POLITECNICO DI MILANO

SCUOLA DI INGEGNERIA DEI SISTEMI

CORSO DI LAUREA IN

INGEGNERIA FISICA

FREE RUNNING SINGLE PHOTON DETECTION BASED ON A NEGATIVE FEEDBACK INGAAS APD

Tesi di Laurea di : Stefano Decadri 736409

Relatore: Prof. Gianluca Valentini

Correlatore: Prof. Hugo Zbinden

Anno Accademico: 2010/2011

Acknowledgments.

Innanzitutto voglio ringraziare la mia famiglia, senza la quale tutto questo lavoro e tutto ciò che lo precede non sarebbe mai stato possibile : grazie papà, mamma, sorellina, nonni, nonne, zie… Poi voglio ringraziare il Prof. Valentini e il Prof. Svelto, per avermi permesso di andare a Ginevra nell'aprile del 2010, insieme a Jack (anche se poi lui ha preferito andarsene a Barcellona…). Ringrazio il Prof. Hugo Zbinden e il Prof. Nicolas Gisin per avermi accettato al GAP per preparare questa tesi e per i preziosi consigli che mi hanno dato , e tutti quelli che a Ginevra mi hanno aiutato, soprattutto quando non sapevo far nulla : Claudio, Olivier, Raphael, Nino, Charles, Patrick, Bruno, Enrico, Natalia…e tutti gli altri membri del GAP. Un ringraziamento particolare a Tommaso, per la correzione della tesi e i suggerimenti, e a Jun, per l'enorme pazienza che ha avuto. Poi ringrazio tutti gli amici, conosciuti (e sparsi) qua e là : Alberto, Pit, Fabio, Luca, Ale, Bea, Fra, La-Sissa-e-il-Mone, Ste, Alice e tutte le Brugheresi, Sadir, Gigi, Pupilla, Gommoni, Gianni e tutti quelli del calcetto e del Poli in generale, Mariachiara, Marta, Chiara, Sandra, Francis, Anika e tutti quelli che ho conosciuto a Ginevra. E poi, last but not least, grazie a Silvia. Per tutto.

Table of contents.

Table of figures and graphs.

which lowers the voltage over the diode and finally quenches the avalanche. (b): Current (upper graph) and voltage (lower graph) evolution in the diode, which shows the retriggering process of a SPAD in a PQC; the speed of recovery depends on the RC constant of the circuit. Extracted from [25].

Figure 4.2. Equivalent circuit of a PQC in the current-output mode configuration[18]. The avalanche signal is subsequently sensed by a comparator, which generates a standard output useful for counting the avalanches triggered. On the right the behaviors of diode current I_d and diode voltage V_d is shown. The diode current rises quickly to its peak value and then it decreases exponentially towards I_f and meanwhile the catode voltage also decreases exponentially; when the avalanche reaches I_Q, it is quenched to zero and thus the voltage starts to slowly return to its initial value.

Figure 4.3 Diode avalanche current for two different R_B values[22].

Figure 4.4 Dependency of α on I_f [22].

Figure 4.5 Basic diagram of an AQC[22].

Figure 4.6. Rapid gating detector set-up using the sine-gating plus filtering technique[64]. BPF : Band pass filter, BSF : Band stop filter

Figure 5.1. The experimental set-up used for measures of dark counts, efficiency and afterpulses (with double-gated method)

Figure 5.2 Photograph of the box represented in Fig. 6.3.1, with descriptions of the external connections; the NFAD is fixed over the Peltier cell.

Figure5.3 Free-running mode. (a) dark counts and (b) effective efficiency as a function of the bias voltage applied on the NFAD.

Figure5.4 Gated mode: (a) Normalized dark counts (per second) and (b) efficiency as a function of the bias voltage applied on the NFAD.

Figure 5.5. Behavior of the noise as a function of the efficiency. For the case of the Free-running Mode the noise is the “raw” Dark Count rate, while for the Gated Mode it is the Normalized Dark Counts Rate (considering only the time during which the diode is active).

Figure 5.6. The double-gated method. The SPAD is biased under its breakdown voltage, thus being normally in a non active state; an arbitrary function generator produces square biasing gates with a full width at half maximum of 150 ns, an amplitude of 3 V and a repetition frequency of 10 kHz. When a biasing gate is applied, the diode is biased over the breakdown voltage, so an avalanche could be triggered, filling the traps and thus generating subsequent afterpulses; these afterpulses are measured during a second bias gate situated at varying delays after the first one. Coincidence windows (shorter than biasing windows) are necessary in order to discriminate, at the output of the discriminator, the pulses due to detection and the derivative pulses originated by the square biasing gates as a consequence of the ac coupling.

Figure 5.7. Afterpulse probability per ns versus delay of the second gate

Figure 5.8 Afterpulse probability per ns versus deadtime. Data from : J.Zhang, R. Thew,J.-D. Gautier, N. Gisin, and H. Zbinden, “ Comprehensive Characterization of InGaAs-InP Avalanche Photodiodes at 1550 nm With an Active Quenching ASIC”, IEEE J. of Quantum Electron., vol. 45, NO. 7, July 2009

Figure 5.9. Set-up used for the measures of afterpulses using an FPGA.

Figure 5.10 Logical behavior of the FPGA. During this experiment the dead time was not already implemented, thus every time there was a detection the FPGA went directly into the histogram mode.

Figure 5.11. Laser Detection Efficiency versus the number of waiting cycles. The measures were registered at -50 C, with a bias voltage corresponding to ~14% of efficiency at 500 W.C.

Figure 5.12 Probability of afterpulses per ns versus time at 223 K, $f_r \approx 97$ kHz, $\eta \approx 10\%$ and 0.1 photons per Pulse. The results of the fit are: $y_0 = 10^{-6}$ (this parameter was fixed before the fit, and was previously obtained with a linear fit of the curve starting from 100 μs after the detection of the laser pulse), $A_1 = 3*10^{-4}$, $\tau_1 = 512.9$ (± 9.8), $A_2 = 1.2*10^{-4}$ ($\pm 3.2*10^{-6}$), $\tau_2 = 3132$ ($\pm$ 78.1), $\chi^2 = 8.4*10^{-12}$, $R^2 = 0.999$

Fig. 5.13. Afterpulse probabilities as a function of time after an avalanche.

Figure 5.14. Set-up used for testing the NFAD with an external-applied hold-off time.

