment, but they remain vulnerable to classical modeling, environmental influences, and potential side-channel attacks.

As a result, the need for an intrinsically random process is greatly desirable, which has in turn generated great interest in QRNGs as quantum mechanics provides a radically different solution to this problem. According to the postulates of quantum mechanics, measurement outcomes are intrinsically probabilistic and cannot be predetermined, even with complete knowledge of the system [172]. This

indeterminacy allows us to construct devices that harness quantum processes to produce fundamentally unpredictable sequences. In contrast to classical approaches, the randomness certified by quantum theory is not a consequence of ignorance about hidden variables but it rather arises from the physical laws governing the system itself [173].

Many QRNG architectures have been developed over the years, differing both in their physical principles, and trust assumptions. The simplest and first implementations of QRNGs were, for example, sending individual photons through a beam splitter and recording the output port [174, 175]. More advanced QRNG designs exploit Continuous Variable (CV) quantum states, such as vacuum fluctuations, or phase noise in coherent laser sources measured by balanced homodyne detection [176–178]. These CV architectures overcome the practical constraints of single-photon methods, offering significant advantages in speed and integration. By utilizing coherent detection with high-bandwidth detectors, CV systems achieve substantially higher generation rates (often Gbps) compared to the rate-limited single-photon detectors. Furthermore, the use of standard photonic components makes them highly suitable for miniaturization and on-chip integration. Homodyne-based vacuum-fluctuation QRNGs have reached tens to hundreds of Gbps and can be built using off-the-shelf telecom components, or using integrated photonic circuits, making them particularly well-suited for large-scale deployment [179–182]. An ongoing issue with the design of QRNGs is the level of trust that must be placed in these devices. In a fully Device Dependent (DD) QRNG, the security relies on an accurate and absolute characterization of the system as deviations from the model can compromise the randomness of the output, and even be exploited to be controlled. As such, DI QRNGs have been proposed, employing Bell inequality violations to certify randomness without requiring any trust in the internal functioning of the devices. While they are theoretically robust, DI approaches impose stringent experimental demands. Loophole-free Bell tests require high detection efficiencies and reliable entangled sources [183, 184]. DI approaches also suffer from very low generation rates, typically on the order of a few bits per second [185–187].

A promising middle ground may thus be found with SDI approaches bridging the gap between the two extremes. By introducing relaxed assumptions—such as bounds on the Hilbert space dimension or constraints on the energy of the prepared states—SDI frameworks allow for partial certification of randomness while retaining a high level of practicality [188–191]. In our context, we sought to keep pushing SDI protocols toward practical deployment by developing a compact, integrated QRNG

architecture that simultaneously ensures high performance and strong security guarantees. Unlike earlier laboratory demonstrations, such as the proof-of-principle experiment by Rusca et al. [135] that provided important theoretical validation but lacked the compactness and robustness needed for field applications, the aim of this work was to realize a system that is not only theoretically secure but also portable and scalable. This required re-thinking the physical implementation at both the photonic and electronic levels to minimize system size.

The specific objectives of this work can therefore be summarized as the following:

- Design and implement an integrated SDI QRNG that leverages photonic integration to place sources, modulators, detectors, and passive optical components on a single optical chip.

- Incorporate SDI self-testing protocol demonstrated in Rusca et al. [135] capable of certifying in real time the randomness output under minimal assumptions, without relying on full device characterization.

- Develop a compact and low-power detection electronics tailored to the integrated platform, enabling robust real-time acquisition of quantum randomness.

- Evaluate the system performance in terms of entropy, bit generation rate, stability, and compliance with standard randomness test suites, while benchmarking against state-of-the-art QRNG implementations.

By pursuing these objectives, this work aimed to demonstrate that SDI QRNGs are not limited to conceptual or proof-of-principle studies but can be realized as compact, integrated devices with potential for widespread practical deployment in cryptography, secure communications, and emerging quantum technologies.

This chapter explores the principles and practical realization of integrated QRNGs, with particular emphasis on SDI approaches. The following sections will introduce the framework of SDI QRNGs, including self-testing techniques that allow partial certification of randomness under minimal assumptions. We then present the experimental attempt at developing such a QRNG based on integrated photonics, detailing the design of the detection electronics, characterization of key photonic components, and the overall system performance.

4.2 Semi-Device Independent Quantum Random Number Generators

As mentioned, to address the difficulty of DI implementations, SDI protocols have been proposed. These, inspired by the P&M implementation of QKD (explained in the previous chapter), allow communication between the quantum devices but require some level of characterization/assumption on the states preparation stage [188–190] or measurement device [191]. The principle behind an SDI approach was to allow for a relatively simple implementation with limited characterization. However, semi-DI devices cannot be fully treated in a black box manner and must satisfy a number of conditions, such as a bound on the dimensionality of the relevant Hilbert space. This is enough to certify non-classical correlations in a P&M scenario without requiring entanglement, allowing the implementation to be far simpler [192,193]. However, this poses some implementation challenges since photons, the most common carriers of quantum information, may have quantum states spanning infinite Hilbert space dimensions and therefore, one cannot assume that information is encoded solely in a few degrees of freedom without a thorough characterization of the devices used. An important work on SDI-QRNG was published in 2017 by Himbeeck et al. [194], where they proposed the use of a well characterizable observable, for example the energy of the state, to constrain the exchanged message and certify its quantum nature. In the following work [195] they present the analysis of a QRNG protocol based on the semi-device-independent scheme introduced in [194]. It is worth discussing these two works since they provide a basis for the security framework for the QRNG work studied in this thesis. This will be done in the following sections.

4.2.1 Self-Testing Semi-Device Independent QRNG

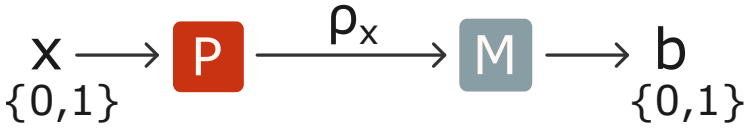


Figure 4.1: Schematic representation of a Self-Testing Semi-Device Independent QRNG.

The protocol used in this work was based on the work of Rusca et al. [29]. Depicted in 4.1, we have the standard P&M protocol. A Preparation stage and a Measurement device are connected via a quantum channel, and the partially characterized

Preparation stage receives a binary input $x \in \{0, 1\}$ that will yield a quantum state ρ_x. This state is sent to and measured at the *uncharacterized* Measurement device M, outputting a binary output $b \in \{0, 1\}$.

To an external observer, with access only to the input and output statistics of the system, the behaviour of the devices can be characterized by

$$P(b|x) = \text{tr}\left[M_b \rho_x\right] \tag{4.1}$$

which can be further generalized by first assuming a hidden random classical parameter/variable (λ) shared between the two devices that generates correlations between the behaviour of the two devices. The probabilities will then, in a trusted scenario, take the form:

$$P(b|x) = \sum_{\lambda} p_\lambda \text{tr}\left[M_b^\lambda \rho_x^\lambda\right] \tag{4.2}$$

As proposed in the work [194], ρ_x is constrained by means of an observable H that represents a physical property of the system we can either control, monitor, or trust. A constraint on H translates to a constraint on its mean value $H_x = tr[H\rho_x]$. However many observables, like H, may be used to describe the constraints on the system, but they all should be well-defined. In [194], H was proposed to be the photon number operator. In this case, the average photon number would then be the physical property to be monitored and characterized:

$$H_x = \text{tr}[H\rho_x] = \sum_{\lambda} p_\lambda \text{tr}[H\rho_x^\lambda] \leq \omega_x, \forall x \tag{4.3}$$

where ω_x is the threshold value for the mean photon number of the state ρ_x. This constraint is reasonable in an optical implementation, as it can be monitored by measuring the optical power at the output of the source. Based on this bound, the authors identified analogues of Bell inequalities, which can be used to discriminate between genuine quantum devices and fully classical/deterministic ones. This can be applied to the specific scenario we worked with. For only one possible measurement M, and the states $\rho_x, x \in \{0, 1\}$ being non-orthogonal (otherwise they would be deterministically distinguishable, rendering this problem trivial) the input/output relations can be described by the 4 probabilities:

$$P(b|x) = \text{tr}[M_b \rho_x] \quad \text{where} \quad x \in \{1, 2\} \quad \text{and} \quad b \in \{\pm 1\} \tag{4.4}$$

Conveniently, these probabilities can be expressed in terms of the correlators

$E_x = P(+1|x) - P(-1|x)$ $(x = 1, 2)$ where $E_x = P(+1|x) - P(-1|x)$ is the expectation value of the observable $M = M_{+1} - M_{-1}$, with $-1 \leq M \leq 1$. The *correlations* E_x characterize the bias of the output b for each input x, and represent how the output of M is correlated to the input of S [194]. If $E_1 = 1$ and $E_2 = -1$, then the output b is perfectly correlated with the input x, and the system is deterministic. We can then calculate a set of classical correlations $\mathcal{C}_{\omega_\infty,\omega_\in}$ under the max average assumption (Equation: 4.3). If the measured correlation $\mathbf{E}$ lies outside this set, then the system is said to exhibit correlations that cannot be modeled by any classical hidden variable theory, thus certifying its quantum nature. In our case, the boundary of this set is defined by:

$$|E_1 - E_2| \leq 2(\omega_1 + \omega_2) \tag{4.5}$$

Now, to quantify the entropy produced by the input and output statistics of the system, in [195] was developed a semi-definite program (SDP) that returns a lower bound on the conditional Shannon Entropy $H(B|X, \Lambda)$. This bound can then be used as a self-test to certify the randomness of the system. Specifically, this self-test is done by fixing the entropy threshold h_{th} before the experiment starts that is adjusted according to the characterization of the devices in use. After running the experiment several times, we check that the calculated witness is greater than h_{th}. If that is the case, then the runs passed the test and genuine quantum randomness can be ensured, with the amount of randomness that can be extracted from the output b being:

$$H(B|X, \Lambda) \geq \frac{1}{n} n \left(h - c\sqrt{\frac{log(\epsilon/2)}{n}} - d\frac{log(\epsilon/2)}{n} \right) \tag{4.6}$$

where n is the number of runs of the experiment, c and d are constants that depend on the specific protocol, and ϵ is the security (confidence) parameter. h refers to a "nominal" entropy rate (in bits) one would get under ideal/asymptotic conditions given the observed correlations.

In summary, the framework established in [195] provides a robust method for certifying the randomness of quantum devices through a combination of Bell inequality violations and entropy bounds. By leveraging these techniques, we can ensure that the outputs of our quantum random number generator are genuinely random and not subject to classical deterministic influences.

4.3 Experimental SDI-QRNG Prototype

4.3.1 Homodyne Detection

In our implementation, we utilized a balanced homodyne detection (BHD) scheme to measure the quadrature of incoming quantum states, following the approach demonstrated by Rusca et al. [29], a Binary Phase Shift Keying (BPSK) scheme was used, where the source prepared two coherent states with a π phase difference and the same average photon number. Although BHD does not achieve the theoretical optimum of minimum-error discrimination measurements, it offers a practical compromise between implementation complexity and measurement effectiveness.

Homodyne detection is a method of extracting the information encoded in one of the phase quadratures of the incoming signal by mixing the signal with a reference signal of the same frequency. The term *Homodyne* refers to this single-frequency approach, while heterodyne measurements employ a reference signal with a different frequency than that of the encoded signal [196]. In both cases and in our context, the incoming quantum state (signal) is interfered with a local oscillator (LO) acting as a reference at a 50:50 BS. The photodiode at the output of the beam-splitter produces a current (I_{out}) which reflects the quadrature fluctuations of the signal state [197]. This relationship can be expressed mathematically as:

$$\Delta n_b = \dot{\bar{n}} T 2 \Delta X_q^2 \tag{4.7}$$

where Δn_b is the variance of the output current, $\dot{\bar{n}}$ is the mean photon rate in a mode for a given integration time T and ΔX_q is the quadrature variance. The quantum nature of the output current can be measured from the shot noise characteristic, defined as how the noise variance scales proportionally with the square root of the integration time ($\sqrt{T}$). Equation 4.7 follows from the standard treatment of balanced homodyne detection in the large local-oscillator limit, where the difference photocurrent variance is proportional to the quadrature variance [196].

A further enhancement in sensitivity comes from employing a balanced detection scheme for homodyne measurements. Similar to homodyne detection, the signal and Local Oscillator (LO) are combined at a 50:50 BS, but now the two output ports are directed to two separate photodiodes. The electrical signal generated by both detectors are then subtracted from each other with a subtraction circuit. This configuration cancels out the common-mode noise, thus improving the SNR. This is very effective in mitigating classical noise (laser power fluctuations and electronic

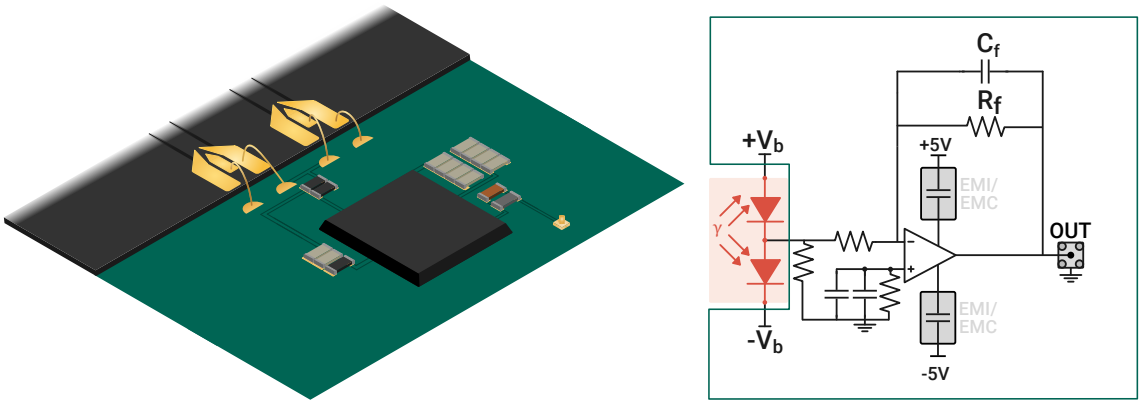
noise) that can mask the quantum signal [198]. Balanced Homodyne Detection (BHD) is widely used in a number of applications such as gravitational wave detection [199, 200], quantum communications [201–203], metrology [204, 205], bioimaging [206, 207] and quantum state tomography [208, 209].

While BHD offers practical advantages through its relatively simple implementation - requiring only a BS, local oscillator, and two photodiodes - it does have some limitations. Experimental analyses [210] have shown that homodyne detection exhibits lower minimum entropy ($H_{\min}$) compared to alternative approaches like Unambiguous State Discrimination (USD) and Unambiguous State Exclusion (USE) when dealing with transmission loss between the source and measurement. This does not represent a problem for QRNG applications since the channel is very short. Due to it using CV quantum states, homodyne detection exploits the use of fast detectors as well as off-the-shelf telecommunication components, making it a practical and cheap choice when compared to discrimination methods working with DV. Using BHD also allows for higher random number generation rates - 113Mbps [189], 270Mbps- when compared to works that have used USD and USE - 16.5Mbps [211].

4.3.2 Design of a Transimpedance Amplifier for Balanced Homodyne Detection

The output signal resulting from BHD is generally too weak to be directly processed by an FPGA device. To overcome this, an amplification circuit must be used to amplify the current of the signal. In our implementation, we opted for a Transimpedance Amplifier (TIA) to convert the current signal into a voltage signal that can be later discriminated by a leading edge discriminator. The correct design of the TIA circuit is crucial for the BHD setup, as it will determine its overall performance. The TIA should have a high bandwidth, low noise, and high gain to ensure the signal is accurately captured and processed.

The TIA circuit represented in Figure4.2 was designed and developed in-house by Claudio Barreiro and I performed the component optimizations. We chose OPA847 from texas instruments as the op-amp due to its combination of high-speed performance and low-noise characteristics. Its high Gain Bandwidth Product (GBP), coupled with minimal input voltage and current noise, provided optimal wideband transimpedance amplification at low to moderate gain levels, while maintaining signal integrity through low distortion. The chosen TIA had an intrinsic bandwidth of 3.9GHz, which should be sufficient for the high-speed detection required by our self-testing SDI-QRNG with an input modulation of up to 1.25GHz.



(a) (b)

Figure 4.2: a) 4D view of the TIA circuit bonded to the two fast photodiodes. b) Electrical schematic of the TIA circuit.

Three critical parameters determine the overall circuit's performance: The diode capacitance (C_D), the desired TIA gain (R_F), and the GBP. To optimize the frequency response, a feedback capacitor (C_F) can be precisely calculated once these three parameters are established. One must however take into account the combined effect of the diode capacitance and parasitic capacitances. For an optimized C_F, the cutoff frequency of the of the TIA circuit is given by [212]:

$$f_{-3dB} = \sqrt{\frac{GBP}{2\pi R_F C_D}} \text{ (Hz)} \tag{4.8}$$

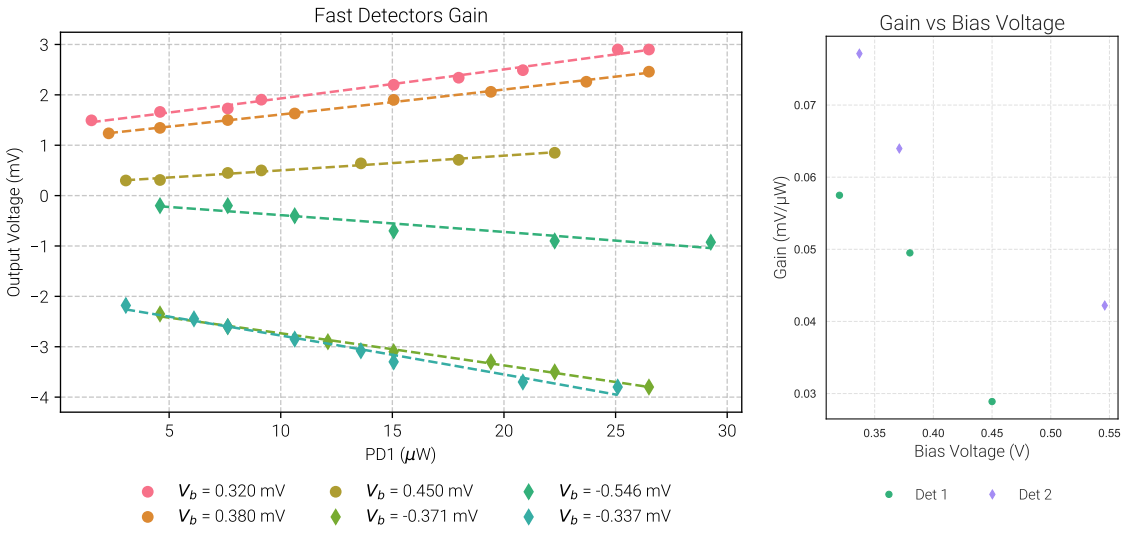
From this equation, we can see the direct effect of the the input source capacitance in the TIA's bandwidth. Indeed, there are many parasitic contributors to this value seen by the TIA circuit apart from the photodiodes themselves. Such parasitic capacitances are a common issue in high-speed circuits, and they can significantly reduce the bandwidth of an amplifier. Parasitic capacitances inevitably occur in transmission lines - whether between different lines themselves, between lines and ground, or between lines and other components. Given these effects, precise circuit design becomes critical to optimize the bandwidth. In our design, to minimize parasitic effects:

- We designed a 4-layer PCB with a characteristic impedance of 50Ω, ensuring proper impedance matching between the photodiode and TIA.
- A dedicated ground plane was incorporated to minimize stray capacitances to ground.
- The TIA was placed as close as possible to the photodiode and was directly

bonded to them.