Figure 5.15 Photograph of the experimental set-up used for the hold-off time characterization.

Figure 5.16. (a)Noise characteristic of the NFAD tested in free-running mode with 10^4 photons per second for three different values of the hold-off time. (b)Zoom of the first part of graph (a).

Fig A5.1 Noise characteristic of the NFAD tested in free-running mode with 10^5 photons per second for three different values of the hold-off time, (a), and zoom of the first part of the graph ,(b).

Figure A5.2. Evolution of quantum, (a) and (b), and effective, (c) and (d), efficiencies of the NFAD tested in free-running mode with 10^5 photons per second for four different values of the hold-off time.

Figure A5.3. Free-running mode; dark counts frequency versus effective efficiency (a), with a zoom of the first part of the graph (b), when 10^5 photons per second are sent to the NFAD.

Figure A6.1. Noise characteristic of the NFAD tested in free-running mode with 10^4 photons per second for three different values of the hold-off time.

Figure A6.2. Evolution of quantum, (a) and (b), and effective, (c) and (d), efficiencies of the NFAD tested in free-running mode with 10^5 photons per second for four different values of the hold-off time.

Figure A6.3. Free-running mode; dark counts frequency versus effective efficiency (a), with a zoom of the first part of the graph (b), when 10^6 photons per second are sent to the NFAD.

List of tables

Table 3.4: Gate width of 50 ns and 10 photons per pulse.

Tables 4: Hold-off time measures for 10^4 photons per second using different hold-off times.

Table 4.1: Data obtained with a hold-off time of 10 μs.

Table 4.2: Data obtained with a hold-off time of 5 μs.

Table 4.3: data obtained with a hold-off time of 1 μs.

Tables 5: Hold-off time measures for 10^5 photons per second using different hold-off times.

Table 5.1: Data obtained with a hold-off time of 10 μs.

Table 5.2: data obtained with a hold-off time of 5 μs.

Table 5.3: Data obtained with a hold-off time of 1 μs are presented.

Tables 6: Hold-off time measures for 10^6 photons per second using different hold-off times.

Table 6.1: Data obtained with a hold-off time of 1 μs.

Table 6.2: data collected with a hold-off time of 0,5 μs.

Table 6.3: data obtained with a hold-off time of 0,2 μs are presented.

Abstract

Il conteggio di singoli fotoni tramite single photon avalanche diode (SPAD) in InGaAs/InP operanti nella terza finestra telecom (1550 nm), è affetto dall'afterpulsing. Tale effetto corrisponde alla generazione di impulsi spuri, costituiti da valanghe spontaneamente innescatesi in seguito alla comparsa di una valanga provocata da un fotone incidente. Questo fenomeno è dovuto al rilascio di portatori intrappolati nella regione di moltiplicazione del diodo. La probabilità di intrappolamento dipende da quanta carica scorre attraverso il diodo durante una valanga. L'afterpulsing può dunque essere mitigato limitando la durata della valanga; ciò è possibile, o polarizzando il diodo solo all'interno di una breve finestra temporale, oppure utilizzando circuiti rapidi (attivi o passivi) che arrestino rapidamente tale valanga. Nei Negative Feedback Avalanche Diodes (NFAD) un resistore, costituito da un film sottile integrato col diodo, garantisce una retroazione negativa passiva in grado di arrestare rapidamente la valanga.

Tale diodo, inserito in un apposito circuito, è stato analizzato per determinarne le caratteristiche principali, quali efficienza, rumore, risoluzione temporale e afterpulsing.

In seguito abbiamo sviluppato un set-up controllato da un FPGA (field–programmable gate array) in grado di effettuare la completa caratterizzazione del dispositivo. Tale set-up permette di caratterizzare, in particolare, l'effetto dell'afterpulsing a partire da condizioni iniziali note: l'impulso laser è applicato al diodo solo se in precedenza, per un dato intervallo di tempo, non sono state rilevate delle valanghe. In seguito, se l'impulso provoca una valanga, i successivi afterpulses sono

registrati in funzione del tempo. E' possibile sia registrare solo gli afterpulses primari, sia gli afterpulses degli afterpulses.

Infine, abbiamo realizzato un circuito esterno in grado di fissare un tempo morto in seguito ad ogni detezione di una valanga, per limitare l'impatto dell'afterpulsing; con tale circuito sono state poi registrate misure di efficienza, per valutare l'impatto dei diversi tempi morti utilizzati e per mostrarne l'effetto nella soppressione dell'afterpulsing.

Introduzione

L'obiettivo di questa tesi è la realizzazione e l'analisi di un detector per la rivelazione di singoli fotoni, realizzato a partire da un Negative Feedback Avalanche Diode (NFAD) fabbricato presso i laboratori della Princeton Lightwave.

La rivelazione di singoli fotoni, aventi lunghezze d'onda appartenenti alla cosiddetta "terza finestra telecom" (centrata a 1550 nm), ha assunto importanza crescente a partire dalla seconda metà degli anni '90, in quanto sono via via maturati numerosi ambiti di applicazione; tra questi, particolare interesse è suscitato dalla Quantum Key Distribution (QKD), ovvero dallo scambio fra due utenti di una chiave che verrà utilizzata per trasmettere in maniera sicura un messaggio che vuol essere mantenuto segreto. Tale scambio avviene seguendo precisi protocolli che sfruttano particolari proprietà della luce, derivanti dalla meccanica quantistica. Nonostante siano già disponibili sul mercato numerosi sistemi di QKD, molti sono ancora i miglioramenti necessari, ed il "collo di bottiglia" per lo sviluppo della QKD è rappresentato proprio della problematiche legate alla rivelazione di singoli fotoni.

Nel capitolo 1 della tesi vengono introdotti i principi che stanno alla base della QKD. Questa tecnologia viene presentata inserendola all'interno del più ampio capitolo della crittografia, un insieme di tecniche atte a rendere un messaggio incomprensibile a qualsiasi utente non autorizzato a conoscerlo. I due utenti del sistema di QKD sono per tradizione battezzati Alice e Bob; essi, per potersi trasmettere a lunga distanza la chiave segreta per criptare o decriptare i propri messaggi, devono essere connessi tramite due canali, il primo classico ed il secondo quantistico. Entrambi i canali sono

implementati con delle fibre ottiche, le cui proprietà sono riassunte al termine del capitolo.

Nel capitolo 2 viene presentato il problema della rivelazione di singoli fotoni e le varie tecniche sviluppate a partire dagli anni ’60 per poterlo risolvere. Tra di esse, particolare attenzione viene dedicata ai “Single-Photon Avalanche Diodes” (SPADs), realizzati a partire da degli “Avalanche PhotoDiode (APDs) operanti in modalità Geiger, ovvero in grado di generare una “valanga” di carica elettrica (costituita da circa 10^6 elettroni) a partire da un unico elettrone fotogenerato. Gli SPAD sono in seguito suddivisi in tre categorie, differenti per l’intervallo di lunghezze d’onda nel quale sono sensibili alla luce e quindi per il materiale fotosensibile con il quale sono realizzati: Silicio, Germanio, lega ternaria di InGaAs. Quest’ultimo caso viene ulteriormente approfondito, in relazione alla complessa struttura del diodo, comprendente vari strati costituiti da diverse leghe ternarie e quaternarie di semiconduttori.

Nel capitolo 3 sono presentati i vari parametri utilizzati per valutare le performances e le condizioni operative degli SPAD. Viene per prima definita l’efficienza ed i vari fattori che la determinano; poi sono introdotti i conteggi di buio, suddivisi rispettivamente in conteggi di buio originati termicamente e conteggi di buio generati da tunneling interbanda. Infine, ampio spazio è dedicato al principale problema che deve essere risolto per migliorare le prestazioni degli SPAD realizzati in InGaAs: il cosiddetto afterpulsing. Esso consiste nell’emissione da parte del diodo di impulsi spuri di carica, dovuti alla moltiplicazione di elettroni rimasti “intrappolati” nella regione di moltiplicazione del diodo durante una precedente valanga. Viene quindi presentato un modello in grado di spiegare la densità di probabilità temporale dell’afterpulsing e, infine, viene spiegato come applicare tale modello all’analisi dei dati misurati.

Una volta generata, la valanga di carica utilizzata per segnalare la presenza di un singolo fotone deve essere rapidamente fermata, riportando il diodo nella condizione iniziale. Non potendo arrestarsi autonomamente, si rendono necessarie delle tecniche di quenching delle valanghe. Tali tecniche sono presentate nel capitolo 4. Innanzitutto è introdotta la prima e più semplice tecnica emersa, il quenching passivo, spiegando le modalità di funzionamento dei circuiti utilizzati per implementarla; particolare riguardo è riservato al calcolo della costante di tempo del quenching e della carica contenuta nella valanga, parametri chiave per determinare la quantità di carica intrappolata in seguito ad una valanga e la relativa probabilità di afterpulsing. In seguito viene introdotta la tecnica del quenching attivo, con uno sguardo rivolto anche alla sua implementazione realizzata con un circuito integrato specifico (ASIC). Infine, viene brevemente introdotta la modalità di funzionamento "gated" del detector, la quale prevede che il diodo sia attivo solamente per un sequenza di brevi intervalli di tempo che si ripetono con una frequenza regolare, in maniera tale da ridurre l'impatto dell'afterpulsing sulle prestazioni del detector; in tale paragrafo vengono anche accennate alcune tecniche recentemente sviluppate in grado di garantire un gating rapido del diodo.

Il capitolo 5 rappresenta il cuore della tesi, in quanto contiene i principali risultati sperimentali ottenuti e un'approfondita spiegazione del funzionamento del programma realizzato per controllare una FPGA (field-programmable gate array), utilizzata per caratterizzare il comportamento del diodo. Nel primo paragrafo, vengono presentate le proprietà innovative dei Negative Feedback AvalancheDiodes (NFADs), categoria alla quale appartiene il diodo in esame; nel secondo viene spiegato il set-up utilizzato per le misure di efficienza, conteggi di buio e afterpulses; nel terzo e quarto

paragrafo sono raccolte ed analizzate tali misure, introducendo anche il metodo “a doppia finestra”, utilizzato misurare la probabilità di afterpulsing. Nel quinto paragrafo viene introdotta la tecnica di misura degli afterpulses basata sull’utilizzo dell’FPGA, spiegando il set-up utilizzato, i risultati ottenuti ed ottenibili con tale tecnica e analizzando i dati raccolti. Nel sesto paragrafo si introduce il concetto di tempo morto (o hold-off time) successivo alla generazione di ogni valanga e se ne spiega la necessità: esso serve a ridurre la probabilità di afterpulsing, evitando che gli elettroni intrappolati, una volta liberati, generino delle valanghe spurie di corrente; viene così illustrato il set-up utilizzato per le misure effettuate al variare dell’hold-off time, discussi ed analizzati i risultati ottenuti. Infine, nel settimo paragrafo sono illustrate le misure relative al jitter del dispositivo.

Il capitolo 6 discute le potenzialità del NFAD testato per la sua applicazione all’interno di un sistema di QKD basato su di un protocollo recentemente proposto dal gruppo di ricerca del GAP-optique; tale analisi è basata su di un report sviluppato da Nino Walenta.

Infine, il capitolo 7 presenta le conclusioni della tesi. Seguono l’appendice, contenente le tabelle che racchiudono tutti i dati presentati ed analizzati nel testo, ulteriori grafici non presentati nel capitolo 5, e la bibliografia.

1. QKD, cryptography and optical fibers.

Cryptography is an ensemble of techniques which provide a message to be unintelligible to any unauthorized party. It is a part of the broader field of cryptology , which also includes cryptoanalysis, the art of code breaking. To achieve this goal, an algorithm (also called a cryptosystem or cipher) is used to combine a message with some additional information (known as the key) in order to produce a cryptogram. If a cryptosystem is secure, it is impossible (or extremely difficult) to unlock the cryptogram without the key[1].

Modern cryptography lies at the interception of many disciplines, like mathematics, computer science, quantum mechanics and electronics[58].

In particular, the scenario of modern cryptography was deeply changed after the early 1980s, when Bennet and Brassard have proposed a solution to the key distribution problem based on quantum physics [59]: the BB84 protocol.

This idea was independently re-proposed by Ekert [60] a few years later and became the beginning of the most promising task of Quantum Cryptography (QC), the so-called Quantum Key Distribution (QKD).

During recent years, QC has rapidly progressed [55][1], thus leading to the development of commercial products [56]. It was shown that this technology is capable to be useful in cryptography, in addition with the classical cryptographic systems[55][57].