4.3.2.1 Amplifier Gain Linearity

As our goal was to measure the quadrature of the signal by discriminating between the different voltage levels at the output of the TIA circuit, we must ensure the linearity of the amplifier gain. To check for this, we measured the output voltage as a function of incoming optical power for each fast Photodiode (PD) at a time. Since we are working with an integrated circuit, and the BS splitting ratio to the photodiodes cannot be tuned, we will always have both PDs illuminated. To account for this, when designing the *PIC mezzanine* we included two neighboring pads at the detector's output (see Figure 4.2a) that can be soldered together or unsoldered, in order to have full control over what detector we are collecting the data from. While blocking one of the PDs, we measured with an oscilloscope the output voltage of the TIA circuit across a range of 30µW. The same was done with the other PD blocked. These measurements were repeated for different values of detector bias, and the results can be seen in Figure 4.3.



(a) (b)

Figure 4.3: a) Output voltage versus incident optical power for the fast photodetectors under different bias voltages. The data demonstrates predominantly linear behavior, with fitted regression lines highlighting the gain associated with each bias condition. Positive slopes correspond to Detector 1, while negative slopes correspond to Detector 2, consistent with the polarity of the applied bias b) Module of the gain as a function of the module of applied bias voltage for the photodetectors.

The data suggests that both photodiodes exhibit an approximately linear re-

sponse with respect to the incident optical power, although the degree of linearity varies between bias conditions. For Detector 1, the extracted gain lies between $0.029 - 0.057\text{mV}/\mu\text{W}$, with the majority of traces showing a mean deviation from linearity of about 2 to 19%. Detector 2 shows a similar behaviour with the gain ranging from -0.042 to $-0.077\text{mV}/\mu\text{W}$, with a mean deviations of 4 to 12%. Overall, the results suggest the TIA stage itself was behaving as designed, as the observed deviations from ideal linearity arose primarily from the photodiodes at certain bias points, rather than from limitations in the TIA.

Another important result to highlight is the decrease of gain with increasing bias voltage for both detectors, indicating an intrinsic issue with these devices. In APDs, the gain is expected to increase with bias voltage due to the enhanced electric field in the depletion region that facilitates impact ionization [213]. The observed decrease in gain with an increase in bias voltage, suggests potential surface leakage or other forms of degradation at the juntion. The increased absolute bias can cause surface defects to disrupt the normal gain mechanism that could result from manufacturing imperfections [214–216]. High load resistances in series with the PDs can also cause the effective gain to decrease with the increase in bias voltage, as the voltage drop across the internal resistor might limit the multiplication process [217]. The first mechanism seemed more likely given the low load series resistors used (as shown in Figure. 4.2b) were 50Ω and 0Ω, and electrical bondings tend to have resistances in the $\text{m}\Omega$ range ($\approx 0.045\Omega$) per millimetre for a 25µm wire [218].

4.3.2.2 Detection Frequency Bandwidth

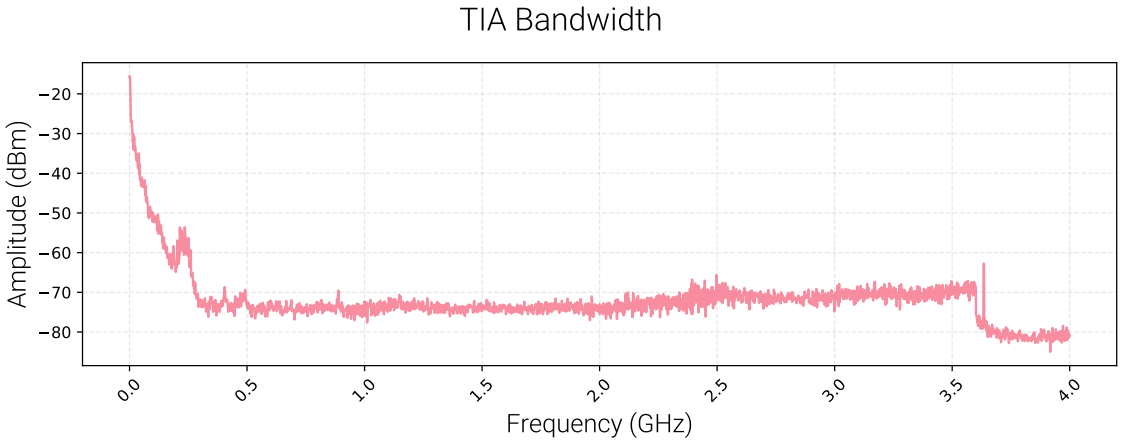


Figure 4.4: Bandwidth measurement of detection circuit with 40mW external laser power.

We then analyzed the noise characteristics in the frequency domain, comparing frequency-domain data of the spectrum analyzer's intrinsic noise, the BHD's electronic noise, and the BHD's output noise spectrum under varying LO power conditions, as shown in Figure 4.4. This measurement was done using a spectrum analyzer with 6GHz bandwidth (Rohde & Schwarz FSC Spectrum analyzer).

From the figure we can see a cutoff at low frequencies as the frequency approaches zero. This is often referred to as the "DC bin" and is a common artifact in Fast Fourier Transform (FFT) spectrum analyzers due to residual DC from the analyzer's own input amplifiers [219]. At low frequencies, near 0Hz and up to 200MHz, spectrum analyzers often exhibit artifacts due to LO signal leakage. When converting input signals to intermediate frequencies, some LO signal can inadvertently leak into the input, generating these large unwanted signals characteristic of spectrum analyzer measurements. For frequencies above 2.2GHz, we see an increase in signal power. This is also an artifact due to signal integrity issues in the measurement setup. When approaching high frequencies, impedance mismatches in connectors and even PCB traces become more apparent, and spectrum analyzers themselves can introduce artifacts due to more LO leakage, mixer nonlinearities, and digitization effects.

Using the data, we can consider our system's upper cut-off frequency to be approximately 3.5GHz. These results demonstrate that we can achieve a high bandwidth with our BHD setup, which is essential for the high-speed operation required by our protocol with a state modulation of up to 1.25GHz.

4.3.2.3 Common Mode Rejection Ratio

To characterize the effectiveness of the subtraction circuit during the BHD, we investigated the Common Mode Rejection Ratio (CMRR) which is calculated by determining the power difference between the differential mode signal and common mode signal in the frequency domain. The CMRR is a crucial parameter for the BHD, as it directly affects the quality of the output signal and the overall performance of the QRNG system. A higher CMRR indicates better noise rejection and improved signal quality, which is essential for achieving high random number generation rates.

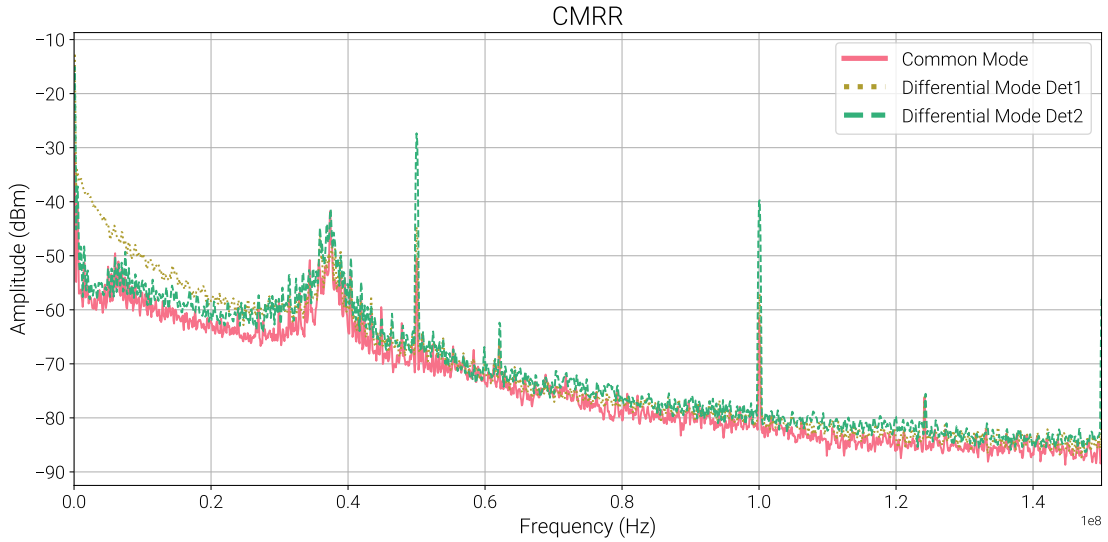


Figure 4.5: Common-mode rejection ratio measurement showing the comparison between common mode (full) and differential mode responses (dashed).

As demonstrated in Fig.4.5, the CMRR was obtained by measuring the output signal of the BHD illuminated with an external laser pulsed at 50MHz in two scenarios: (dashed lines) only one PD was illuminated at a time while the signal of the other is blocked, and (full line) both PDs were illuminated. The external laser was coupled into the chip via a lensed fiber and since the CMRR has no relation to the input optical power [220], an arbitrary input power of 25mW was chosen. To eliminate the common mode signal as effectively as possible, the gain of both detectors must be finely adjusted to achieve a smaller residual signal. The CMRR can be calculated based on the maximum difference of the fundamental harmonic spectral power. In our system, the CMRR reached 46.2dB.

4.3.2.4 Quantum to Classical Noise Ratio

Lastly, to quantify the noise output by our BHD, we must measure Root Mean Square (RMS) of the signal as our output is AC-coupled. We considered the total noise to be a combination of electronic noise and quantum noise. The electronic noise includes background noise of the measuring oscilloscope, the BHD, and the power supply. We then calculated the quantum noise by subtracting the electronic noise from the total noise. The Quantum to Classical Noise Ratio (QCNR) was then calculated as the ratio of the quantum noise to the electronic noise. We measured the RMS for different powers of the LO with a 6GHz oscilloscope (MSO64B - Tektronix). The time scale was set to 10µs and the sampling rate to 20GHz. The trigger was

set to falling edge and the trigger level to 50%. In the oscilloscope's trigger settings we added Low Frequency (LF) noise reject to remove the 1Hz noise sine wave being picked up in the ground signal. For each LO power, we measured the RMS 10 times and considered the average result. As seen in 4.6, the clearance is maximum at around 19dBm. The measurement also shows a good linearity of the BHD until LO optical powers of 260µW where it appears to saturate. Such high clearance is a good indicator of the quality of the BHD and from this, high randomness generation rates can be expected.

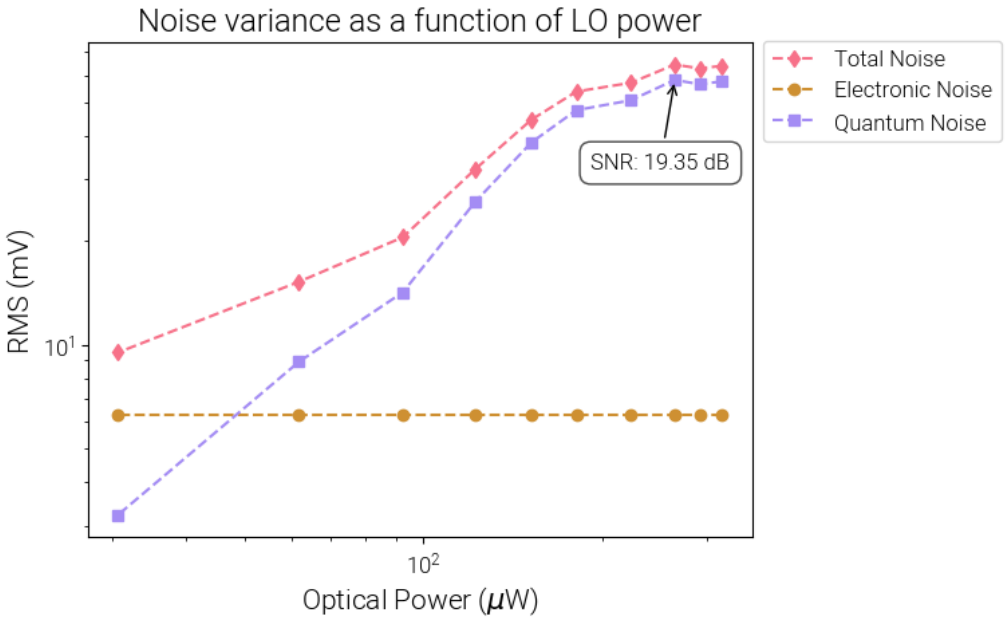


Figure 4.6: Noise variance as a function of LO power. The total noise increases with optical power and eventually saturates, while the electronic noise remains constant and independent of LO power. The quantum noise scales linearly with optical power and dominates at the electronic noise higher powers. The maximum measured SNR is 19.35dB.

4.3.3 Overview of the QRNG System

The experimental setup used for the Integrated SDI-QRNG was adapted form the work of Rusca et al. [29] and is shown in FIG4.7. A Binary Phase Shift Keying (BPSK) scheme was used, where the source prepares two coherent states with a π phase difference and the same average photon number. A homodyne measurement is then used to project the quadrature of the incoming states and discriminate the two.

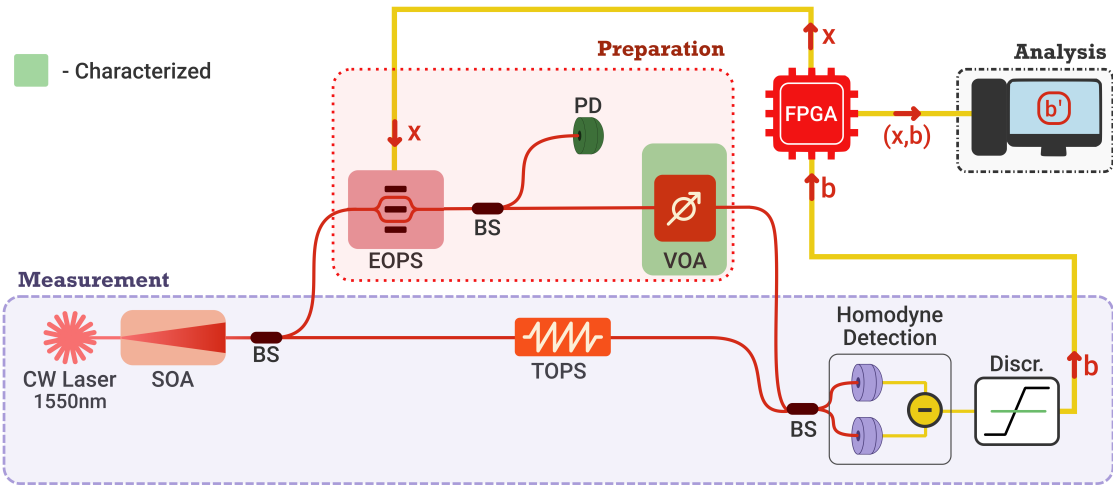


Figure 4.7: Experimental setup of implementation of the self-testing SDI protocol. CW - Continuous Wave; SOA - Semiconductor Optical Amplifier; EOPS - Electro Optical Phase Shifter; BS - Beam Splitter; VOA - Variable Optical Attenuator; PD - Photodiode; TOPS - Thermo-Optic Phase Shifter; Discr. - Discriminator.

The optical setup was completely integrated on-chip. The chip housed a continuous-wave laser at 1550nm, amplified by a Semiconductor Optical Amplifier (SOA). The laser light was coupled into a balanced MZI with a 50:50 beam splitter at the input. As shown in Fig.4.7, the lower arm of the MZI corresponds to the LO and the top arm (the signal) corresponds to the Preparation stage of our experiment. In the signal arm, one of two quantum states $|\alpha\rangle$ or $|-\alpha\rangle$ was prepared based on the input given by the FPGA. These states were then attenuated to the desired mean photon level via a VOA. We note that this attenuation was the only part of the setup that required comprehensive characterization and monitoring, thus the SDI designation. The light coming from both arms was then recombined at the output of the MZI and coupled into two fast integrated PDs. Outside the chip, the detectors were bonded in a differential configuration, and the output signal was discriminated between positive and negative values with a raising edge discriminator. This corresponded to the discrimination of the respective quadrature values, generating the binary b output.

The output b and input x information were processed by the FPGA and forwarded to a PC that, in real-time, evaluated the conditional probabilities $P(b|x)$ and calculated the extraction rate certified by the semi-device independent protocol. The monitoing in real-time of these probabilities was crucial to ensure the security and self-testing probabilities of the system. If the probabilities deviate from the expected values, it could indicate a malfunction or tampering with the device,

which could compromise the randomness of the output. In the previous fiber-based prototype by Rusca et al. [29], a digital optimization setup was used to stabilize the phase of the interferometer. However, due to the integrated nature of our set-up, this phase stabilization can be automatically achieved through the design of the photonic integrated circuit.

4.3.4 Interfacing Electronics

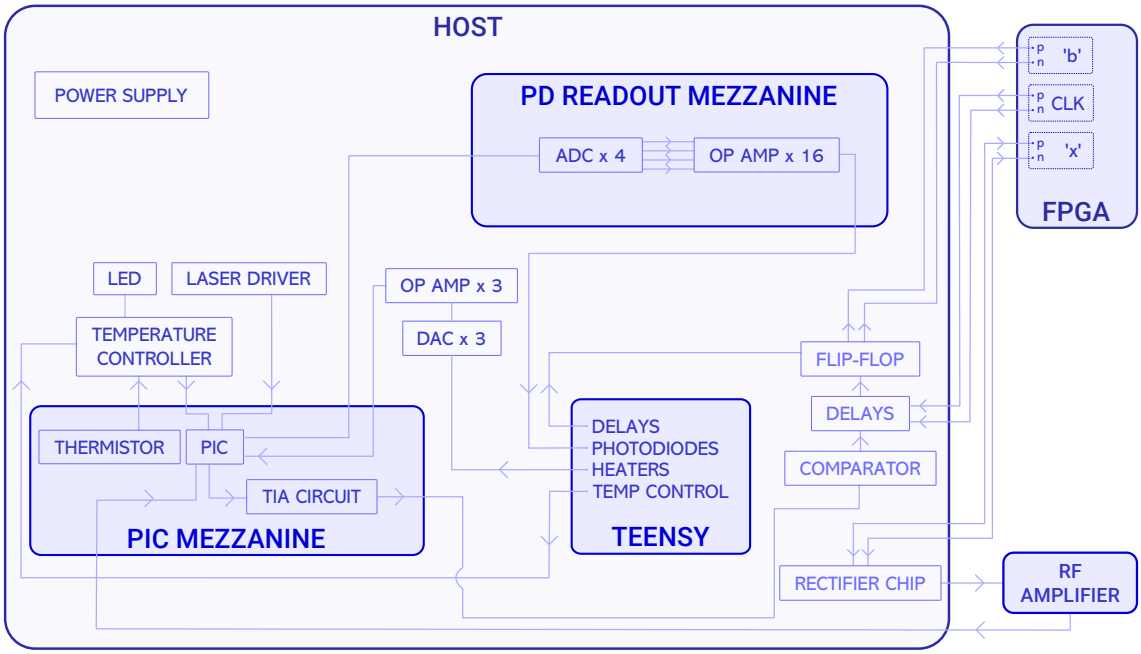


Figure 4.8: Block diagram illustrating the electrical architecture of the QRNG control system. The system integrates the microcontroller units (PIC Mezzanine, Teensy), analog-to-digital and digital-to-analog converters, temperature regulation components, photodiode detection circuitry, and signal processing elements. This custom electrical setup enables the full control of the self-testing SDI-QRNG.

As with the integrated QKD system, we developed a custom electronic control system around the QRNG PIC using three PCBs designed for both modularity and flexibility: The *PIC mezzanine*, *PD ReadOut mezzanine*, and *host* PCB, connected via high-speed FX23-60S-0.5SV floating connectors. The full electronic circuit is shown in Figure 4.8. The PIC mezzanine contained the PIC and TIA circuit, the PD ReadOut mezzanine contained the Op-Amps and Analog-to-Digital Converters (ADCs) connected to the slow PDs, and the host PCB contained the microcontroller, TEC driver modules, power supplies, and other components for signal processing. The host PCB also provided the power supply and data communication interfaces

for the entire system. The three-board architecture allowed easy PIC swapping and independent adjustment of the photodiode's (PD) amplifier gain without full system disassembly. Specifically, we separated out the ADC and OP-AMP components onto a dedicated board to minimize the risk of chip damage during gain tuning.

The whole system operated from a single 12V power supply. After the usual high frequency-noise filtering, the input voltage was converted through multiple DC/DC converters to generate the various voltage levels required by system components, like the microcontroller (Teensy 4.1), PIC, amplifiers, temperature controllers etc.

The PIC mezzanine received the set-point information for its optical components: V_{bias} for the BHD, and supply current for phase-shifter, laser and SOA. Conversely, it sent the PD's current measurements back through the host PCB to the PD ReadOut mezzanine where they are converted to digital signals using 16-bit ADCs. The digital signals were then processed by the microcontroller for monitoring.

The aforementioned microcontroller managed multiple system functions: setting phase-shifter currents ($0-18$mA), interfacing with the TEC driver (MTD415L), and transmitting the user-defined delay signals to the FPGA. All communications with the microcontroller were done via 16-bit SPI connections.

A comparator, delay chip and flip-flop, were combined to create a leading edge discriminator. The comparator compared the input signal to a user-defined threshold voltage set dependent the TIA circuit. The delay chip determined the precise timing for signal sampling. The flip-flop then sampled the signal at the specified time, generating a differential signal that is read by the FPGA.

From Figure 4.8, we see the FPGA is responsible for generating the random bit sequence x that is sent to the PIC and the CLK signal used for discrimination. It also collects the processed B signal from the PIC. Both x and B are then sent to the PC for further analysis. The differential signal x from the FPGA was converted to a single-ended signal on the host PCB to mitigate potential signal integrity issues at high speeds. The processed signal was then connected via a SMA cable to the PIC mezzanine where the PIC is bonded.

4.3.5 Photonic Integrated Circuit Design and Characterization

In order to achieve such extensive integration of the optical components, a number of key design choices were made. For the PIC material selection, InP was chosen due to its direct bandgap property, which allows for efficient light emission and

absorption at the relevant wavelengths that enables on-chip light sources, optical amplifiers, signal modulators and detectors in the C-band [221]. The bandgap of InP can also be engineered to cover a wide range of wavelengths, where different regions of the same chip can have different bandgaps optimized for specific functions: 1100 nm bandgap for low-loss waveguides, 1550 nm bandgap for gain, 1650 nm bandgap for photodetectors [222]. InP also has a wider bandgap than other common active materials like Gallium Arsenide (GaAs), which allows for operation at higher frequencies [222]. The bandgap has direct effects in electron mobility and electron saturation velocity, both increasing with an increase in bandgap. Wider bandgap materials also allow for the devices to operate at higher voltages and electric fields, which can drive electrons to higher speeds. InP's bandgap is considered moderately wide (1.35ev), which gives it a good balance of high-speed performance, efficient light emission/detection in the NIR, and relatively low operating voltages, especially when compared with wide-bandgap materials [222]. The PIC was fabricated using standard foundry recipes: Epitaxial growth steps; lithographic masks; electron beam lithography for grating structures; selective area growth for different functional regions; and passivation and metallization.

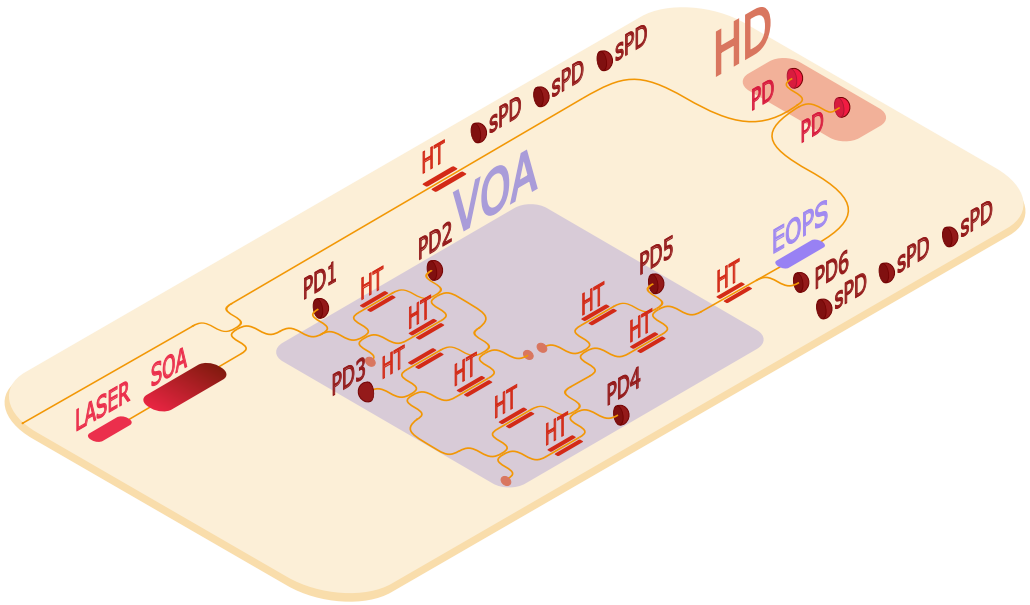


Figure 4.9: Schematic of QRNG Photonic Integrated Circuit. SOA - Semiconductor Optical Amplifier; EOPS - Electro-Optic Phase Shifter; HT - Heater; PD - Photodiode; HD - Homodyne Detection; VOA - Variable Optical Attenuator; sPD - Stray Photodiode.

In the following sections, we will first examine the characteristics and perfor-

mance of the integrated laser and the SOA, followed by the characterization of the VOA and the Electro-Optic Phase Shifter (EOPS). Finally, we will discuss the effect of stray light in the chip on system performance. All mentioned specifications were provided by the chip manufacturer (Fraunhofer HHI).

4.3.5.1 Laser and Semiconductor Optical Amplifier

We used a Distributed Feedback (DFB) laser based on a Multiple Quantum Well (MQW) 2.5µm wide ridge-waveguide structure. The MQWs generate photons through the recombination of electrons and holes confined in the quantum wells. The DFB grating, etched onto and on top of the MQWs, is the periodic structure responsible for providing optical feedback to the laser, ensuring that the laser operates at a single wavelength. This wavelengths can be tuned by a 45Ω heater placed in the DFB, with supported wavelengths: 1525nm, 1536nm, 1548nm, 1558nm, 1568nm, 1578nm; with an accuracy of 5nm and side-mode suppression greater than 50dB and a linewidth smaller than 5MHz for all possible wavelengths. The laser is designed to output 10mW at an operating current of 150mA.

Similarly to the laser, the SOA was based on a MQW ridge-waveguide structure. The structure included a ridge-waveguide for lateral light confinement, MQWs in the active region for optical gain, and electrical contacts for current injection. The quantum wells enhance carrier-photon interactions and improve carrier confinement. Upon forward-biasing, injected carriers populate the quantum wells, and incoming photons stimulate coherent light emission, amplifying the input signal. MQW-based SOA structures typically offer higher differential gain, lower threshold current, better temperature stability, and wider gain bandwidth when compared to bulk semiconductor optical amplifiers that use a thicker, homogeneous active region rather than discrete quantum wells [223, 224]. This device operated in a $1350 - 1625$nm range, achieving a nominal gain of 25dB/mm for wavelengths ranging from 1540nm to 1575nm, biased between $0.15 - 0.35$mA, and saturating at around 13dBm. When unbiased, it had a propagation loss of -30dB/mm.

With reference to Fig. 4.10, to characterize the laser and SOA, in PD1, we measured the output power of the source as a function of the current applied to the laser, for different current values applied to the SOA. Given the first BS has a splitting ratio of 50:50, the power measured in PD1 would be 1/4 of the total power. Details of the characterization of the photodiodes can be seen in A The results are shown in Fig. 4.10. We can see, for sufficiently large I_{SOA}, the lasing threshold around I_L=0.20mA. From then on, the power increases linearly with I_L,

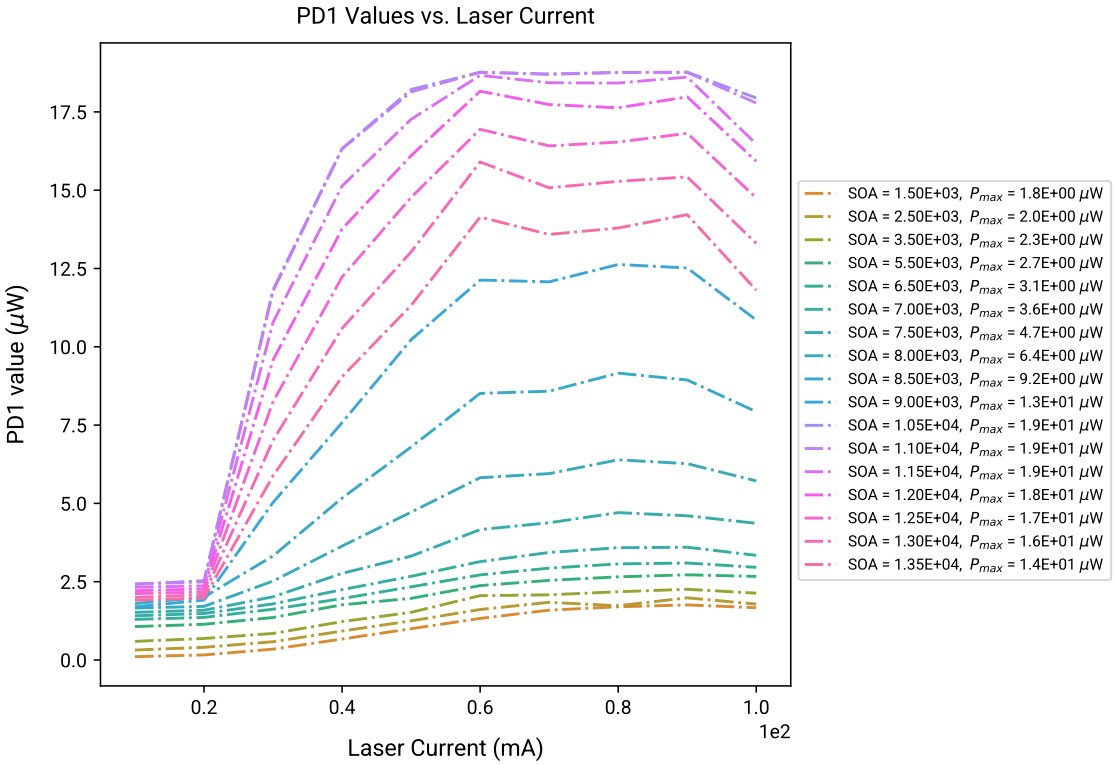


Figure 4.10: Plot of the characterization measurement of the Laser and SOA.

until it plateaus at around 0.60mA and for I_L>0.90mA, the power starts to decrease. A similar pattern can be observed with the supply current to the SOA, though without a plateau, for I_{SOA}>11kAu, the power starts to decrease. The maximum power achieved was 12.5μW for I_L between 0.6mA and 0.9mA and I_{SOA} = 11kAu.

We noticed that the measured values deviate significantly from the manufacturer's, which can be attributed to several critical manufacturing challenges still fairly common in InP-based devices. The behavior highlighted at operating points beyond I_L =90mA and I_{SOA}=11kAu likely stems from imperfect material quality and local thermal management issues [225]. In InP devices, crystal defects and impurities can lead to non-radiative recombination that directly impacts the device's output power. Furthermore, the complete signal suppression at higher currents may suggest thermal effects that are typical in InP-based devices, where heat generation can significantly affect performance if not properly managed. This thermal sensitivity, combined with issues in optical confinement - non-negligible in this chip and will be addressed later in this chapter - and the interface quality between the laser and SOA sections, could also explain the unexpected behavior at higher injection currents. The interaction between all these factors - material defects, thermal effects, and imperfect isolation between the laser and SOA - creates a complex interplay that

requires detailed investigation with the proper material characterization techniques to understand the cause of the performance discrepancy.

Unfortunately, the power achieved using the integrated laser and the SOA were not sufficiently high to observe any quantum noise effects at the BHD. Having experienced this in previous iterations of the chip, a second waveguide was designed to bypass the laser and SOA sections, allowing for external light to be coupled directly into the input of the MZI, although no additional structures was added to improve coupling efficiency. Using a edge-coupled external laser source with 40mW, we were able to achieve up to 220µW of optical power at PD1, equating to a total coupled power about 900µW, corresponding to a $\sim 2\%$ coupling efficiency. While this alternative is not ideal, it allowed us to proceed with the characterization of optical components downstream. The results presented in the following sections were obtained using this external laser source.

4.3.5.2 Variable Optical Attenuator

The VOA was composed of four cascaded MZIs. Each stage featured a photodiode for monitoring the output power, while both arms of each MZI were equipped with heaters (TOPSs) to adjust the phase of the transmitted light. These heaters operated based on thermo-optic effects. The 2µm wide waveguide was covered by a Silicon Nitride (SiN) isolation layer. Positioned above this isolation layer was a thin metal with a resistance of 185Ω/mm, which functioned as the heater element. Current passing through the metal generated heat through the Joule effect, which was then conductively transfered to the waveguide.

The heaters were designed to provide a response of $\mathrm{I}_\pi \times \mathrm{L} = 200\mathrm{mA} * \mathrm{mm}$. The primary thermo-optic effect at play was the temperature-dependent change in refractive index. For InP, the thermo-optic coefficient ($\frac{\delta n}{\delta T}$) is approximately $2.01 \times 10^{-5} K^{-1}$. Additionally, the thermal expansion of the InP material slightly increases the physical path length of the waveguide. While smaller than the refractive index effect, this thermal expansion also contributes to the overall phase change [226].

To characterize the attenuation profile of the VOA, we measured the output power at each MZI stage (PD_n) as a function of the current applied to the respective heaters (H_n). We chose to characterize each MZI stage individually to better understand their individual performance characteristics and due to practical time constraints. A measurement set for two heaters requires approximately 1 hour; characterizing all heaters simultaneously would have required approximately 1.8×10^8 hours due to the exponential scaling of the measurement parameter space. A de-

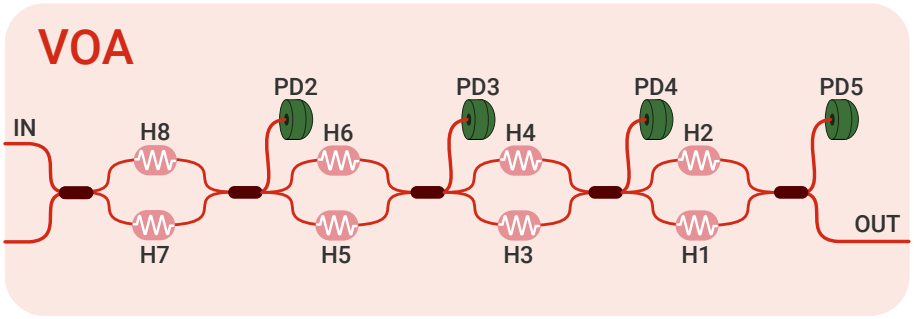


Figure 4.11: Simplified schematics of the components of the VOA.

tailed characterization and the resulting plots can be found in Appendix B. Table 4.1 presents the performance of each individual MZI stage. The measured power values are normalized to the power in PD1 to account for fluctuations in the input coupled light, ensuring consistent comparison across measurements. The first MZI stage (PD2) demonstrated the highest attenuation, with the performance progressively diminishing at subsequent stages. For each measurement beyond the first stage, the transmission in all preceding MZI stages was maximized (attenuation minimized) to isolate and characterize the performance of each individual stage. The attenuation measurements had a 600nW noise floor that came from stray light, which will be adressed later in this section.

PD_n	**PD2**	**PD3**	**PD4**	**PD5**
Attenuation (dB)	25.8	12.3	7.7	4.6
Max. Power (µW)	27.1	12.1	6.0	2.5
Min. Power (µW)	0.6	0.7	1.0	0.9

Table 4.1: Attenuation performance of each MZI stage in the variable optical attenuator

These results were somewhat unexpected, as we anticipated similar attenuation values across all MZI stages given that they are identical in design. The observed trend could suggest that the measured attenuation was limited by the optical power one can couple into each stage, meaning we were out of the power range of the photodiodes. However, if this were the case, we would expect to see a distortion of the signal. As the measured power starts dropping, we would expect a deviation from the sine function in the form of a truncation for negative values. Instead, what was observed was merely an attenuation of the maximum power without any distortion of the signal shape. This can indicate that the splitting ratios may not be perfectly balanced at 50:50, and as we progress down the cascaded architecture,

this imbalance becomes more apparent, affecting the overall visibility.

Another contributing factor that must be taken into acount, especially since we are considering an integrated setup, is the crosstalk between the different heaters in each MZI stage. The thermal crosstalk between adjacent heaters can lead to unintended phase shifts in neighboring MZI stages, which can affect the visibility of the interferometers. This phase shift would not be seen in the first stage, since its heaters would just compensate for this. However, starting from the second stage, the heater values of the previous stages would affect not only the incoming phase, but also the power going into the current stage. This would be more apparent the more stages we have.

4.3.5.3 On-Chip Stray Light

Another issue known from previous iterations of the chip was stray light in the PIC. In our context, stray light refers to unwanted light that scattered from components and discontinuities and/or leakage from waveguide modes. In our specific application, this stray light can prove to be quite troublesome as different light powers across the chip can create unbalances in the homodyne detection and mischaracterization of the VOA's attenuation, as well as couple into monitoring detectors, leading to incorrect signal power measurements during random number generation. All of these effects can compromise the certification of randomness.

As this stray light effect was known before the development of the latest chip version, several design choices were made to mitigate the issue. For example, absorptive materials were placed in strategic locations to absorb any stray light that may have escaped the waveguides, and "stray" photodiodes (not connected to waveguides) were placed along the edges of the PIC for monitoring the stray light (see Figure 4.9). Additionally, some of the PDs are shielded with the same absorptive material used for preventing light leakages from the chip's active components.

	S1	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**
Power (µW)	5.1	6.9	6.2	5.5	5.3	6.4	7.8	8.2

Table 4.2: Power measured at stray photodiodes.

With the external laser input power set to 40mW, correponding roughly to 80µW of coupled power, we measured the optical power in all stray photodiodes. From the measurements we can see there is no difference whether the PD is shielded or not, indicating either a wrong choice of material, improper shielding placement, or

insufficient shielding. Another interesting observation was that the power detected by the stray photodiodes was larger than the power measured at the photodiode placed at the output of the VOA. This suggests that most of the stray light was not being coupled back into the waveguide, but rather that a significant portion was being lost to the substrate or scattered in other directions, which ended up being beneficial for the QRNG process.

4.3.5.4 Electro Optical Phase Shifter

Finally, a particularly important component of the chip was the EOPS. This device was responsible for rapidly switching phase difference (from 0 to π)between the two states in the signal arm of the MZI. For our chip, we had a twin Electro-Optic Modulator (EOM) with a Traveling Wave Electrode (TWE) in a push-pull configuration. This works by modulating the phase of light traveling through the waveguide using an electrical signal that propagates along the electrode at the same speed as the optical signal. The electric field changes the refractive index of the material and this in turn alters the phase of the light traveling through the waveguide.