1.1 Basics of Quantum Key Distribution (QKD).

1.1.1 General settings.

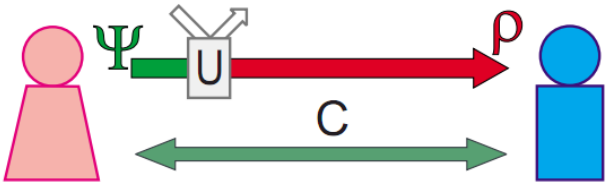


Figure 1.1 The setting of QKD. Alice and Bob are connected by a quantum channel (the upper one) on which Eve can tap without any restriction other than the laws of physics, and by an authenticated classical channel (the lower one), which Eve can only listen to [58].

The generic settings of QKD are schematically depicted in Fig. 1.1; Traditionally the two users, who want to establish a secret key at a certain distance, are named Alice and Bob. This two authorized partners need to be connected by two channels: a quantum channel, which allows them to share quantum signals and a classical channel, which they can use to send and receive classical messages.

The first problem which has to be solved is the authentication of the classical channel, that is the capability for Alice and Bob to identify themselves; a third person can only listen to the conversation, but cannot participate in it. However, the quantum channel is totally transparent to any third person who wants to manipulate the quantum signal flowing on it.

The task of Alice and Bob is thus to guarantee security against an adversarial eavesdropper, usually called Eve (from the assonance between these two words)., who would tap on the quantum channel and listen to the messages sent on the classical one.

The term security has to be intended in this way: “a non-secret key is never used”, that is either Alice and Bob can indeed create and share a secret key, or they abort the protocol.

Therefore, after they have transmitted a sequence of symbols, Alice and Bob can estimate the amount of information has leaked out to Eve. This estimation is impossible in classical information: if someone is tapping on a classical communication channel, the information goes on unchanged. It is evident at this point how quantum physics can help us: in a quantum channel, leakage of information is quantitatively related to a degradation of communication[58].

1.1.2 The security of QKD.

One of the basic statements of quantum mechanics says: “one cannot take a measurements without perturbing the system”[1] (unless the quantum state is compatible with the measurement). If we apply this axiom to the problem of QKD between Alice and Bob, it shows why it can be useful: since it applies also to Eve, the eavesdropper, she cannot get any information about

the communication without perturbing the system and thus revealing her presence. Alternatively, we can apply the no-cloning theorem [61], which states that one cannot duplicate an unknown quantum state while keeping the original intact[58].

The fact that security of QKD is based on general principle of quantum mechanics suggests the possibility of unconditional security, that is the possibility of guaranteeing security to a communication system, without restricting the power of the eavesdropper. Today, unconditional security has been proved for several QKD protocols.

1.2 The choice of light and optical fibers.

In general, quantum information processes could be implemented with many system, like ions, atoms, light, spin, etc…This statement is valid for QKD, too: it can be performed with electrons, molecules, ions, however, light is the only practical choice[58]. In fact, light does not interact easily with matter and quantum states of light can be transmitted to distant locations without decoherence[58].

Another problem remains to solve: the scattering, i.e. losses. Losses affects QKD by imposing limits to the maximum key-rate (it cannot scale with distance better than the transmittivity of the line) and on the achievable distance (at a certain distance the signal reaches the noise level)[58].

In summary, QKD is always implemented with light and there is no reason to believe that things will be change in future[58]. Therefore, the quantum channel might

be a medium that propagates light with reasonably low losses : typically, either an optical fiber or free space.

The choice of the medium is directly connected with the questions of the source and detectors; since they have to be compatible, the crucial choice is that of the wavelength: we can choose both a wavelength around 800 nm, for which photon counters in Si are available (see section 2.2.1), or a wavelength which can exploit telecom optical fibers, i.e. around 1300 or 1550 nm, for which we can use Ge and InGaAs detectors(see section 2.2.2 and 2.2.3)[1].

The first choice requires free-space transmission or special optical fibers, hence the installed telecom networks cannot be use; the second choice required the improvement of detectors not based on silicon, which must be transparent above a wavelength of 1000 nm.

Actually, the second choice seems the best one and many industries and research groups are working on it. A good reason to support this solution is that the quality of telecommunication fibers is higher than that of any special fiber[1], in particular the attenuation is much lower: at 800 nm, the attenuation is about 2 dB/km (i.e., half of the photons are lost after 1.5 km), while it is only of the order of 0.35 and 0.20 dB/km at 1300 and 1550 nm, respectively (half of the photons are lost after about 9 and 15 km). Around this two wavelength, in fact, the two “telecom windows” are located, and the attenuation is minimal, as we can see in Fig. 1.2.

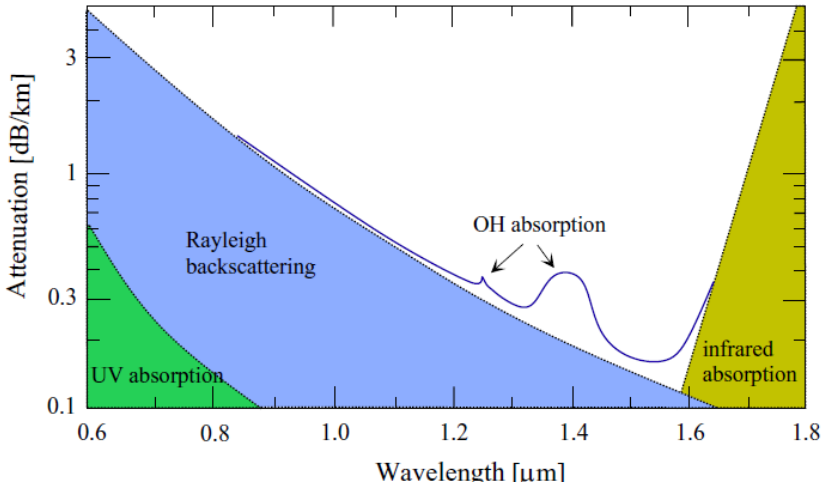


Figure 1.2. Transmission losses as a function of wavelength in optical fibers. At lower wavelengths, absorption is due to electronic transitions in SiO_2, while at higher wavelengths absorption is due to vibrational modes. Superimposed is the absorption due to Rayleigh backscattering and to the transitions in OH groups. Modern telecommunications systems work on wavelengths around 1300 nm (the second telecommunications window) and 1550 nm (the third telecommunications window)[1].