According to the manufacturer, the phase modulator provides a phase shift $V_\pi \times L = 8V \times mm$. For us this corresponded to $V_\pi = 16V$ given the length of our bias section is 500µm. During testing, it became clear that the coupled RF signal was leaking into the ground plane of both the chip and the PCB. Several potential causes were considered. These included poor chip design, incorrect bonding and issues with the PCB. To troubleshoot, all bond wires were first removed from the chip to isolate the possible sources of the leakage. Without the chip bonded, the ground signal was clean, confirming that the PCB was not the source of the problem. The next thing to look into was the IM bondings as they could be creating an antenna effect and thus cause the leakage. During rebonding we shortened the bond length, added an additional bonding to the ground to reduce coupling capacitance [227], and increased physical separation between the signal and ground bondings. Despite these measures, the issue persisted.

These results suggest that the leakage originated from the large gold pads on the chip that connect the RF signal into the IM. If not carefully designed, such pads can introduce a significant pad-to-substrate parasitic capacitance that are antenna-like, coupling the signal into the ground plane [228].

Due to the semi-insulating nature of the InP substrate, the RF fields can leak through the substrate. This is especially problematic without isolation trenches, guard rings, or proton-implant isolation [229, 230]. Unfortunately, this meant that

the chip design was found to be unsable in its current state: All previously mentioned deviations from specifications could be worked around, but this issue was both fundamental to the chip's design and added severe noise to the ground reference of the entire PIC, as well as PCB. Further review of the chip design and fabrication recipe is necessary to address this issue.

4.4 Discussion and Outlook

In this section I will provide a critical reflection on the experimental work carried out, with the aim of understanding why the setup did not achieve the intended performance and what can be learned from the process. The discussion begins with an analysis of the experimental shortcomings and system's severe limitations, followed by potential future improvements that could address these issues.

The system's performance was hindered by several factors that prevented it from even being usable. As mentioned before, these included fabrication issues that resulted in lower-than-expected laser and SOA output power, which was insufficient to observe quantum noise at the BHD. The EOPS exhibited significant RF signal leakage into the ground plane, rendering the entire system unusable when an external modulating signal was connected. Though less problematic, the stray light within the substrate also degraded the system SNR.

At the debugging and engineering level, a number of challenges were also observed. The chip design restricted access to many critical nodes, making the debugging both time consuming and incomplete. The limited input access to the chip, and no optical output access, meant that issues could not always be localized, forcing the reliance on indirect indicators that were difficult to interpret.

The main limitation encountered however, was the lack of fabrication reliability. The growth defects and yield problems meant that most devices did not function within specification. The packaging also introduced complications, particularly with external optical coupling, where the lack of waveguide structures for mode-matching significantly increased mechanical sensitivity of the coupling alignment.

A key consideration when discussing future directions of this work is the choice of integration platform. InP offers a direct bandgap, making it well-suited for active photonic functions such as lasers, modulators, and photodetectors. Its high electron mobility also enables high-speed operation. Hence why it was chosen in the first place. However, InP fabrication is relatively costly, device yield can be inconsistent, and large-scale integration remains challenging when compared to more

mature platforms. By contrast, silicon photonics benefits from the vast infrastructure of CMOS fabrication, offering high yield, low cost, and excellent scalability. Passive components such as waveguides, couplers, and filters are highly efficient. Yet, because silicon is an indirect bandgap material, it cannot support efficient light sources [231, 232]. The use of silicon would require hybrid integration with III-V materials for lasers and homodyne detectors. This adds complexity but provides a pontentially promising compromise.

Many improvements can be implemented in future designs to address the aforementioned issues, namely:

- Providing more access points for both light and electrical signals to facilitate easier debugging. This is especially crucial for optical signals, where direct access, and bypassing of elements can help identify issues more quickly, even if this comes at the expense of compactness.
- Exploring alternative waveguide designs to bypass problematic elements and minimize stray light.
- Implementing significant shielding and grounding practices to reduce noise interference.
- Ensuring proper electrical isolation near the IM to prevent RF leakage.
- Considering a hybrid integration approach to benefit from the maturity of silicon technologies and the direct bandgap of III-V materials, potentially improving performance.

Despite these challenges, we successfully integrated most of the crucial electronic and optical components for a compact and scalable self-testing SDI-QRNG device. However, it is clear that a more considered and cautious workflow must be adopted with regards to PIC design and especially fabrication. Without this, even seemingly well-designed, high-performance prototypes risk not being fully characterized due to hidden or unobservable failure modes, rendering the devices unusable in any application. With the right improvements in the PIC development, we hope that our platform can more closely approach being ready to deploy in applications requiring reliable and certifiable entropy generation.

Chapter 5

Conclusion and Outlook

In this thesis, we have addressed three interconnected pillars of quantum communication technologies: single-photon detection, QKD, and QRNG. Each of these elements represents both a challenge and an opportunity in the path towards practical and scalable secure communication systems. Correspondingly, the work presented here contributes to bridging the gap between laboratory feasibility and field-deployable solutions, through advances in single-photon detector design, integrated QKD system development, and prototype integrated QRNG implementations.

Chapter 2 focused on the role of single-photon detectors in enabling high-speed quantum communication. While SNSPDs define the current state-of-the-art in single-photon detection performance, their cryogenic requirements limit widespread deployment. In contrast, SPADs offer a practical path towards compact and cost-effective systems. We investigated the performance of dual-anode SPADs, a novel detector architecture designed to mitigate afterpulsing effects and improve timing performance. Through detailed characterization, including dark count rates, photon detection efficiency, afterpulsing probability, and jitter, I demonstrated their suitability for high-speed quantum applications such as asynchronous heralded single-photon sources. This work highlights both the promise and the challenges of SPADs as room-temperature alternatives for single photon detection in the NIR.

Chapter 3 presented the development of an integrated QKD prototype system, motivated by the need to move beyond proof-of-principle demonstrations, and towards deployable platforms. Building on previous efforts of the group, we did a redesign and miniaturization of system components, including compact custom electronics, an improved receiver chip with tunable basis selection, and the replacement of external bulk filtering with on-transmitter-chip ring resonators, yielding significant size reduction and improved performance. Furthermore, the susceptibility of

time-bin encoded schemes to chromatic dispersion in fiber was addressed by adapting the system repetition rate, thereby eliminating the need for bulky dispersion-compensating fibers. The prototypes fit within standard 19in 3U and 1U rack units, marking a significant step towards compact and practical QKD systems. The field deployment of this system demonstrated the feasibility of integrated solutions while also revealing the limitations that must be addressed before large-scale adoption, including further improvements in the detector technologies.

Chapter 4 turned to the critical requirement of true randomness for cryptographic applications. We examined a SDI-QRNG scheme based on homodyne detection and optimized the associated high-speed electronics, including a custom transimpedance amplifier tailored for high-bandwidth balanced homodyne detection. Although the development process terminated prematurely near the final characterization stages with the integrated EOM, a lot was learned about the practical challenges of developing integrated QRNG systems with detectors and lasers on-chip. The work underscored the importance of careful design, fabrication, and characterization of both optical and electronic components to achieve high-quality randomness generation. At the same time, the work emphasized the importance of integration for future QRNG systems, particularly their potential for direct embedding in photonic integrated circuits and larger communication platforms. Such integration will be essential for moving QRNGs from research prototypes to components that can be seamlessly deployed within consumer and industrial technologies.

Looking Ahead

Taken together, the three strands of this thesis suggest that the maturation of quantum communication will depend less on singular breakthroughs and more on the careful convergence of multiple disciplines. Semiconductor physics, photonic integration, electronics, and cryptographic theory each bring different constraints and possibilities. Bridging them requires not only technical innovation but also a sensitivity to practical circumstances: cost, scalability, and interoperability. As the field transitions to real-world deployment, success will increasingly be measured not only in terms of bit rates and distances, but also in terms of reliability, ease of use, and the ability to coexist with existing communication infrastructure.

For SPAD technologies, ongoing research into new materials such as Ge and GeSn alloys, combined with advanced quenching electronics, could further suppress afterpulsing and extend sensitivity into the near-infrared regime. For integrated QKD, the next step will be adding a wavelength division multiplexing layer which

will allow for the implementation of multiple quantum channels in a single optical fiber, significantly increasing the overall capacity of the system. The main challenge will be to manage crosstalk and maintain low error rates in the presence of multiple channels. QKD systems will also benefit from advances in integrated photonics, such as the incorporation of on-chip sources and detectors, along with the development of robust error correction and privacy amplification modules that can operate in real time at the required high key rates. Finally, for QRNG, combining SDI certification with fully integrated implementations offers a clear path towards compact, low-power, and certifiably secure randomness sources.

Beyond these technical directions, the broader challenge remains the transition from laboratory demonstrations to scalable industrial systems. This requires not only advances in device performance but also component cost reductions, standardization, and interoperability with existing telecommunication infrastructures. It is important not to disregard the fact that the adoption of QKD and QRNG technologies will inevitably depend as much on user trust, regulatory frameworks, and economic incentives as on physical performance metrics. As such, future progress in quantum communication must be pursued in close dialogue between physicists, engineers, cryptographers, and policymakers.

Final Remarks

In conclusion, this thesis has sought to contribute towards the larger vision of practical quantum communication. From improving the performance of semiconductor single-photon detectors, to engineering a more compact QKD prototype, to exploring certifiable quantum randomness, the results collectively highlight the opportunities and the remaining challenges in bringing quantum technologies out of the laboratory. More broadly, they reflect the shift in quantum communication from a primarily scientific pursuit to one with clear real-world relevance. The promise of theoretically secure communication is no longer a distant theoretical ideal but an emerging reality, one that will require continued multi-disciplinary collaboration to fully realize.

Bibliography

[1] R. Sax, A. Boaron, G. Boso, S. Atzeni, A. Crespi, F. Grünenfelder, D. Rusca, A. Al-Saadi, D. Bronzi, S. Kupijai, H. Rhee, R. Osellame, and H. Zbinden, *High-speed integrated QKD system*, Photon. Res. **11**(6), 1007–1014 (Jun 2023).

[2] J. Ling, *Brian Hayden. The power of ritual in prehistory: Secret societies and origins of social complexity (Cambridge: Cambridge university press, 2018, 410pp., 64 illustr., hbk, ISBN 978-1-10-857207-1)*, Eur. J. Archaeol. **22**(4), 599–603 (Nov. 2019).

[3] S. Blakely, *5 Toward an Archaeology of Secrecy: Power, Paradox, and the Great Gods of Samothrace*, Archeological Papers of the American Anthropological Association **21** (03 2011).

[4] Suetonius, *Lives of the Caesars, Volume I: Julius. Augustus. Tiberius. Gaius Caligula*, volume 31 of *Loeb Classical Library*, Harvard University Press, Cambridge, MA, 1914.

[5] A. Hodges, *Alan Turing: The Enigma*, Princeton University Press, Princeton, NJ, 2014, Updated edition.

[6] R. Namase, Cybercrime Statistics 2025: Rising AI Threats & Global Impact, July 2025, Accessed: 2025-09-05, https://sqmagazine.co.uk/cybercrime-statistics.

[7] V. Lyskoit, Cybersecurity statistics and facts you need to know, January 2025, Accessed: 2025-09-05, https://nordvpn.com/blog/cybersecurity-statistics.

[8] C. Hale, A shocking number of businesses don't have cyber insurance - here's why you should fix that immediately, August 2025, Accessed: 2025-09-05, https://www.techradar.com/pro/security/a-shocking-number-of-

businesses-dont-have-cyber-insurance-heres-why-you-should-fix-that-immediately.

[9] R. Overbeck and N. Sendrier, Code-based cryptography, in *Post-Quantum Cryptography*, pages 95–145, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[10] R. L. Rivest, A. Shamir, and L. Adleman, *A method for obtaining digital signatures and public-key cryptosystems*, Commun. ACM **26**(1), 96–99 (Jan. 1983).

[11] P. W. Shor, *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer*, SIAM Journal on Computing **26**(5), 1484–1509 (1997).

[12] W. Diffie, I and M. E. Hellman, New Directions in Cryptography, in *Democratizing Cryptography*, pages 365–390, ACM, New York, NY, USA, Aug. 2022.

[13] W. K. Wootters and W. H. Zurek, *A single quantum cannot be cloned*, Nature **299**(5886), 802–803 (Oct. 1982).

[14] J. L. Park, The concept of transition in quantum mechanics, https://philpapers.org/rec/PARTCO-16, 1970, Accessed: 2025-9-9.

[15] C. H. Bennett and G. Brassard, *Quantum cryptography: Public key distribution and coin tossing*, Theoretical Computer Science **560**, 7–11 (Dec. 2014).

[16] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, *Fundamental limits of repeaterless quantum communications*, Nature Communications **8** (2017).

[17] A. Boaron, G. Boso, D. Rusca, C. Vulliez, C. Autebert, M. Caloz, M. Perrenoud, G. Gras, F. Bussières, M.-J. Li, D. Nolan, A. Martin, and H. Zbinden, *Secure Quantum Key Distribution over 421 km of Optical Fiber*, Phys. Rev. Lett. **121**, 190502 (Nov 2018).

[18] F. Grünenfelder, A. Boaron, G. V. Resta, M. Perrenoud, D. Rusca, C. Barreiro, R. Houlmann, R. Sax, L. Stasi, S. El-Khoury, E. Hänggi, N. Bosshard, F. Bussières, and H. Zbinden, *Fast single-photon detectors and real-time key distillation enable high secret-key-rate quantum key distribution systems*, Nature Photonics **17**, 422–426 (5 2023).

[19] G. V. Resta, L. Stasi, M. Perrenoud, S. El-Khoury, T. Brydges, R. Thew, H. Zbinden, and F. Bussières, *Gigahertz detection rates and dynamic photon-number resolution with superconducting nanowire arrays*, Nano Lett. **23**(13), 6018–6026 (July 2023).

[20] M. Grundmann, *The Physics of Semiconductors*, Springer Berlin, Heidelberg, 2nd edition edition, 2010.

[21] F. Thorburn, X. Yi, Z. M. Greener, J. Kirdoda, R. W. Millar, L. L. Huddleston, D. J. Paul, and G. S. Buller, *Ge-on-Si single-photon avalanche diode detectors for short-wave infrared wavelengths*, JPhys Photonics **4**(1), 012001 (Jan. 2022).

[22] P. Vines, K. Kuzmenko, J. Kirdoda, D. C. S. Dumas, M. M. Mirza, R. W. Millar, D. J. Paul, and G. S. Buller, *High performance planar germanium-on-silicon single-photon avalanche diode detectors*, Nature Communications .

[23] A. Giunto, *GeSn as next-generation material for short-wave infrared single-photon detection*, PhD thesis, EPFL, Lausanne, 2023.

[24] S. Cova, S. Member, A. Lacaita, G. Ripamonti, and S. Cova, *Trapping Phenomena in Avalanche Photodiodes on Nanosecond Scale H o l d Off Time Filling P u l s e*, IEEE ELECTRON DEVICE LETTERS **12**.

[25] A. C. Giudice, M. Ghioni, S. Cova, and F. Zappa, A process and deep level evaluation tool: afterpulsing in avalanche junctions.

[26] G. Acconcia, I. Rech, A. Gulinatti, and M. Ghioni, *High-voltage integrated active quenching circuit for single photon count rate up to 80 Mcounts/s*, Opt. Express **24**(16), 17819–17831 (Aug. 2016).

[27] D. Rusca, A. Boaron, F. Grunenfelder, A. Martin, and H. Zbinden, *Finite-key analysis for the 1-decoy state QKD protocol*, Applied Physics Letters **112**(17) (2018).

[28] C. Park, S.-B. Cho, C.-Y. Park, S. Baek, and S.-K. Han, *Dual anode single-photon avalanche diode for high-speed and low-noise Geiger-mode operation*, Opt. Express **27**(13), 18201–18209 (Jun 2019).

[29] D. Rusca, H. Tebyanian, A. Martin, and H. Zbinden, *Fast self-testing Quantum Random Number Generator based on homodyne detection*, (4 2020).

[30] A. Boaron, B. Korzh, R. Houlmann, G. Boso, D. Rusca, S. Gray, M. J. Li, D. Nolan, A. Martin, and H. Zbinden, *Simple 2.5 GHz time-bin quantum key distribution*, Applied Physics Letters **112** (4 2018).

[31] Z. L. Yuan, B. E. Kardynal, A. W. Sharpe, and A. J. Shields, *High speed single photon detection in the near infrared*, Appl. Phys. Lett. **91**(4), 041114 (July 2007).

[32] J. Zhang, R. Thew, C. Barreiro, and H. Zbinden, *Practical fast gate rate InGaAs/InP single-photon avalanche photodiodes*, Appl. Phys. Lett. **95**(9), 091103 (Aug. 2009).

[33] L. C. Comandar, B. Fröhlich, J. F. Dynes, A. W. Sharpe, M. Lucamarini, Z. Yuan, R. V. Penty, and A. J. Shields, *GHz-gated InGaAs/InP single-photon detector with detection efficiency exceeding 55% at 1550 nm*, arXiv [quant-ph] (Dec. 2014).

[34] L. C. Comandar, B. Fröhlich, J. F. Dynes, A. W. Sharpe, M. Lucamarini, Z. L. Yuan, R. V. Penty, and A. J. Shields, *Gigahertz-gated InGaAs/InP single-photon detector with detection efficiency exceeding 55% at 1550 nm*, J. Appl. Phys. **117**(8), 083109 (Feb. 2015).

[35] K. A. Patel, J. F. Dynes, A. W. Sharpe, Z. L. Yuan, R. V. Penty, and A. J. Shields, *Gigacount/second photon detection with InGaAs avalanche photodiodes*, arXiv [physics.ins-det] (Jan. 2012).

[36] N. Namekata, S. Sasamori, and S. Inoue, *800 MHz single-photon detection at 1550-nm using an InGaAs/InP avalanche photodiode operated with a sine wave gating*, Opt. Express **14**(21), 10043–10049 (Oct. 2006).

[37] Y.-Q. Fang, W. Chen, T.-H. Ao, C. Liu, L. Wang, X.-J. Gao, J. Zhang, and J.-W. Pan, *InGaAs/InP single-photon detectors with 60% detection efficiency at 1550 nm*, arXiv [physics.app-ph] **91**(8) (July 2020).

[38] A. Tada and S. Inoue, Sinusoidally Gated InGaAs/InP Avalanche Photodiode with 53% Photon Detection Efficiency at 1550 nm, in *Conference on Lasers and Electro-Optics*, page FTu4C.2, Optica Publishing Group, June 2016.

[39] A. Losev, V. Zavodilenko, A. Koziy, Y. Kurochkin, and A. Gorbatsevich, *Dependence of functional parameters of sine-gated InGaAs/InP single-photon*

avalanche diodes on the gating parameters, IEEE Photonics J. **14**(2), 1–9 (Apr. 2022).

[40] H. Young, A. Ford, and R. Freedman, *Sears & Zemansky's University Physics With Modern Physics*, Addison-Wesley Longman, Incorporated, 2014.

[41] R. H. Haitz, *Model for the electrical behavior of a microplasma*, J. Appl. Phys. **35**(5), 1370–1376 (May 1964).

[42] M. S. Arman Vassighi, Thermal Runaway and Thermal Management, in *Thermal Runaway and Thermal Management of Integrated Circuits*, pages 119–148, Springer Nature, Cham, Switzerland, 2006.

[43] C. Bartolo-Perez, A. Ahamed, A. S. Mayet, A. Rawat, L. McPhillips, S. Ghandiparsi, J. Bec, G. Ariño-Estrada, S. Cherry, S.-Y. Wang, L. Marcu, and M. S. Islam, *Engineering the gain and bandwidth in avalanche photodetectors*, Optics Express **30**, 16873 (5 2022).

[44] G. E. Stillman and C. M. Wolfe, *Avalanche photodiodes*, Semicond. Semimet. (1977).

[45] F. Ceccarelli, G. Acconcia, A. Gulinatti, M. Ghioni, I. Rech, and R. Osellame, *Recent advances and future perspectives of single-photon avalanche diodes for quantum photonics applications*, (10 2020).

[46] A. Panglosse, P. Martin-Gonthier, O. Marcelot, C. Virmontois, O. Saint-Pé, P. Magnan, S. Member, and S. Member, *Dark Count Rate Modeling in Single-Photon Avalanche Diodes*, **2020**.

[47] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, *Progress in silicon single-photon avalanche diodes*, IEEE Journal on Selected Topics in Quantum Electronics **13**, 852–862 (2007).

[48] G. Acconcia, F. Ceccarelli, A. Gulinatti, and I. Rech, *Timing measurements with silicon single photon avalanche diodes: principles and perspectives [Invited]*, Opt. Express **31**(21), 33963–33999 (Oct. 2023).

[49] F. Sun, Y. Xu, Z. Wu, and J. Zhang, *A Simple Analytic Modeling Method for SPAD Timing Jitter Prediction*, IEEE Journal of the Electron Devices Society **7**, 276–281 (2019).

[50] M. Assanelli, A. Ingargiola, I. Rech, A. Gulinatti, and M. Ghioni, *Photon-timing jitter dependence on injection position in single-photon avalanche diodes*, IEEE Journal of Quantum Electronics **47**, 151–159 (2011).

[51] A. Lacaita, S. Cova, M. Ghioni, and F. Zappa, Single-Photon Avalanche Diode with Ultrafast Pulse Response Free from Slow Tails, 1993.

[52] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. Cova, *Progress in Silicon Single-Photon Avalanche Diodes*, IEEE J. Sel. Top. Quantum Electron. **13**(4), 852–862 (2007).

[53] J. Yin, Y. Cao, Y.-H. Li, J.-G. Ren, S.-K. Liao, L. Zhang, W.-Q. Cai, W.-Y. Liu, B. Li, H. Dai, M. Li, Y.-M. Huang, L. Deng, L. Li, Q. Zhang, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, *Satellite-to-ground entanglement-based quantum key distribution*, Phys. Rev. Lett. **119**(20), 200501 (Nov. 2017).

[54] G. A. Howland and J. C. Howell, *Efficient high-dimensional entanglement imaging with a compressive sensing, double-pixel camera*, arXiv [quant-ph] (Dec. 2012).

[55] A. I. Lvovsky, B. C. Sanders, and W. Tittel, *Optical quantum memory*, arXiv [quant-ph] (Feb. 2010).

[56] X.-L. Wang, L.-K. Chen, W. Li, H.-L. Huang, C. Liu, C. Chen, Y.-H. Luo, Z.-E. Su, D. Wu, Z.-D. Li, H. Lu, Y. Hu, X. Jiang, C.-Z. Peng, L. Li, N.-L. Liu, Y.-A. Chen, C.-Y. Lu, and J.-W. Pan, *Experimental ten-photon entanglement*, arXiv [quant-ph] (May 2016).

[57] F. Kaneda, K. Garay-Palmett, A. B. U'Ren, and P. G. Kwiat, *Heralded single-photon source utilizing highly nondegenerate, spectrally factorable spontaneous parametric downconversion*, arXiv [quant-ph] (Mar. 2016).

[58] P. Senellart, G. Solomon, and A. White, *High-performance semiconductor quantum-dot single-photon sources*, Nat. Nanotechnol. **12**(11), 1026–1039 (Nov. 2017).

[59] T. Meany, M. Gräfe, R. Heilmann, A. Perez-Leija, S. Gross, M. J. Steel, M. J. Withford, and A. Szameit, *Laser written circuits for quantum photonics: Laser written quantum circuits*, Laser Photon. Rev. **9**(4), 363–384 (July 2015).

[60] N. Yao, Q. Yao, X.-P. Xie, Y. Liu, P. Xu, W. Fang, M.-Y. Zheng, J. Fan, Q. Zhang, L. Tong, and J.-W. Pan, *Optimizing up-conversion single-photon detectors for quantum key distribution*, Optics Express **28**, 25123 (8 2020).

[61] S. K. Liao, H. L. Yong, C. Liu, G. L. Shentu, D. D. Li, J. Lin, H. Dai, S. Q. Zhao, B. Li, J. Y. Guan, W. Chen, Y. H. Gong, Y. Li, Z. H. Lin, G. S. Pan, J. S. Pelc, M. M. Fejer, W. Z. Zhang, W. Y. Liu, J. Yin, J. G. Ren, X. B. Wang, Q. Zhang, C. Z. Peng, and J. W. Pan, *Long-distance free-space quantum key distribution in daylight towards inter-satellite communication*, Nature Photonics **11**, 509–513 (8 2017).

[62] H. Xu, L. Ma, A. Mink, B. Hershman, and X. Tang, *1310-nm quantum key distribution system with up-conversion pump wavelength at 1550 nm*, Opt. Express **15**(12), 7247–7260 (Jun 2007).

[63] P. A. Hiskett, G. S. Buller, A. Y. Loudon, J. M. Smith, I. Gontijo, A. C. Walker, P. D. Townsend, and M. J. Robertson, *Performance and design of InGaAs /InP photodiodes for single-photon counting at 1.55 microm*, Appl. Opt. **39**(36), 6818–6829 (Dec. 2000).

[64] A. Lacaita, P. A. Francese, F. Zappa, and S. Cova, *Single-photon detection beyond 1 μm: performance of commercially available germanium photodiodes*, Appl. Opt. **33**(30), 6902–6918 (Oct. 1994).

[65] X. Jiang, M. A. Itzler, R. Ben-Michael, and K. Slomkowski, *InGaAsP–InP Avalanche Photodiodes for Single Photon Detection*, IEEE J. Sel. Top. Quantum Electron. **13**(4), 895–905 (2007).

[66] K. Nishida, K. Taguchi, and Y. Matsumoto, *InGaAsP heterostructure avalanche photodiodes with high avalanche gain*, Applied Physics Letters **35**, 251–253 (1979).

[67] M. A. Itzler, R. Ben-Michael, C. F. Hsu, K. Slomkowski, A. Tosi, S. Cova, F. Zappa, and R. Ispasoiu, *Single photon avalanche diodes (SPADs) for 1.5 μm photon counting applications*, Journal of Modern Optics **54**, 283–304 (1 2007).

[68] P. A. Houston, *Growth and characterization of InGaAsP lattice-matched to InP*, J. Mater. Sci. **16**(11), 2935–2961 (Nov. 1981).

[69] C. D. Yerino, B. Liang, D. L. Huffaker, P. J. Simmonds, and M. L. Lee, *Review Article: Molecular beam epitaxy of lattice-matched InAlAs and InGaAs layers on InP (111)A, (111)B, and (110)*, J. Vac. Sci. Technol. B Nanotechnol. Microelectron. **35**(1), 010801 (Jan. 2017).

[70] J. Zhang, M. A. Itzler, H. Zbinden, and J. W. Pan, *Advances in InGaAs/InP single-photon detector systems for quantum communication*, Light Sci. Appl. **4** (May 2015).

[71] T. Lunghi, C. Barreiro, O. Guinnard, R. Houlmann, X. Jiang, M. A. Itzler, and H. Zbinden, *Free-running single-photon detection based on a negative feedback InGaAs APD*, J. Mod. Opt. **59**(17), 1481–1488 (Oct. 2012).

[72] M. M. Hayat, M. A. Itzler, D. A. Ramirez, and G. J. Rees, Model for passive quenching of SPADs, in *Quantum Sensing and Nanophotonic Devices VII*, edited by M. Razeghi, R. Sudharsanan, and G. J. Brown, SPIE, Jan. 2010.

[73] D. A. Ramirez, M. M. Hayat, G. J. Rees, X. Jiang, and M. A. Itzler, *New perspective on passively quenched single photon avalanche diodes: effect of feedback on impact ionization*, receiver array architecture, " Proc. of SPIE **5338**, 56–64 (2004).

[74] R. T. Thew, D. Stucki, J.-D. Gautier, A. Rochas, and H. Zbinden, *Free-running InGaAs/InP avalanche photodiode with active quenching for single photon counting at telecom wavelengths*, arXiv [quant-ph] , 201114 (Jan. 2008).

[75] M. Stipcević, *Active quenching circuit for single-photon detection with Geiger mode avalanche photodiodes*, Appl. Opt. **48**(9), 1705–1714 (Mar. 2009).

[76] A. Giudici, G. Acconcia, I. Labanca, M. Ghioni, and I. Rech, *4 ns dead time with a fully integrated active quenching circuit driving a custom single photon avalanche diode*, Rev. Sci. Instrum. **93**(4), 043103 (Apr. 2022).

[77] J. Liu, Y. Xu, Z. Wang, Y. Li, Y. Gu, Z. Liu, and X. Zhao, *Reducing after-pulsing in InGaAs(P) single-photon detectors with hybrid quenching*, Sensors (Basel) **20**(16), 4384 (Aug. 2020).

[78] F. Lin, C. Jackson, M. Mac Sweeney, M. Manning, M. M. Sheehan, and A. Mathewson, Hybrid CMOS compatible active/passive quenching module, in *Optoelectronic Integrated Circuits VIII*, edited by L. A. Eldada and E.-H. Lee, SPIE, Feb. 2006.

[79] H. Wang, Y. Shi, Y. Zuo, Y. Yu, L. Lei, X. Zhang, and Z. Qian, *High-performance waveguide coupled Germanium-on-silicon single-photon avalanche diode with independently controllable absorption and multiplication*, Nanophotonics **12**(4), 705–714 (Feb. 2023).

[80] L. F. Llin, J. Kirdoda, F. Thorburn, L. L. Huddleston, Z. M. Greener, K. Kuzmenko, P. Vines, D. C. S. Dumas, R. W. Millar, G. S. Buller, and D. J. Paul, *High sensitivity Ge-on-Si single-photon avalanche diode detectors*, Opt. Lett. **45**(23), 6406–6409 (Dec. 2020).

[81] Z. Lu, Y. Kang, C. Hu, Q. Zhou, H.-D. Liu, and J. C. Campbell, *Geiger-mode operation of Ge-on-Si avalanche photodiodes*, IEEE J. Quantum Electron. **47**(5), 731–735 (May 2011).

[82] R. E. Warburton, G. Intermite, M. Myronov, P. Allred, D. R. Leadley, K. Gallacher, D. J. Paul, N. J. Pilgrim, L. J. M. Lever, Z. Ikonic, R. W. Kelsall, E. Huante-Ceron, A. P. Knights, and G. S. Buller, *Ge-on-Si single-photon avalanche diode detectors: Design, modeling, fabrication, and characterization at wavelengths 1310 and 1550 nm*, IEEE Trans. Electron Devices **60**(11), 3807–3813 (Nov. 2013).

[83] G. Ribordy, J. D. Gautier, N. Gisin, O. Guinnard, and H. Zbinden, Automated 'plug and play' quantum key distribution, 1998, https://arxiv.org/abs/quant-ph/9812052.

[84] R. J. Hughes, G. L. Morgan, and C. G. Peterson, *Practical quantum key distribution over a 48-km optical fiber network*.

[85] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, *Quantum Key Distribution over 67 km with a plug&play system*, (2002).

[86] H. Kosaka, A. Tomita, Y. Nambu, T. Kimura, and K. Nakamura, *Single-photon interference experiment over 100 km for quantum cryptography system using a balanced gated-mode photon detector*.

[87] C. Gobby, Z. L. Yuan, and A. J. Shields, Quantum key distribution over 122 km of standard telecom fiber, 2004.

[88] N. Walenta, A. Burg, D. Caselunghe, J. Constantin, N. Gisin, O. Guinnard, R. Houlmann, P. Junod, B. Korzh, N. Kulesza, M. Legré, C. C. W. Lim, T. Lunghi, L. Monat, C. Portmann, M. Soucarros, P. Trinkler, G. Trolliet,

F. Vannel, and H. Zbinden, *A fast and versatile QKD system with hardware key distillation and wavelength multiplexing*, (9 2013).

[89] B. Fröhlich, M. Lucamarini, J. F. Dynes, L. C. Comandar, W. W.-S. Tam, A. Plews, A. W. Sharpe, Z. Yuan, and A. J. Shields, *Long-distance quantum key distribution secure against coherent attacks*, Optica **4**, 163 (1 2017).

[90] B. Korzh, N. Walenta, T. Lunghi, N. Gisin, and H. Zbinden, *Free-running InGaAs single photon detector with 1 dark count per second at 10% efficiency*, Applied Physics Letters **104**(8), 081108 (02 2014).

[91] H. de Riedmatten, I. Marcikic, W. Tittel, H. Zbinden, D. Collins, and N. Gisin, *Long Distance Quantum Teleportation in a Quantum Relay Configuration*, Physical Review Letters **92**(4) (Jan. 2004).

[92] B. P. Williams, J. M. Lukens, N. A. Peters, B. Qi, and W. P. Grice, *Quantum secret sharing with polarization-entangled photon pairs*, Physical Review A **99** (6 2019).

[93] J. Bogdanski, N. Rafiei, and M. Bourennane, *Experimental quantum secret sharing using telecommunication fiber*, Physical Review A - Atomic, Molecular, and Optical Physics **78** (12 2008).

[94] J. F. Dynes, Z. L. Yuan, A. W. Sharpe, and A. J. Shields, *A high speed, post-processing free, quantum random number generator*, Applied Physics Letters **93**(3), 031109 (07 2008).

[95] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, Quantum repeaters based on atomic ensembles and linear optics, 2009, https://arxiv.org/abs/0906.2699.

[96] M. Savanier and S. Mookherjea, *Generating photon pairs from a silicon microring resonator using an electronic step recovery diode for pump pulse generation*, Applied Physics Letters **108** (6 2016).

[97] G. Acconcia, A. Giudici, J. A. Smith, I. Labanca, R. J. Hare, M. Ghioni, and I. Rech, *Toward high-performance SPAD arrays for space-based atmosphere and ocean profiling LiDARs*, Journal of Applied Remote Sensing **15**(1), 017501 (2021).

[98] A. Incoronato, M. Locatelli, and F. Zappa, *Statistical modelling of SPADs for time-of-Flight LiDAR*, Sensors (Basel) **21**(13), 4481 (June 2021).

[99] F. Rutz, A. Wörl, A. Bächle, J. Niemasz, and R. Rehm, Fabrication of InGaAs/InP single-photon avalanche diodes for SWIR active imaging, in *Emerging Imaging and Sensing Technologies for Security and Defence VIII*, edited by G. S. Buller, P. M. Alsing, N. A. Salmon, R. C. Hollins, R. A. Lamb, M. Laurenzis, M. L. Fanto, P. Walther, and F. Gumbmann, volume 12740, page 127400C, International Society for Optics and Photonics, SPIE, 2023.

[100] H. Liu, T. Zhao, and M. Zhang, *OTDR Development Based on Single-Mode Fiber Fault Detection*, Sensors **25**(14) (2025).

[101] E. Slenders, M. Castello, M. Buttafava, F. Villa, A. Tosi, L. Lanzanò, S. V. Koho, and G. Vicidomini, *Confocal-based fluorescence fluctuation spectroscopy with a SPAD array detector*, Light Sci. Appl. **10**(1), 31 (Feb. 2021).

[102] M. Tillmann, F. Koberling, T. Roehlicke, M. Wahl, C. Saudan, H. A. Homulle, I. M. Antolovic, and R. Erdmann, Small SPAD-arrays for confocal fluoresence lifetime imaging (Conference Presentation), in *Multiphoton Microscopy in the Biomedical Sciences XXIII*, edited by A. Periasamy, P. T. C. So, and K. König, volume PC12384, page PC123840G, International Society for Optics and Photonics, SPIE, 2023.

[103] M. A. Itzler, X. Jiang, B. Nyman, and K. Slomkowski, InP-based negative feedback avalanche diodes, in *Quantum Sensing and Nanophotonic Devices VI*, edited by M. Razeghi, R. Sudharsanan, and G. J. Brown, volume 7222, pages 462–473, SPIE, Jan. 2009.

[104] M. Sanzaro, N. Calandri, A. Ruggeri, and A. Tosi, *InGaAs/InP SPAD With Monolithically Integrated Zinc-Diffused Resistor*, IEEE J. Quantum Electron. **52**(7), 1–7 (May 2016).

[105] A. Restelli, J. C. Bienfang, and A. L. Migdall, *Single-photon detection efficiency up to 50% at 1310 nm with an InGaAs/InP avalanche diode gated at 1.25 GHz*, Appl. Phys. Lett. **102**(14), 141104 (Apr. 2013).

[106] C. Scarcella, G. Boso, A. Ruggeri, and A. Tosi, *InGaAs/InP single-photon detector gated at 1.3 GHz with 1.5% afterpulsing*, IEEE J. Sel. Top. Quantum Electron. **21**(3), 17–22 (May 2015).

[107] G. Wu, Y. Jian, E. Wu, and H. Zeng, *Photon-number-resolving detection based on InGaAs/InP avalanche photodiode in the sub-saturated mode*, Opt. Express **17**(21), 18782–18787 (Oct. 2009).

[108] J. Zhang, P. Eraerds, N. Walenta, C. Barreiro, R. Thew, and H. Zbinden, *2.23 GHz gating InGaAs/InP single-photon avalanche diode for quantum key distribution*, arXiv [quant-ph] (Feb. 2010).

[109] Y. Cao, Y. Zhao, Q. Wang, J. Zhang, S. X. Ng, and L. Hanzo, *The evolution of quantum key distribution networks: On the road to the qinternet*, IEEE Commun. Surv. Tutor. **24**(2), 839–894 (Jan. 2022).

[110] Y.-A. Chen, Q. Zhang, T.-Y. Chen, W.-Q. Cai, S.-K. Liao, J. Zhang, K. Chen, J. Yin, J.-G. Ren, Z. Chen, S.-L. Han, Q. Yu, K. Liang, F. Zhou, X. Yuan, M.-S. Zhao, T.-Y. Wang, X. Jiang, L. Zhang, W.-Y. Liu, Y. Li, Q. Shen, Y. Cao, C.-Y. Lu, R. Shu, J.-Y. Wang, L. Li, N.-L. Liu, F. Xu, X.-B. Wang, C.-Z. Peng, and J.-W. Pan, *An integrated space-to-ground quantum communication network over 4,600 kilometres*, Nature **589**(7841), 214–219 (Jan. 2021).

[111] D. Y. He, S. Wang, J. L. Chen, W. Chen, Z. Q. Yin, G. J. Fan-Yuan, Z. Zhou, G. C. Guo, and Z. F. Han, *2.5 GHz Gated InGaAs/InP Single-Photon Avalanche Diode with 44 ps Time Jitter*, Advanced Devices and Instrumentation **4** (Jan. 2023).

[112] Y. Liang, Q. Fei, Z. Liu, K. Huang, and H. Zeng, *Low-noise InGaAs/InP single-photon detector with widely tunable repetition rates*, Photon. Res., PRJ **7**(3), A1–A6 (Mar. 2019).

[113] A. Restelli, J. C. Bienfang, and A. L. Migdall, *Time-domain measurements of afterpulsing in InGaAs/InP SPAD gated with sub-nanosecond pulses*, J. Mod. Opt. **59**(17), 1465–1471 (Oct. 2012).

[114] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, *Avalanche photodiodes and quenching circuits for single-photon detection*, .

[115] A. Ingargiola, M. Assanelli, A. Gallivanoni, I. Rech, M. Ghioni, and S. Cova, Avalanche buildup and propagation effects on photon-timing jitter in Si-SPAD with non-uniform electric field, in *Advanced Photon Counting Techniques III*, volume 7320, page 73200K, SPIE, 5 2009.