During all the experiments which will be presented in this thesis, we have worked at a frequency of 1550 nm, that is in the third telecom window.

2. The problem of single-photon detection and the Single-Photon Avalanche Diodes.

The ability to detect the light at a single-photon level is the ultimate level of sensitivity for many optical techniques, and nowadays there are many applications which rely on this, because they involve physical processes in which only a very small number of photons, often just one, are available for detection [45].

With the availability of pseudo-single-photon and photon-pair sources, the success of quantum cryptography essentially relies on the availability of single-photon detectors [1]. Other quantum optics applications of single-photon detectors are quantum computing [48] and fundamental studies of quantum physics [49].

This kind of devices has also been employed successfully in optical fiber characterization by optical time domain reflectometry[2][3], semiconductor material studies, faint object imaging for astronomy applications, single-photon three-dimensional LADAR imaging[46], free-space optical communications in photon-starved environments [47] or even for detect brain activity using a functional diffuse reflectance spectroscopy[4].

Photon counters can also be used for detection of very fast optical signals exploiting their fast optical response; as an example of such application, there is the measurement of fast fluorescent decays by means of the technique named Time-Correlated Single-Photon Counting (TCSPC)[5][6].

The ideal detector should fulfill the following requirements:

- The ability to detect photons of different wavelengths with the same efficiency The probability of generating noise, that is a signal without a photon arriving on the diode, should be small.
- The time jitter of the whole system should be small, in order to have a time between detection of a photon and generation of the corresponding signal as constant as possible.
- The recovery time of the diode (i.e., the dead time) should be short, in order to allow high repetition frequencies for arriving data.
- The diode must be practical, so that it will be possible to commercially develop it.

In the next paragraphs, the most important techniques developed for single-photon counting will be shown e shortly described. Firstly, the PhotoMultiplier Tubes (PMTs) will be introduced, then there will be a short description of Superconducting Single Photon Detectors (SSPDs) and finally Semiconductor Avalanche PhotoDiodes (APDs), from which SPADs have been originated, will be described.

2.1 Techniques for single-photon counting

There are a lot of devices for counting single-photons, for example photomultiplier tubes, avalanche photodiodes and superconducting Josephson junctions.

PhotoMultiplier Tubes (PMTs) exploit the internal electron multiplication process (cascaded secondary electron emissions) and attain an internal gain of about 10^6, thus producing electrical signals which can be discriminated from the noise of the read-out electronics. Compared with microelectronic solid-state detectors, PMTs present many disadvantages, both practically (size, difficult of use, etc. …) and related with their performances (lower quantum efficiency) [13].

Despite their disadvantages, both photon counting and time-correlated photon counting techniques were developed thanks PMTs. High performance PMTs were produced industrially from the 1960s[25].

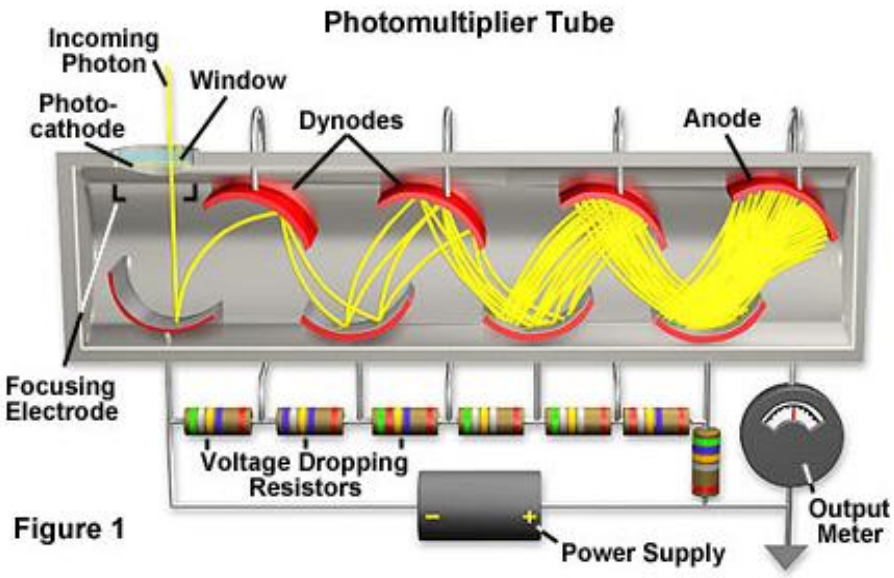


Figure 1

Fig 2.1 Photomultiplier Tube (PMT). Photomultipliers acquire light through a glass or quartz window that covers a photosensitive surface, called a photocathode, which then releases electrons that are multiplied by electrodes known as metal channel dynodes. At the end of the dynode chain is an anode or collection electrode. Over a very large range, the current flowing from the anode to ground is directly proportional to the photoelectron flux generated by the photocathode [33].

The Superconducting Single Photon Detectors (SSPDs) present low noise and low time jitter (estimated response time 30 ps, registered time due to apparatus limitations 150 ps), but work at cryogenic temperatures (2,4 K) [19]. It can reach high count rates (in the GHz range), but it has a small active area (10 μm * 10 μm) and it requires bulky cryostats.

The Semiconductor Avalanche PhotoDiodes (APDs) have the typical advantages characterizing the solid-state devices (small size, low bias voltage, low power consumption, ruggedness and reliability). However, the internal gain of APDs operated in linear mode (which means that the output current is proportional to the input optical power) is not sufficient to detect single-photons. This problem can be solved using the avalanche photodiodes (APD's) operated in the so-called Geiger mode (Single Photon Avalanche Diodes, or SPADs).

There are three different kinds of such detectors, depending to the material used to build them: silicon, germanium or ternary alloys (InGaAs).

The most important categories for our purposes are InGaAs/InP APD's, since they are able to detect photon's in the so-called "near-infrared wavelength regime" (within 1000nm and 1650nm), thus covering both the second and three third telecom windows of optical fibers, which are situated around 1310nm and 1550nm respectively.

Compared with PMTs, SPADs became available to research groups much later, in the 1990s. However, the studies which would lead to the development of these solid-state detectors started much earlier, in the 1960s, during studies of avalanche phenomenon in p-n junctions at the Shockley laboratory[25].

The group of Shockley scholars, and in particular R. H. Haitz, was the first which gave an insight into the phenomenon of triggering of macroscopic voltage pulses by the absorption of single-optical photons, which happened in p-n junctions reverse biased above the breakdown level. (see for example [26] [27] and [28]). In particular they identified the physical phenomena involved in the generation of the pulses, that is pulses due to detected photons and pulses due to noise (dark counts and afterpulses)[28], and then they depicted a model of the device biased above breakdown[25].

However, to see the application of these devices as practical solid-state avalanche detectors of single photons, it was necessary to wait until the 1990s. In fact, the features of the junctions and of the avalanche-quenching circuits (see section…) available at those times, strongly limited the application of these devices as photodetectors[25].

In order to complete this overview of single-photon detectors, we can also mention novel proposals such as using a single-electron transistor consisting of a semiconductor quantum dot [38].

2.2 Single-photon avalanche diodes (SPAD) or Geiger mode avalanche photodiodes (GAPD): introduction.

Single-photon avalanche diodes (SPADs) is a semiconductor device (a p-n junction) that is able to detect single photons; in contrast to linear avalanche photodiodes (LAPDs) it operates in Geiger-mode, in which the applied reverse bias voltage V_a overcome the breakdown voltage V_{bd} [8], and thus it is more similar not to an amplifier but to a bi-stable circuit. In literature they are also known as Geiger-mode avalanche photodiodes (since their operation principle is similar to that of the Geiger-Muller counters of ionizing radiations) or triggered avalanche detectors.

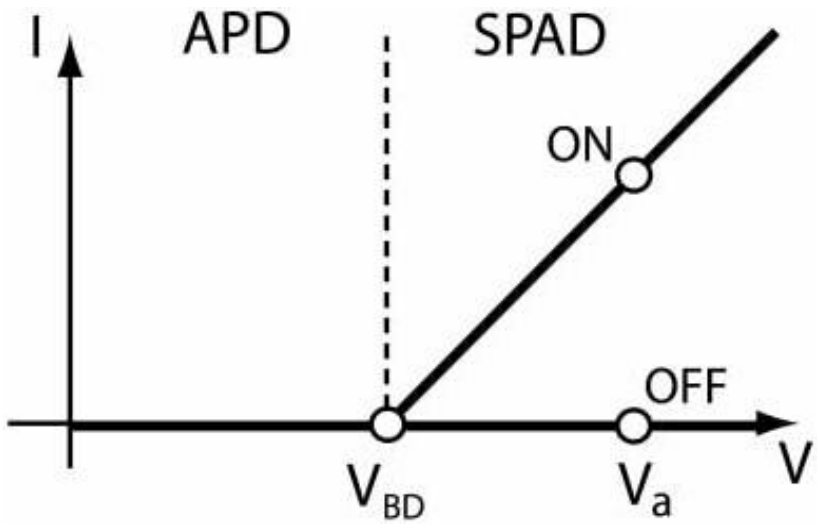


Figure 2.2 Regions of ADP and SPAD operation in the (reverse) I-V characteristics of a p-n junction. (from [25])

In the quiescent state the SPAD is biased at a voltage V_a above the breakdown voltage V_{bd} but no current flows in it, since, even if in the depleted region of the junction the electric field is very high, no free carriers are present. This is the OFF-state.

When even a single charge carrier is injected into this high-field region, afterwards it is accelerated and, due to impact ionization, it can generate a secondary e-h pairs; then, the secondary e-h pairs themselves may generate even more e-h pairs, so triggering a self-sustaining avalanche. This avalanche leads to an electric signal that can be registered and counted as a detection event. The fast onset of the current marks the time of arrival of the photon which have previously generated the first charge-carrier[25]. This is the ON-state.

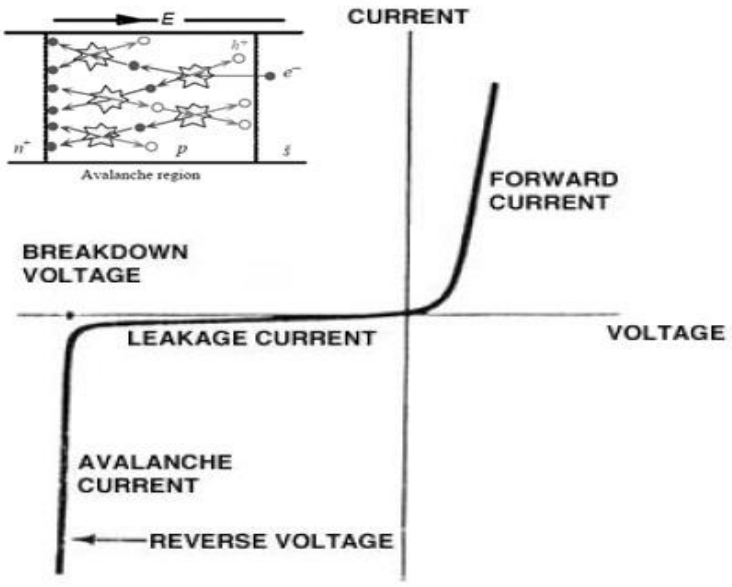


Fig. 2.3 I-V curve of a diode; inserted figure depicts the process of avalanche multiplication of an electron generated by a single-photon [32].

After the detection of the avalanche, this macroscopic current must be quenched by lowering the applied bias voltage below the breakdown voltage V_{bd}; after quenching, the bias voltage must be restored, in order to make the diode able to detect another photon. Doing that, the diode is finally came back to its initial OFF-state.

All these operations are done by a circuit, usually referred as the quenching circuit, which is crucial for establishing the operating conditions and performances of the diode. There are two main categories of such circuits: passive and active quenching circuits; the diode can also work in gated-mode. All these possibilities will be discussed later. It is also possible to use two different kind of circuits for the operations of quenching and resetting of the bias voltage, for example a passive quenching circuit coupled with an active resetting one.

It is important to note that the diode is insensitive to any subsequent photon arriving in the time-interval between the avalanche onset and the voltage-reset, because during this period its voltage Is lower than its breakdown voltage; this time interval is thus an intrinsic dead time for the detector[25].

Finally, it is evident that, regarding a SPAD, speaking of gain has not any sense just as in the case of a bi-stable circuit.