[116] G. Boso, A. D. Mora, A. D. Frera, and A. Tosi, *Fast-gating of single-photon avalanche diodes with 200 ps transitions and 30 ps timing jitter*, Sensors and Actuators, A: Physical **191**, 61–67 (2013).

[117] A. Spinelli and A. L. Lacaita, *Physics and numerical simulation of single photon avalanche diodes*, IEEE Trans. Electron Devices **44**(11), 1931–1943 (1997).

[118] J. Bude, N. Sano, and A. Yoshii, *Hot-carrier luminescence in Si*, Physical Review B **45**, 5848–5856 (1992).

[119] A. Tosi, F. Stellari, F. Zappa, and S. Cova, Hot-carrier luminescence: Comparison of different CMOS technologies, in *European Solid-State Device Research Conference*, pages 351–354, IEEE Computer Society, 2003.

[120] L. Selmi, SILICON LUMINESCENCE TECHNIQUES FOR THE CHARACTERIZATION OF HOT-CARRIER AND DEGRADATION PHENOMENA IN MOS DEVICES, 1995.

[121] F. Acerbi, A. Tosi, A. Dalla Mora, M. Anti, and F. Zappa, Experimental characterization of afterpulsing and timing jitter of InGaAs/InP SPAD, in *Optical Components and Materials VIII*, edited by M. J. F. Digonnet, S. Jiang, J. W. Glesener, and J. C. Dries, volume 7934, pages 103–110, SPIE, Feb. 2011.

[122] A. Tosi, C. Scarcella, G. Boso, and F. Acerbi, *Gate-free InGaAs/InP single-photon detector working at up to 100 mcount/s*, IEEE Photonics Journal **5** (2013).

[123] N. Namekata, S. Adachi, and S. Inoue, *Ultra-low-noise sinusoidally gated avalanche photodiode for high-speed single-photon detection at telecommunication wavelengths*, IEEE Photonics Technol. Lett. **22**(8), 529–531 (Apr. 2010).

[124] Y. Liang, E. Wu, X. Chen, M. Ren, Y. Jian, G. Wu, and H. Zeng, *Low-Timing-Jitter Single-Photon Detection Using 1-GHz Sinusoidally Gated InGaAs/InP Avalanche Photodiode*, IEEE Photonics Technol. Lett. **23**(13), 887–889 (July 2011).

[125] D. F. Walls and G. J. Milburn, *Quantum optics*, Graduate texts in physics, Springer Nature Switzerland, Cham, 2025.

[126] M. A. Pereira, M. Wu, A. S. Raja, R. N. Wang, T. Kippenberg, H. Zbinden, T. Brydges, and R. Thew, *Integrated Telecom Wavelength Heralded Single-Photon Source based on GHz gated detectors*, arXiv [quant-ph] (Sept. 2025).

[127] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, *Quantum cryptography*, Rev. Mod. Phys. **74**(1), 145–195 (Mar. 2002).

[128] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner, *Leftover hashing against quantum side information*, IEEE Trans. Inf. Theory **57**(8), 5524–5535 (Aug. 2011).

[129] J. L. Carter and M. N. Wegman, Universal classes of hash functions (Extended Abstract), in *Proceedings of the ninth annual ACM symposium on Theory of computing - STOC '77*, New York, New York, USA, 1977, ACM Press.

[130] M. N. Wegman and J. L. Carter, *New hash functions and their use in authentication and set equality*, J. Comput. Syst. Sci. **22**(3), 265–279 (June 1981).

[131] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, *Experimental quantum cryptography*, J. Cryptology **5**, 3–28 (Feb. 1991).

[132] M. Esmann, S. C. Wein, and C. Antón-Solanas, Solid-state single-photon sources: recent advances for novel quantum materials, 2023, https://arxiv.org/abs/2312.09280.

[133] X.-B. Wang, *Beating the Photon-Number-Splitting Attack in Practical Quantum Cryptography*, Physical Review Letters **94**(23) (June 2005).

[134] C.-H. F. Fung and H.-K. Lo, *Security proof of a three-state quantum-key-distribution protocol without rotational symmetry*, Physical Review A **74**(4) (Oct. 2006).

[135] D. Rusca, A. Boaron, F. Grünenfelder, A. Martin, and H. Zbinden, *Finite-key analysis for the 1-decoy state QKD protocol*, Applied Physics Letters **112**(17) (Apr. 2018).

[136] X. Ma, B. Qi, Y. Zhao, and H.-K. Lo, *Practical decoy state for quantum key distribution*, Physical Review A **72**(1) (July 2005).

[137] D. Rusca, A. Boaron, M. Curty, A. Martin, and H. Zbinden, *Security proof for a simplified Bennett-Brassard 1984 quantum-key-distribution protocol*, Phys. Rev. A **98**, 052336 (Nov 2018).

[138] A. Boaron, B. Korzh, R. Houlmann, G. Boso, D. Rusca, S. Gray, M.-J. Li, D. Nolan, A. Martin, and H. Zbinden, *Simple 2.5GHz time-bin quantum key distribution*, Applied Physics Letters **112**(17), 171108 (04 2018).

[139] T. Kobayashi, A. Tomita, and A. Okamoto, *Evaluation of the phase randomness of a light source in quantum-key-distribution systems with an attenuated laser*, Phys. Rev. A **90**, 032320 (Sep 2014).

[140] J. Martinez-Mateo, C. Pacher, M. Peev, A. Ciurana, and V. Martin, *Demystifying the information reconciliation protocol Cascade*, arXiv [quant-ph] (July 2014).

[141] E. Hecht, *Optics*, Pearson, Upper Saddle River, NJ, 5 edition, Dec. 2015.

[142] S. A. Yadgeer, S. Ray, and P. Joglekar, Study of chromatic dispersion in single-mode optical fiber, in *Proceedings of the International Conference on Optical Communication Systems*, volume 2426, Feb. 2023.

[143] G. Agrawal, *Nonlinear fiber optics*, Elsevier Academic Press, 6 edition, 2019.

[144] R. Ramaswami, K. N. Sivarajan, and G. H. Sasaki, chapter 5 - Transmission System Engineering, in *Optical Networks (Third Edition)*, edited by R. Ramaswami, K. N. Sivarajan, and G. H. Sasaki, pages 289–365, Morgan Kaufmann, Boston, third edition edition, 2010.

[145] E. Lomonte, M. Stappers, L. Krämer, W. H. Pernice, and F. Lenzini, *Scalable and efficient grating couplers on low-index photonic platforms enabled by cryogenic deep silicon etching*, Scientific Reports **14** (12 2024).

[146] R. Marchetti, C. Lacava, L. Carroll, K. Gradkowski, and P. Minzioni, *Coupling strategies for silicon photonics integrated chips [Invited]*, Photonics Research **7**, 201 (2 2019).

[147] H. Yang, P. Zheng, P. Liu, G. Hu, B. Yun, and Y. Cui, *Design of polarization-insensitive 2×2 multimode interference coupler based on double strip silicon nitride waveguides*, Optics Communications **410**, 559–564 (2018).

[148] F. Morichetti, C. Ferrari, A. Canciamilla, and A. Melloni, *The first decade of coupled resonator optical waveguides: bringing slow light to applications*, Laser Photon. Rev. **6**(1), 74–96 (Jan. 2012).

[149] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, *Silicon microring resonators*, Laser Photon. Rev. **6**(1), 47–73 (Jan. 2012).

[150] S. Liu, J. Feng, Y. Tian, H. Zhao, L. Jin, B. Ouyang, J. Zhu, and J. Guo, Thermo-optic phase shifters based on silicon-on-insulator platform: state-of-the-art and a review, 12 2022.

[151] B. J. Frey, D. B. Leviton, and T. J. Madison, Temperature-dependent refractive index of silicon and germanium.

[152] D. E. Hagan, M. Nedeljkovic, W. Cao, D. J. Thomson, G. Z. Mashanovich, and A. P. Knights, *Experimental quantification of the free-carrier effect in silicon waveguides at extended wavelengths*, Opt. Express **27**(1), 166–174 (Jan 2019).

[153] J. V. Campenhout, W. M. J. Green, S. Assefa, and Y. A. Vlasov, *Integrated NiSi waveguide heaters for CMOS-compatible silicon thermo-optic devices*, Opt. Lett. **35**(7), 1013–1015 (Apr 2010).

[154] Y. Sobu, T. Simoyama, S. Tanaka, Y. Tanaka, and K. Morito, 70 Gbaud Operation of All-Silicon Mach-Zehnder Modulator based on Forward-Biased PIN Diodes and Passive Equalizer.

[155] A. Rahim, A. Hermans, B. Wohlfeil, D. Petousi, B. Kuyken, D. V. Thourhout, and R. Baets, Taking silicon photonics modulators to a higher performance level: State-of-the-art and a review of new technologies, 3 2021.

[156] G. Corrielli, A. Crespi, and R. Osellame, *Femtosecond laser micromachining for integrated quantum photonics*, Nanophotonics **10**(15), 3789–3812 (2021).

[157] G. Corrielli, A. Crespi, and R. Osellame, *Femtosecond laser micromachining for integrated quantum photonics*, Nanophotonics **10**(15), 3789–3812 (Nov. 2021).

[158] G. Corrielli, S. Atzeni, S. Piacentini, I. Pitsios, A. Crespi, and R. Osellame, *Symmetric polarization-insensitive directional couplers fabricated by femtosecond laser writing*, Opt. Express **26**(12), 15101–15109 (June 2018).

[159] L. A. Fernandes, J. R. Grenier, P. R. Herman, J. S. Aitchison, and P. V. S. Marques, *Stress induced birefringence tuning in femtosecond laser fabricated waveguides in fused silica*, Opt. Express **20**(22), 24103–24114 (Oct. 2012).

[160] C. P. Ho, Z. Zhao, Q. Li, S. Takagi, and M. Takenaka, *Tunable grating coupler by thermal actuation and Thermo-optic effect*, IEEE Photonics Technol. Lett. **30**(17), 1503–1506 (Sept. 2018).

[161] B. Korzh and H. Zbinden, Low temperature performance of free-running InGaAs/InP single-photon negative feedback avalanche diodes, in *Advanced Photon Counting Techniques VIII*, edited by M. A. Itzler and J. C. Campbell, volume 9114, SPIE, June 2014.

[162] H. T. Yen, S. D. Lin, and C. M. Tsai, *A simple method to characterize the afterpulsing effect in single photon avalanche photodiode*, J. Appl. Phys. **104**(5), 054504 (Sept. 2008).

[163] M. A. Itzler, X. Jiang, M. Entwistle, M. Owens, K. Slomkowski, A. Ferri, J. Hughes, Y. H. Lo, and J. C. Campbell, *Advances in InGaAsP-based avalanche diode single-photon detectors*, Journal of Modern Optics **58**(3-4), 174–200 (2011).

[164] Z. Liu, Q. Liu, X. Wang, and J. Zhang, *Temperature dependence of afterpulsing in InGaAs/InP single-photon avalanche diodes*, Applied Optics **52**(20), 4901–4906 (2013).

[165] J.-Y. Wu, C.-H. Lin, and S.-H. Wang, *An Analysis of Temperature-Dependent Timing Jitter Factors in the Structural Design of CMOS SPADs*, Sensors **25**(4), 912 (2025).

[166] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, *Secure quantum key distribution with realistic devices*, Rev. Mod. Phys. **92**(2) (May 2020).

[167] J. A. Dolphin, T. K. Paraïso, H. Du, R. I. Woodward, D. G. Marangon, and A. J. Shields, *A hybrid integrated quantum key distribution transceiver chip*, Npj Quantum Inf. **9**(1), 1–8 (Sept. 2023).

[168] P. Sibson, C. Erven, M. Godfrey, S. Miki, T. Yamashita, M. Fujiwara, M. Sasaki, H. Terai, M. G. Tanner, C. M. Natarajan, R. H. Hadfield, J. L. O'Brien, and M. G. Thompson, *Chip-based quantum key distribution*, Nat. Commun. **8**(1), 13984 (Feb. 2017).

[169] S. Beppu, D. J. Elson, S. Murai, A. Murakami, H. Yamamuro, Y. Wakayama, N. Yoshikane, and T. Tsuritani, Coexistence Transmission of 33.4-Tb/s O-band Coherent Classical Channels and a C-band QKD Channel over 80 km, in *2025 Optical Fiber Communications Conference and Exhibition (OFC)*, pages 1–3, IEEE, Mar. 2025.

[170] P. Zhang, J. Sagar, E. Hastings, M. Stefko, S. Joshi, and J. Rarity, *End-to-end demonstration for CubeSatellite quantum key distribution*, IET Quantum Communication **5**(3), 291–302 (Sept. 2024).

[171] T. Hiemstra, D. Hasler, D. Paone, F. Reichert, F. Heine, and J. Struck, *The European satellite-based QKD system EAGLE-1*, arXiv [quant-ph] (May 2025).

[172] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, 2010.

[173] M. Herrero-Collantes and J. C. Garcia-Escartin, *Quantum random number generators*, Rev. Mod. Phys. **89**(1), 015004 (Feb. 2017).

[174] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden, *Optical quantum random number generator*, J. Mod. Opt. **47**(4), 595–598 (Mar. 2000).

[175] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger, *A fast and compact quantum random number generator*, Rev. Sci. Instrum. **71**(4), 1675–1680 (Apr. 2000).

[176] T. Gehring, C. Lupo, A. Kordts, D. Solar Nikolic, N. Jain, T. Rydberg, T. B. Pedersen, S. Pirandola, and U. L. Andersen, *Homodyne-based quantum random number generator at 2.9 Gbps secure against quantum side-information*, Nat. Commun. **12**(1), 605 (Jan. 2021).

[177] J. Li, Z. Huang, C. Yu, J. Wu, T. Zhao, X. Zhu, and S. Sun, *Quantum random number generation based on phase reconstruction*, Opt. Express **32**(4), 5056–5071 (Feb. 2024).

[178] Z. Zheng, Y. Zhang, W. Huang, S. Yu, and H. Guo, *6 Gbps real-time optical quantum random number generator based on vacuum fluctuation*, Rev. Sci. Instrum. **90**(4), 043105 (Apr. 2019).

[179] F. Cavaliere, E. Prati, L. Poti, I. Muhammad, and T. Catuogno, *Secure quantum communication technologies and systems: From labs to markets*, Quantum Rep. **2**(1), 80–106 (Jan. 2020).

[180] X. Wang, Y. Shi, N. Wang, J. Yun, J. Li, Y. Jia, S. Liu, Z. Lu, J. Zou, and Y. Li, *Highly integrated broadband entropy source for quantum random number generators based on vacuum fluctuations*, arXiv [quant-ph] (Apr. 2025).

[181] B. Bai, J. Huang, G.-R. Qiao, Y.-Q. Nie, W. Tang, T. Chu, J. Zhang, and J.-W. Pan, *18.8 Gbps real-time quantum random number generator with a photonic integrated chip*, Appl. Phys. Lett. **118**(26), 264001 (June 2021).

[182] C. Bruynsteen, T. Gehring, C. Lupo, J. Bauwelinck, and X. Yin, *100-gbit/s integrated quantum random number generator based on vacuum fluctuations*, PRX quantum **4**(1), 010330 (Mar. 2023).

[183] X.-W. Fei, Z.-Q. Yin, W. Huang, B.-J. Xu, S. Wang, W. Chen, Y.-G. Han, G.-C. Guo, and Z.-F. Han, *Tighter bound of quantum randomness certification for independent-devices scenario*, Sci. Rep. **7**(1), 14666 (Nov. 2017).

[184] Y. Zhang, E. Knill, and P. Bierhorst, *Certifying quantum randomness by probability estimation*, Phys. Rev. A (Coll. Park.) **A98**(4) (2018).

[185] Y. Liu, Q. Zhao, M.-H. Li, J.-Y. Guan, Y. Zhang, B. Bai, W. Zhang, W.-Z. Liu, C. Wu, X. Yuan, H. Li, W. J. Munro, Z. Wang, L. You, J. Zhang, X. Ma, J. Fan, Q. Zhang, and J.-W. Pan, *Device-independent quantum random-number generation*, Nature **562**(7728), 548–551 (Oct. 2018).

[186] L. K. Shalm, Y. Zhang, J. C. Bienfang, C. Schlager, M. J. Stevens, M. D. Mazurek, C. Abellán, W. Amaya, M. W. Mitchell, M. A. Alhejji, H. Fu, J. Ornstein, R. P. Mirin, S. W. Nam, and E. Knill, *Device-independent randomness expansion with entangled photons*, Nat. Phys. **17**(4), 452–456 (Apr. 2021).

[187] P. Bierhorst, E. Knill, S. Glancy, Y. Zhang, A. Mink, S. Jordan, A. Rommal, Y.-K. Liu, B. Christensen, S. W. Nam, M. J. Stevens, and L. K. Shalm, *Experimentally generated randomness certified by the impossibility of superluminal signals*, Nature **556**(7700), 223–226 (Apr. 2018).

[188] G. Vallone, D. G. Marangon, M. Tomasin, and P. Villoresi, *Quantum randomness certified by the uncertainty principle*, Physical Review A - Atomic, Molecular, and Optical Physics **90** (11 2014).

[189] M. Avesani, H. Tebyanian, P. Villoresi, and G. Vallone, *Semi-Device-Independent Heterodyne-based Quantum Random Number Generator*, (4 2020).

[190] D. G. Marangon, G. Vallone, and P. Villoresi, *Source-Device-Independent Ultrafast Quantum Random Number Generation*, Physical Review Letters **118** (2 2017).

[191] Y.-Q. Nie, J.-Y. Guan, H. Zhou, Q. Zhang, X. Ma, J. Zhang, and J.-W. Pan, *Experimental measurement-device-independent quantum random number generation*, (12 2016).

[192] J. Bowles, M. T. Quintino, and N. Brunner, *Certifying the dimension of classical and quantum systems in a prepare-and-measure scenario with independent devices*, (11 2013).

[193] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. A. Methot, and V. Scarani, *Testing the Hilbert space dimension*, (2 2008).

[194] T. van Himbeeck, E. Woodhead, N. J. Cerf, R. García-Patrón, and S. Pironio, *Semi-device-independent framework based on natural physical assumptions*, Quantum **1** (11 2017).

[195] T. V. Himbeeck and S. Pironio, *Correlations and randomness generation based on energy constraints*, (5 2019).

[196] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, Cambridge University Press, Cambridge, UK, 1995.

[197] D.-G. Welsch, W. Vogel, and T. Opatrný, II Homodyne Detection and Quantum-State Reconstruction, volume 39 of *Progress in Optics*, pages 63–211, Elsevier, 1999.

[198] P. Marecki, *Balanced homodyne detectors in quantum field theory*, Phys. Rev. A **77**, 012101 (Jan 2008).

[199] P. Fritschel, M. Evans, and V. Frolov, *Balanced homodyne readout for quantum limited gravitational wave detectors*, Opt. Express **22**(4), 4224–4234 (Feb. 2014).

[200] J. Heinze, K. Danzmann, B. Willke, and H. Vahlbruch, *10 dB quantum-enhanced Michelson interferometer with balanced homodyne detection*, Phys. Rev. Lett. **129**(3), 031101 (July 2022).

[201] Y.-M. Chi, B. Qi, W. Zhu, L. Qian, H.-K. Lo, S.-H. Youn, A. I. Lvovsky, and L. Tian, *A balanced homodyne detector for high-rate Gaussian-modulated coherent-state quantum key distribution*, New J. Phys. **13**(1), 013003 (Jan. 2011).

[202] T. Gabbrielli, F. Cappelli, N. Bruno, N. Corrias, S. Borri, P. De Natale, and A. Zavatta, *Mid-infrared homodyne balanced detector for quantum light characterization*, Opt. Express **29**(10), 14536–14547 (May 2021).