Geiger mode avalanche photodiodes are useful in many applications as almost ideal photon counting detectors. They can be used in an array for spatially resolved single-photon counting, both in two dimensions [15](photon counting imager for faint

object imaging, for example in astronomy applications) and in one dimension for time correlated photon counting experiments with spectral sensitivity.

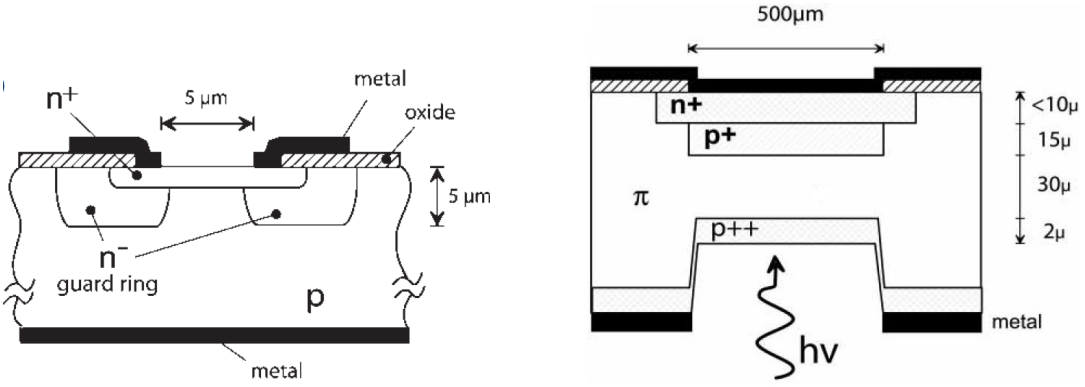


Figure 2.4 (a) Schematic cross-section of planar p-n diode developed by Haitz *et al.* [26][27][28]; (b) Schematic cross-section of an APD with reach-through structure developed by McIntyre and Webb[29]. Figures are from [25].

2.2.1 Silicon SPAD devices.

Since Silicon has been widely used in the semiconductor industry and in research laboratories, considerable progress has been achieved in designing and fabricating of Si-SPADs and nowadays devices with very good characteristics are commercially available[18]. Thanks to these progresses, now it is possible to fabricate Si-SPADs with a low defects concentration, thus reducing effects such as trap-assisted

tunneling (responsible for dark counts) and afterpulses (which are caused by carriers trapped in defects).

Si-SPAD can be divided into three groups, according to the different widths of the depletion layer of the p-n junction:

- a thick type (around 30 μm), also named Slik™ device (Slik™ stands for "super low k", where k is the ratio of the ionization coefficient of holes to that of electrons), which was developed at the former RCA ElectroOptics, now PerkinElmer Optoelectronics. It was successfully employed to produce single-photon counting modules (SPCM) and it is an evolution of the reach-through avalanche diode structure developed by McIntyre and Webb [29][30].
- Two thin types (depletion layer width of ≈ 1μm) which can be fabricated either by dedicated epitaxial fabrication methods or exclusively CMOS processes[8].

These two types of Si-SPAD have the advantage that, for them, monolithically integration between diode and electronics is possible[8][31].

Comparing thin and thick types of Si-SPAD, we can summarize that the first ones are more robust and rugged, costless, compatible with array detectors and integrated circuits, not requiring cooling due to low power dissipation and finally they have a better photon timing resolution (a few 10 ps versus a few 100 ps). On the other hand, thick Si-SPADs present a higher photon detection efficiency and a larger sensitive area diameter (200 μm versus less than 30 μm)[25].

Silicon can detect photon with a wavelength between 450 nm and 900 nm, since it has a band gap of $E_G \approx 1.1$ eV, thus it is not possible to use Si-SPADs to detect photons in the infrared spectral range.

2.2.2 Germanium SPAD devices.

If it is necessary to detect photons with wavelength beyond 1.1 μm , Si-SPADs are no longer usable and semiconductor materials with bandgaps smaller than 1.1 eV must be employed. There are two possibilities: germanium and III-V compound semiconductors; in both cases, detector cooling below room temperature is necessary[25].

Despite Germanium devices are similar to silicon devices, Germanium has worse properties and technology based on this material is less developed; commercially available devices operating, in Geiger mode, are obtained by selecting samples with sufficiently uniform breakdown voltage over the active area, among devices which were designed for operating as APDs.

The main problem of Ge-SPADs is the high dark-counting rate due to the very strong afterpulsing effects: the trap concentration is high and the trap lifetime at low temperature is quite long, 10 μs or more [25]. Detecting photons with wavelength of 1.3 μm is possible biasing the detector with a voltage of a few 0.1 V over the breakdown, obtaining detection efficiencies of a few per cent and a dark counting rate of a few 10^3 per second[25].

Using a quite long dead-time, at least a few microseconds, it is possible to achieve good performances in free-running mode; gated-mode operation is easy and effective in reducing dark-current due to afterpulsing. Good resolution with better than 85 ps FWHM has been experimentally reported.[18]

2.2.3 InGaAs SPAD.

When it is necessary to work with a wavelength belonging to the third telecom window (1,55 μm), the only option is to work with an InGaAs/InP diode [1]. This kind of diodes are useful also for the second telecom window (1,3 μm), where it possible to use APD's made of germanium, too.

The SPADs described in this section evolved from avalanche photodiodes (APDs) originally designed for use in high-bandwidth fiber-optic telecommunications receivers [45].

Traditionally, Ge-made APD's have been implemented in the domain of long-distance quantum communication, but today it is very difficult to find this kind of diodes on the market, since they have been replaced by InGaAs APD's.

Due to the pioneering research reported in 1985 (Levine *et alia*,1985 [9]), now the research focuses on InGaAs APD's in order to detect single-photon in both telecom windows; these detectors have been already used for implementations in the field of quantum cryptography.

Nevertheless many problems remain to solve, first of all the high amount of afterpulses, in order to have single-photon detectors in the third telecom window having the same performances of silicon APD's at lower wavelengths.