[203] Y. Jia, X. Wang, X. Hu, X. Hua, Y. Zhang, X. Guo, S. Zhang, X. Xiao, S. Yu, J. Zou, and Y. Li, *Silicon photonics-integrated time-domain balanced homodyne detector in continuous-variable quantum key distribution*, arXiv [quant-ph] (May 2023).

[204] D. Xie, C. Xu, X. Yao, and A. M. Wang, *Quantum metrology with quantum Wheatstone bridge composed of Bose systems*, arXiv [quant-ph] (Aug. 2022).

[205] Y. Miao, F. Xie, W. Feng, Y. Zhu, X. Zhang, and F. Liu, *High-Precision Interferometric Measurements of Gas Refractive Index Using Homodyne Detection*, Sensors **25**(11) (2025).

[206] J. Lu, Y. Gao, Z. Ma, H. Zhou, R. K. Wang, and Y. Wang, *In vivo photoacoustic imaging of blood vessels using a homodyne interferometer with zero-crossing triggering*, J. Biomed. Opt. **22**(3), 36002 (Mar. 2017).

[207] J. Lu, Y. Gao, Z. Ma, H. Zhou, R. K. Wang, and Y. Wang, *In vivo photoacoustic imaging of blood vessels using a homodyne interferometer with zero-crossing triggering*, J. Biomed. Opt. **22**(3), 36002 (Mar. 2017).

[208] S. Grandi, A. Zavatta, M. Bellini, and M. G. A. Paris, *Experimental quantum tomography of a homodyne detector*, New J. Phys. **19**(5), 053015 (May 2017).

[209] M. Kalash and M. V. Chekhova, *Wigner function tomography via optical parametric amplification*, arXiv [quant-ph] (July 2022).

[210] D. Rusca, Security of quantum cryptography: from quantum random key generation to quantum key distribution, 2021.

[211] J. B. Brask, A. Martin, W. Esposito, R. Houlmann, J. Bowles, H. Zbinden, and N. Brunner, *Megahertz-Rate Semi-Device-Independent Quantum Random Number Generators Based on Unambiguous State Discrimination*, Physical Review Applied **7** (5 2017).

[212] Wideband, Ultra-Low Noise, Voltage-Feedback OPERATIONAL AMPLIFIER with Shutdown, 2002, www.ti.com.

[213] R. Wang, Y. Tian, Q. Li, and Y. Zhao, *High gain and low excess noise InGaAs/InP avalanche photodiode with lateral impact ionization*, Appl. Opt. **59**(7), 1980–1984 (Mar. 2020).

[214] A. A. Dadey, J. A. McArthur, A. Kamboj, S. R. Bank, D. Wasserman, and J. C. Campbell, *High-gain low-excess-noise MWIR detection with a 3.5-μm cutoff AlInAsSb-based separate absorption, charge, and multiplication avalanche photodiode*, APL Photonics **8**(3) (Mar. 2023).

[215] J. Zhang, H. Lin, M. Liu, and Y. Yang, *Research on the leakage current at sidewall of mesa Ge/Si avalanche photodiode*, AIP Adv. **11**(7), 075320 (July 2021).

[216] P. Cao, H. Peng, T. Wang, V. Srivastava, M. Kesaria, M. You, Q. Zhuang, and W. Zheng, *Surface passivation of random alloy AlGaAsSb avalanche photodiode*, Electron. Lett. **59**(18) (Sept. 2023).

[217] Si Avalanche Photodiodes Technical Information, Technical Report KAPD9007E, Hamamatsu Photonics K.K., 2017, Accessed: 2025-09-26.

[218] G. Harman, *Wire Bonding in Microelectronics*, McGraw-Hill Professional, New York, 3rd edition, 2010.

[219] K. Technologies, Spectrum Analysis Basics (AN150), www.keysight.com.

[220] X. Zhang, Y.-C. Zhang, Z. Li, S. Yu, and H. Guo, *1.2 GHz Balanced Homodyne Detector for Continuous-Variable Quantum Information Technology*, (June 2018).

[221] M. Smit, K. Williams, and J. van der Tol, *Past, present, and future of InP-based photonic integration*, APL Photonics **4**(5), 050901 (May 2019).

[222] L.-Q. Yue, Y.-L. Shi, N.-F. Sun, S. Qiang, J.-K. Qin, L. Zhen, and C.-Y. Xu, *InP low-dimensional nanomaterials for electronic and optoelectronic device applications: A Review*, Adv. Sens. Res. **2**(10), 2200101 (Oct. 2023).

[223] X. Yang, A. K. Mishra, D. Lenstra, F. M. Huijskens, H. D. Waardt, G. D. Khoe, and H. J. Dorren, *Sub-picosecond all-optical switch using a multi-quantum-well semiconductor optical amplifier*, Optics Communications **236**, 329–334 (6 2004).

[224] C. St-Arnault, S. Bernal, R. Gutierrez-Castrejn, E. Berikaa, W. Li, Z. Wei, Y. Hu, M. S. Alam, J. Rautert, S. V. Poltavtsev, A. E. Gubenko, V. V. Belykh, V. S. Mikhrin, A. R. Kovsh, and D. V. Plant, *Performance and Characterization Comparison of QD SOA, QW SOA, Bulk SOA and PDFA for Multi-Tbps O-Band WDM Links*, Journal of Lightwave Technology (2024).

[225] M. V. Lysevych, Design, Growth, Fabrication and Characterisation of High Power Single Mode InGaAsP/InP Lasers, 2013.

[226] F. G. D. Corte, G. Cocorullo, M. Iodice, and I. Rendina, *Temperature dependence of the thermo-optic coefficient of InP, GaAs, and SiC from room temperature to 600 K at the wavelength of 1.5 um*, Applied Physics Letters **77**, 1614–1616 (9 2000).

[227] S.-K. Yun and H.-Y. Lee, *Parasitic impedance analysis of double bonding wires for high-frequency integrated circuit packaging*, IEEE Microw. Guid. Wave Lett. **5**(9), 296–298 (1995).

[228] M.-D. Ker, H.-C. Jiang, and C.-Y. Chang, Design of low-capacitance bond pad for high-frequency I/O applications in CMOS integrated circuits, in *Proceedings of 13th Annual IEEE International ASIC/SOC Conference (Cat. No.00TH8541)*, pages 293–296, IEEE, 2002.

[229] T. Blalack, Y. Leclercq, and C. P. Yue, On-Chip RF-Isolation Techniques, 2003, Accessed: September 26, 2025, https://www.eetimes.com/on-chip-rf-isolation-techniques/.

[230] S. H. Voldman, C. Nicholas Perez, and A. Watson, *Guard rings: Structures, design methodology, integration, experimental results, and analysis for RF CMOS and RF mixed signal BiCMOS silicon germanium technology*, J. Electrostat. **64**(11), 730–743 (Oct. 2006).

[231] H. Mekawey, M. Elsayed, Y. Ismail, and M. A. Swillam, *Optical interconnects finally seeing the light in silicon photonics: Past the hype*, Nanomaterials (Basel) **12**(3), 485 (Jan. 2022).

[232] P. Kaur, A. Boes, G. Ren, T. G. Nguyen, G. Roelkens, and A. Mitchell, *Hybrid and heterogeneous photonic integration*, APL Photonics **6**(6), 061102 (June 2021).

Appendix A

QRNG Photodiode Characterization

As provided by the manufacturer, the PDs used in this experiment have a responsivity of approximately 0.8A/W at a wavelength of 1550nm. This means that for every watt of optical power incident on the photodiode, it generates a photocurrent of 0.8 amperes. Given the optical in chip was expected to be in the range of µW to tens of mW, the expected photocurrent generated would be in the range of nA to mA rendering it too small to be effectively measured with a ADC. For this reason, each photodiode is connected to a tunable amplification stage, which is implemented using an operational amplifier.

In order to accurately set the gain and characterize the amplification stage for each PD, we removed the optical chip from the setup and connected a tunable current source to the input of each amplification stage. Then, while performing a sweep of the input current, we recorded the absolute units measured by the ADC. This allowed us to determine the relationship between the input current and the output voltage for each amplification stage. For each PD we fit the data to a linear, quadratic, cubic and power law function to determine the best fit. The results are presented in Figures A.1, A.2, A.3, A.4, A.5, A.6.

The best fit for each PD was determined by comparing the corresponding R^2 values. For all PDs, the cubic model provided the best fit. Finally, from all the current equations and power equations were derived. These results are the ones used in the main text to convert the absolute units measured by the ADC to photocurrent and optical power.

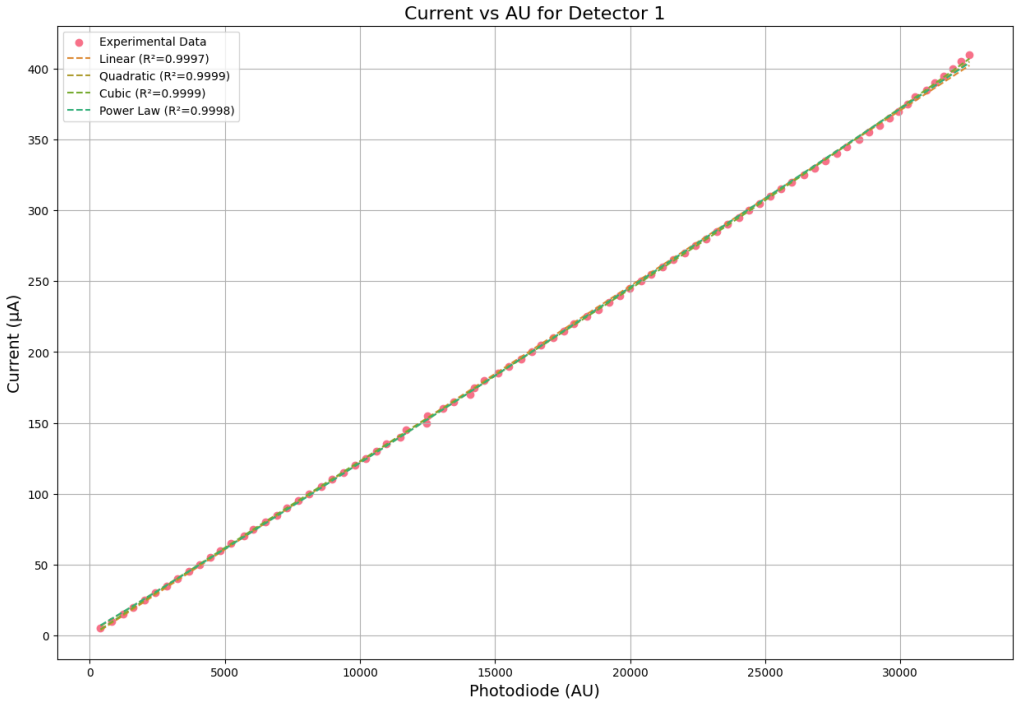


Figure A.1: PD1 amplification circuit and ADC characterization

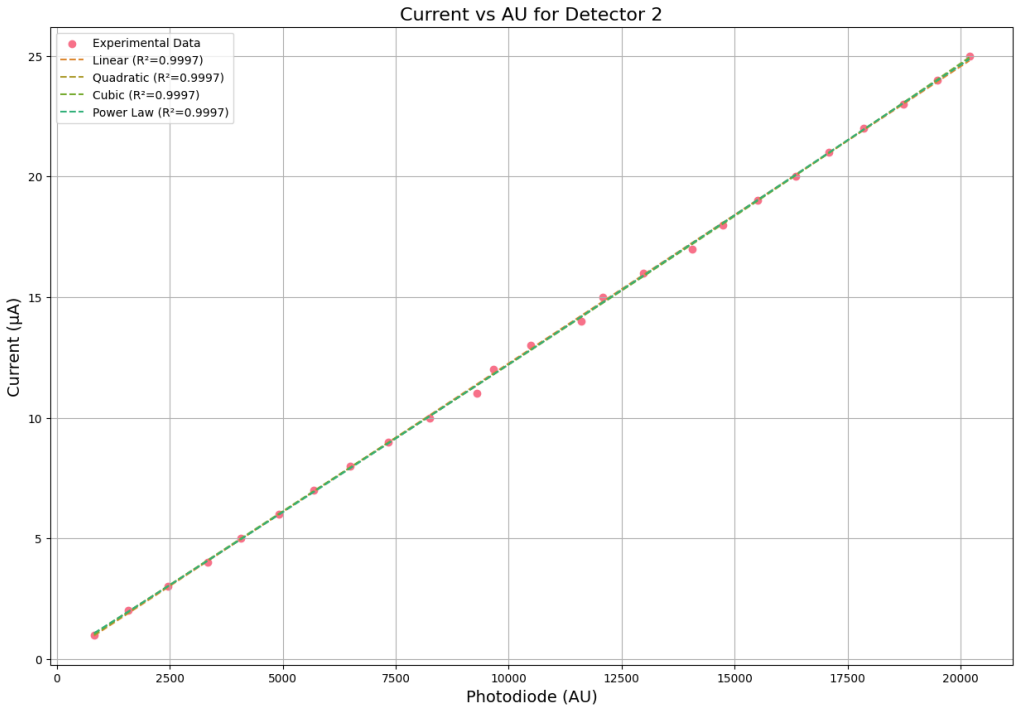


Figure A.2: PD2 amplification circuit and ADC characterization

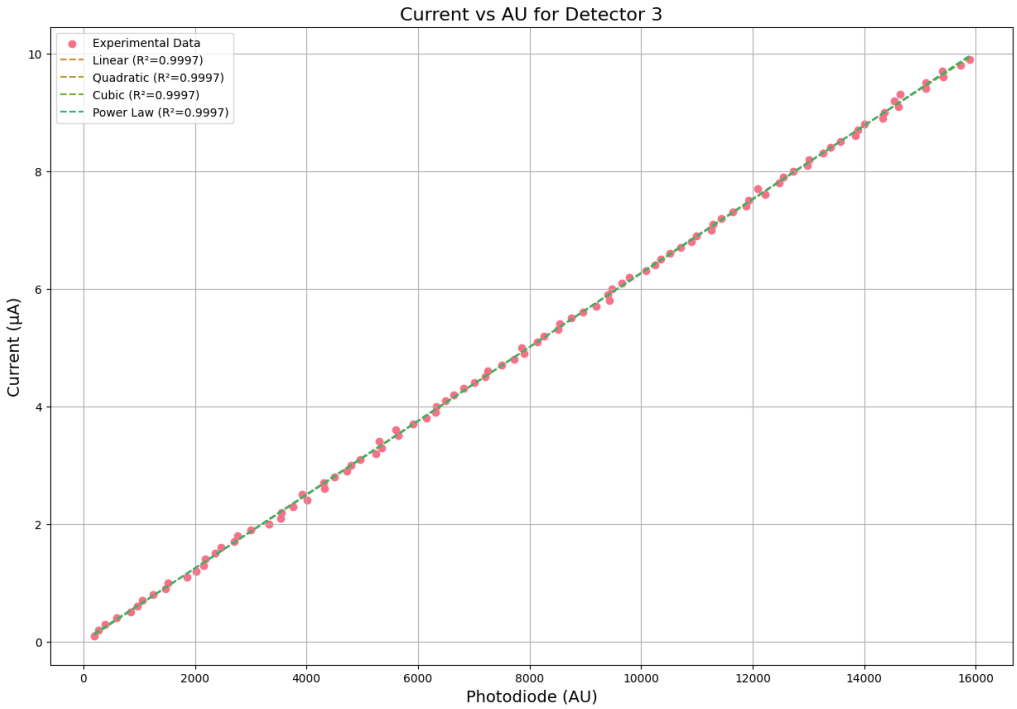


Figure A.3: PD3 amplification circuit and ADC characterization

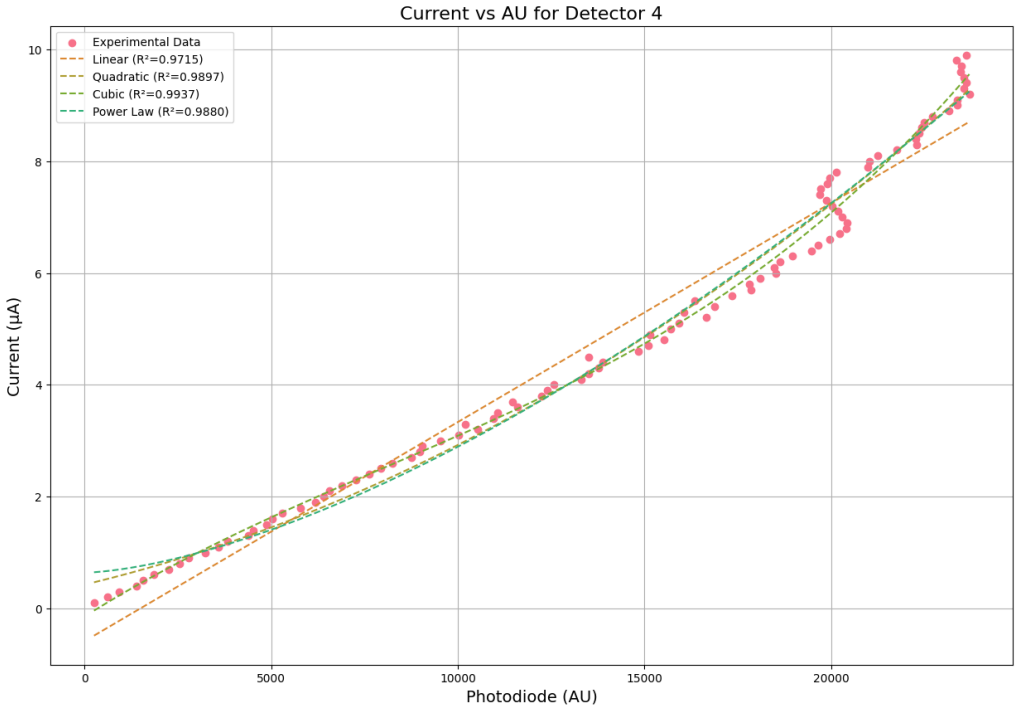


Figure A.4: PD4 amplification circuit and ADC characterization

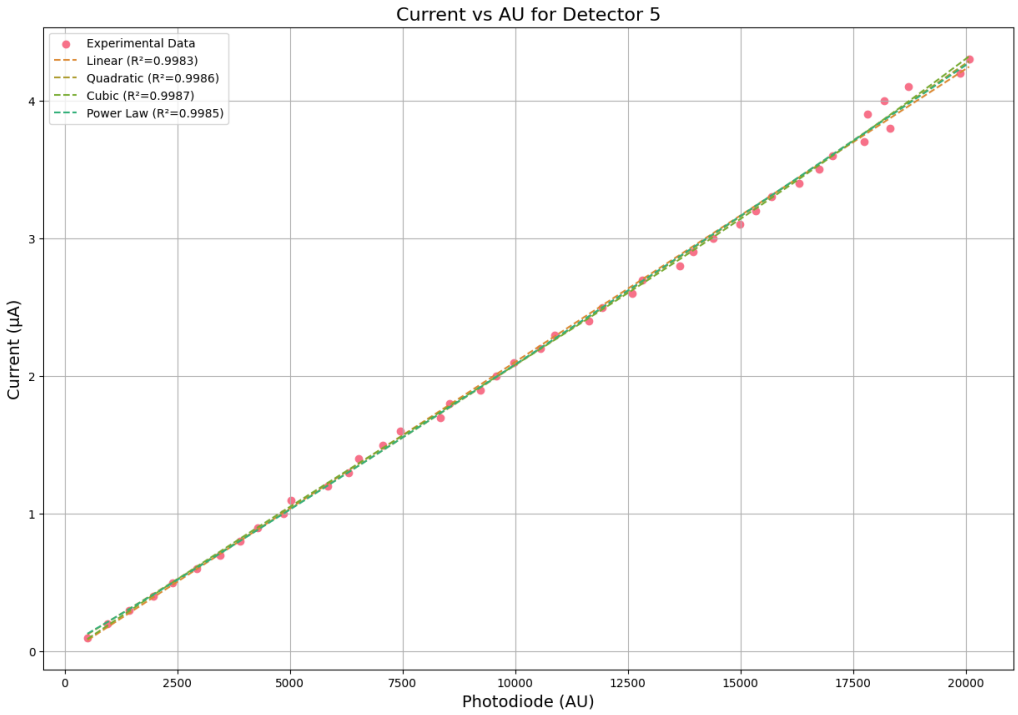


Figure A.5: PD5 amplification circuit and ADC characterization

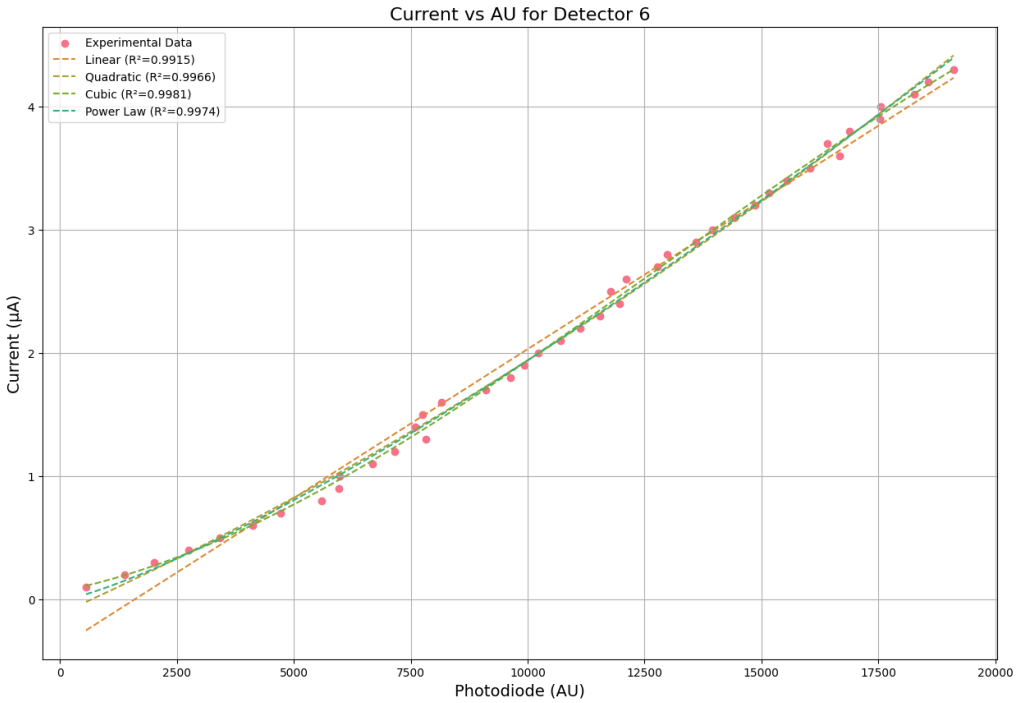


Figure A.6: PD6 amplification circuit and ADC characterization

Appendix B

QRNG Variable Optical Attenuator Characterization

The VOA used in the signal arm of the QRNG chip consisted of four stages of MZIs, each equipped with two thermo-optic heaters to control the phase shift and, consequently, the attenuation level. The characterization of each stage's heater was performed to understand the relationship between the applied voltage and the resulting attenuation. The characterization involved measuring the output power at each stage of the VOA, with the PDs placed in the output arm that is not connected to the following MZI stage, while varying the voltage applied to each heater individually, keeping the other heaters at a constant voltage. The input power was monitored to ensure consistency during the measurements. The results of the characterization are presented in Figures B.1,B.2,B.3 and B.4, showing the attenuation provided by each stage of the VOA as a function of the applied voltage to the respective heater. For all stages except the first one, the attenuation in the previous stages was minimized. Since this characterization was performed with a laser source external to the chip, in order to not be influenced by the fluctuations in the coupling efficiency, the output power of each stage was normalized to the input power measured at the input of the VOA.

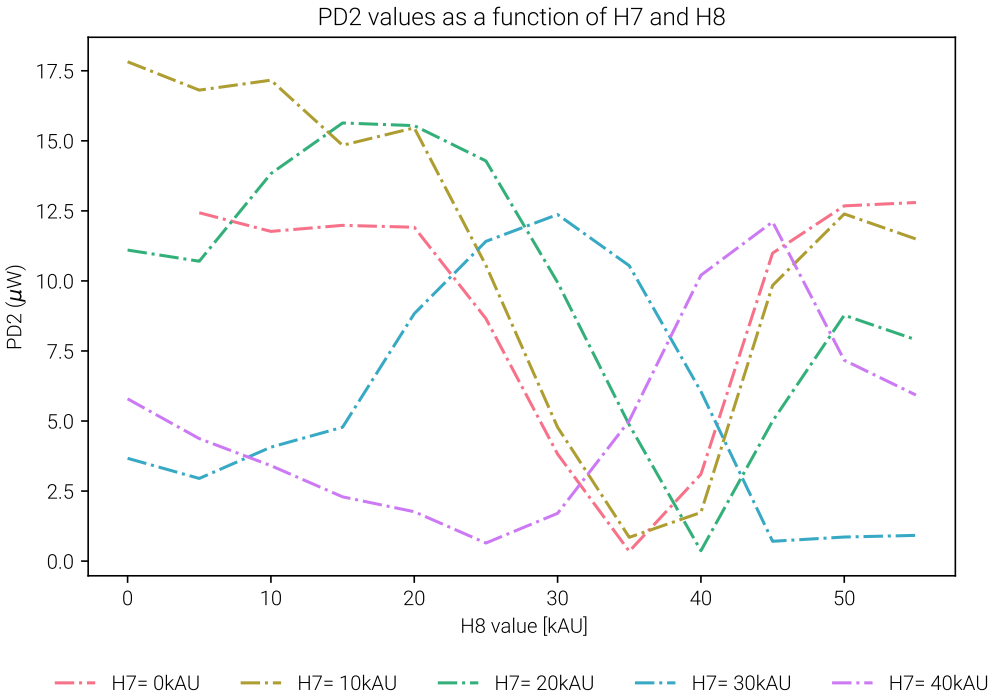


Figure B.1: First Stage Heater Characterization with the power measured at the output of the VOA normalized to the input power.

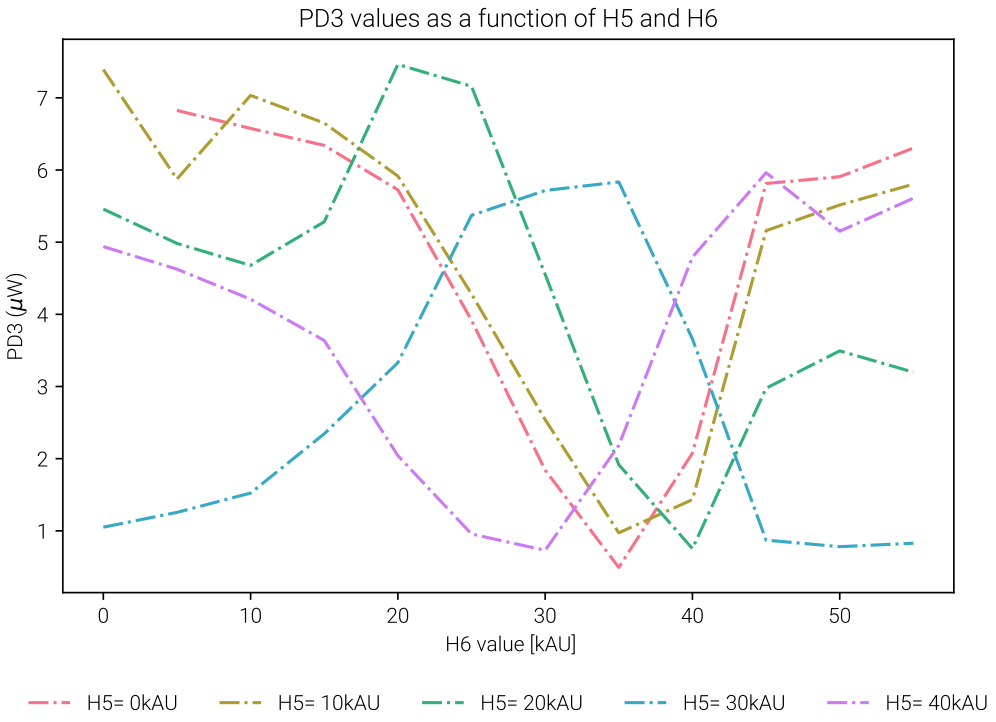


Figure B.2: Second Stage Heater Characterization with the power measured at the output of the VOA normalized to the input power.

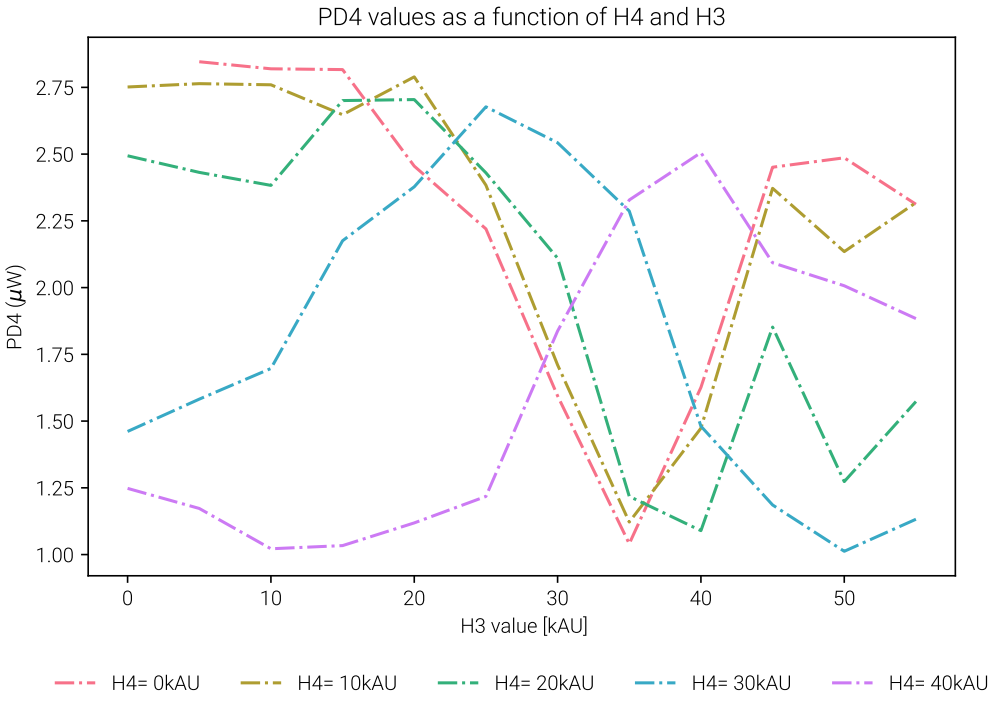


Figure B.3: Third Stage Heater Characterization with the power measured at the output of the VOA normalized to the input power.

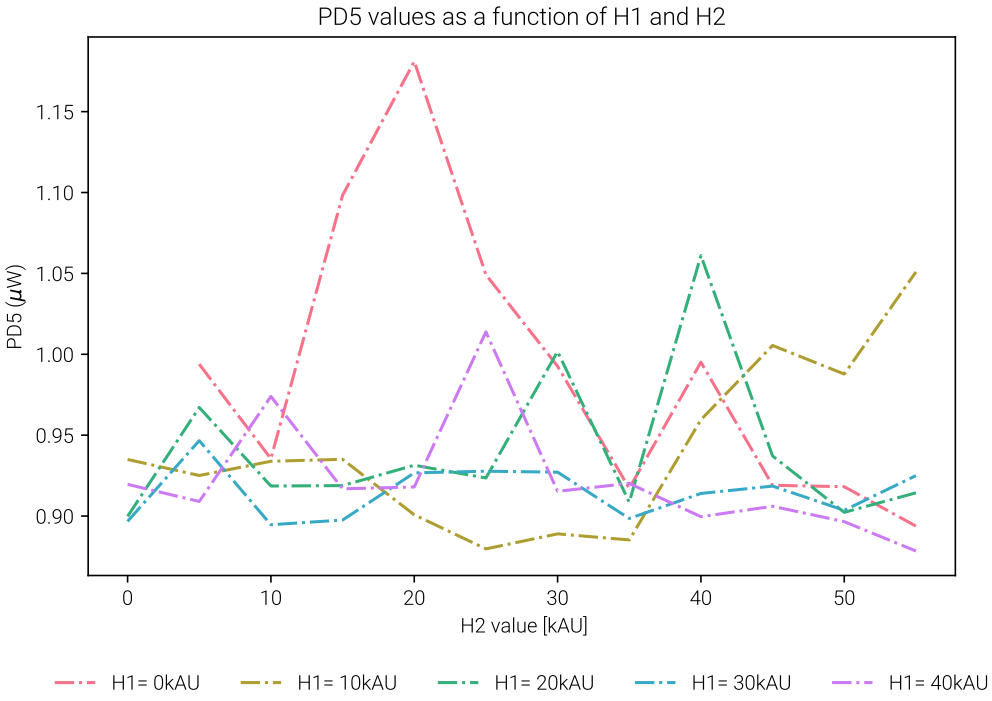


Figure B.4: Fourth Stage Heater Characterization with the power measured at the output of the VOA normalized to the input power.

Appendix C

Transmitter's Variable Optical Attenuator Characterization

The integrated VOA in the transmitter was designed to be the last component in the optical path before the light exits the chip. It provided the attenuation of the encoded states to the desired mean photon number per pulse. The VOA consisted of a balanced MZI with TOPSs in each arm to control the interference at the output coupler, as shown in Figure C.1. By controlling this interference, the output power can be attenuated up to 31dB.

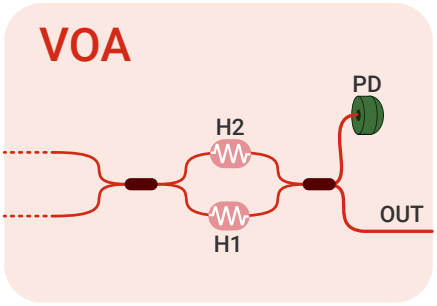


Figure C.1: Schematics of the integrated VOA in the transmitter.

To characterize the integrated VOA of the transmitter, we measured the output optical power while varying the current applied to each of the heaters. Since they

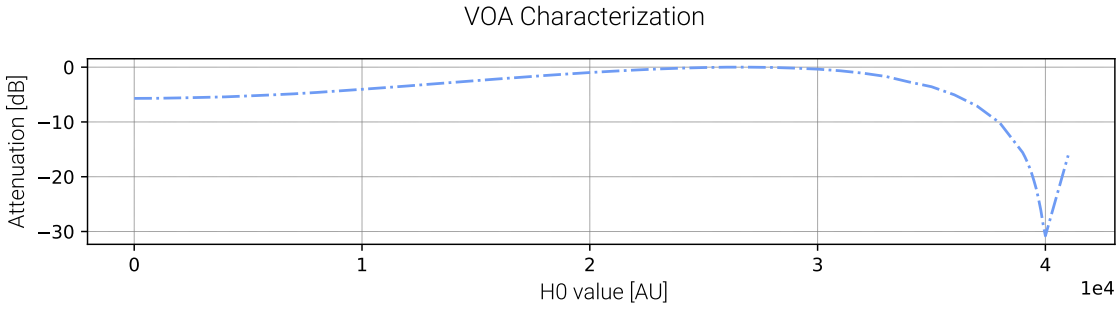


Figure C.2: Plot of the attenuation in output optical power [dB] versus the current applied to one of the heaters of the integrated VOA in the transmitter (in kAU). The other heater was kept at 0kAU.

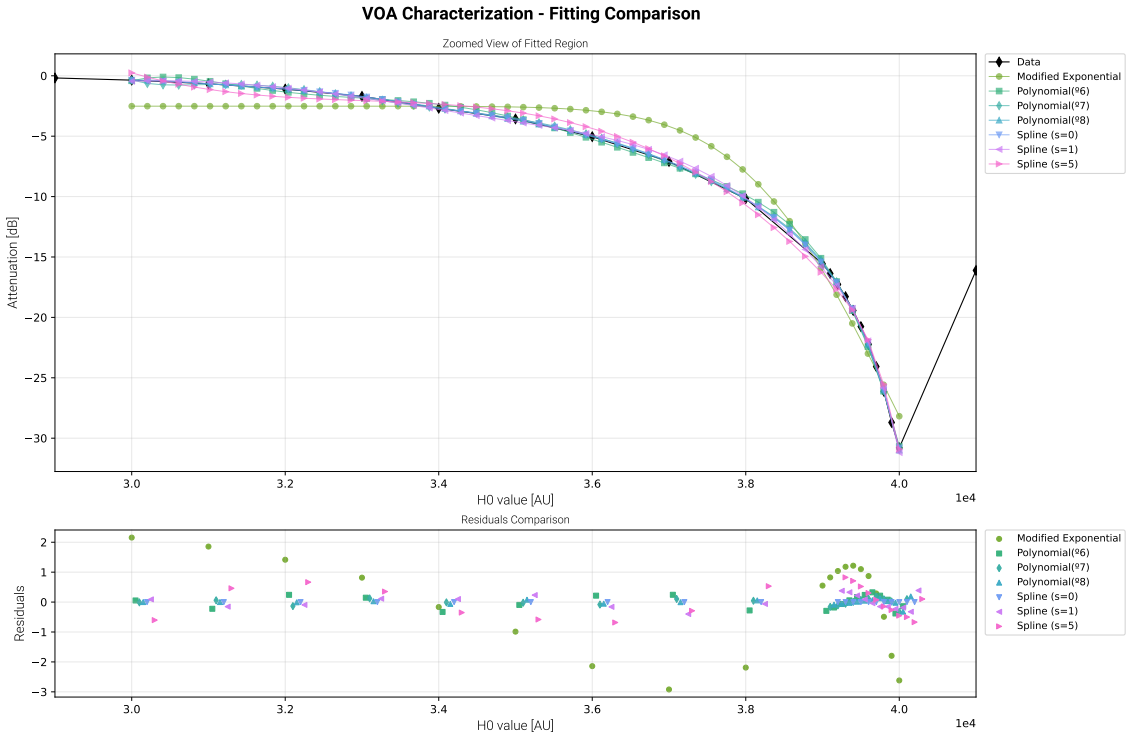


Figure C.3: Comprehensive comparison of different fit types for the VOA characterization data.

worked symmetrically, we only show the characterization of the one heater we ended up using in the final QKD experiments. The measurement results are shown in Figure C.2. Due to the unconventional shape of the data, we decided to crop the data to focus on what we consider the region of interest, where the attenuation ranges from zero (20kAU) to the maximum attenuation (40kAU). Upon a first look, it seemed that the data followed an exponential trend, however, after trying to fit the data with different types of exponential decay functions with different parameters, we found that none of provided an accurate enough fit. We then decided to try different types of fits, including polynomials of different degrees and splines with different smoothing factors. The results of the different fits are shown in Figure C.3. Due to the importance of having a high degree of confidence in the mean photon number per pulse, we ended up choosing the polynomial fit of degree 8, which provided a R^2 value < 0.99 and a low maximum residual. This fit was then used to calculate the heater value needed to achieve the desired mean photon number per pulse during the QKD experiments.