2.2.3.1 InGaAs/InP: layer structure.

InGaAs, or indium gallium arsenide, is an alloy of gallium arsenide and indium arsenide. In a more general sense, it belongs to the InGaAsP quaternary system that consists of alloys of indium arsenide (InAs), gallium arsenide (GaAs), indium phosphide (InP), and gallium phosphide (GaP). As gallium and indium belong to Group III of the Periodic Table, and arsenic and phosphorous belong to Group V, these binary materials and their alloys are all III-V compound semiconductors [16]

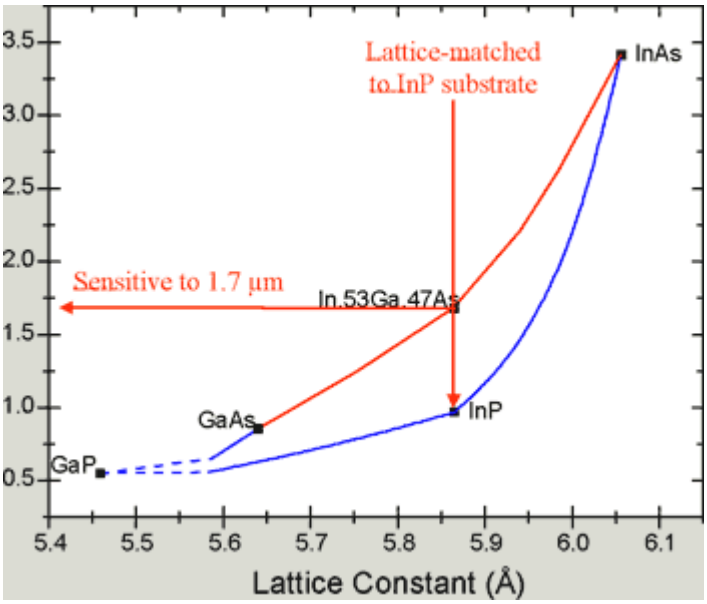


Figure 2.5 Relationship between the long wavelength cut-off and the lattice constants for alloys of the InGaAsP quaternary system.

Figure 2.5 shows the relationship between the lattice constant and the long wavelength cut-off of the ternary alloys which belong to the InGaAsP family. If two or more of the binary alloys of this family are mixed, the resulting ternary or quaternary

compound has electrical and optical properties that are tunable to intermediate values; indeed, these properties rely on energy band-gap and whether the band-gap is direct or indirect. Furthermore, also the lattice constant depends on the material combination of the binary compounds used to make the ternary or quaternary alloy.

Acting on this material combination it is possible to obtain an alloy with appropriate values of long wavelength cut-off (the maximum value for the wavelength of a detectable photon) and of lattice constant; this is useful if it is necessary to match two different compounds, like in the InGaAs/InP SPAD's.

The most suitable material combination for telecom wavelengths is composed by an $In_{0.53}Ga_{0.47}As$ narrow-bandgap layer ($E_g \approx 0.75$ eV @ 295 K) in which photons are absorbed, that is lattice-matched to a wider-bandgap InP layer ($E_g \approx 1.35$ eV) in which avalanche multiplication occurs.

This structure is called "Separate Absorption, Charge and Multiplication" (SACM-structure) and it is necessary because InGaAs is not suitable for multiplication, in contrast with Si-SPAD's, in which the same material is used both for absorption and multiplication. In fact, if we wanted to obtain the multiplication process in this alloy, the electric fields would be so high that they would induce a strong band-to-band tunneling of carriers, due to the small width of the gap. As a consequence, the dark count rate would be very high. The solution is to put another layer in the structure, the InP, which, due to its wider band gap, reduces the dark count rate to tolerable level.

Several InGaAsP grading layers with intermediate lattice constants are sandwiched between the other two, in order to ease the crossing of the heterobarrier by the photogenerated holes. The difference in the band-gap energy among the

absorption and the multiplication layers, in fact, leads to a potential barrier for holes in the valence band; these carriers are thus detained and can pass the barrier with a probability proportional to $e^{-Ebarrier/KbT}$. The figure 3 shows the typical layer order and electric field distribution.

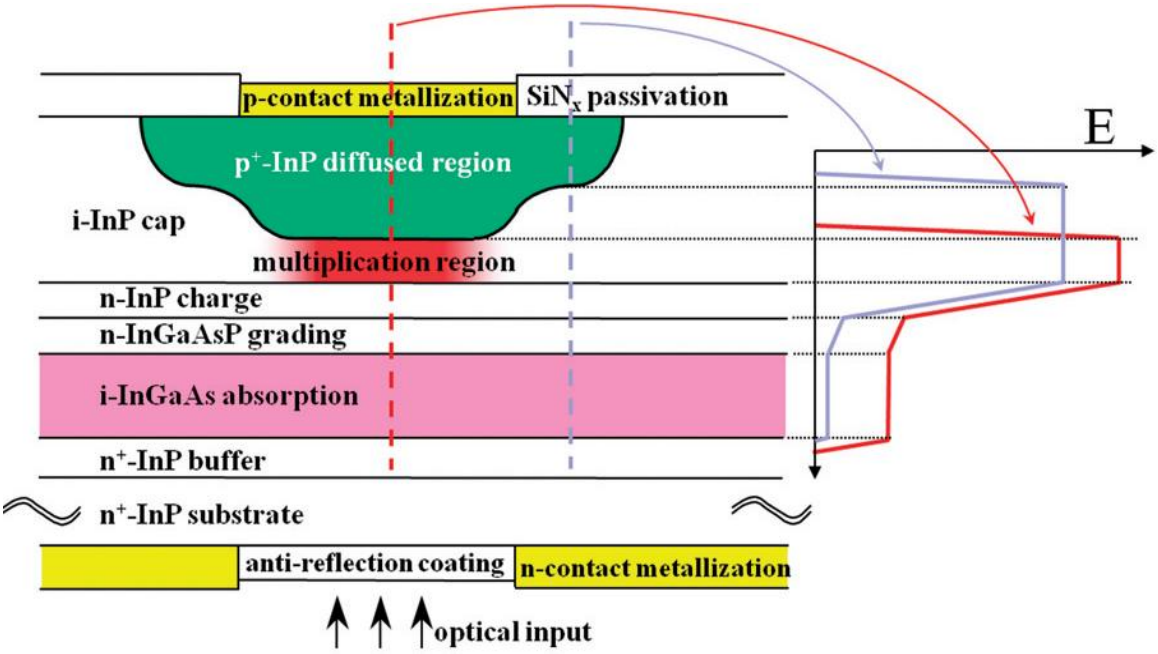


Figure 2.6 Internal structure of InGaAs/InP avalanche diode with SACM structure [11]

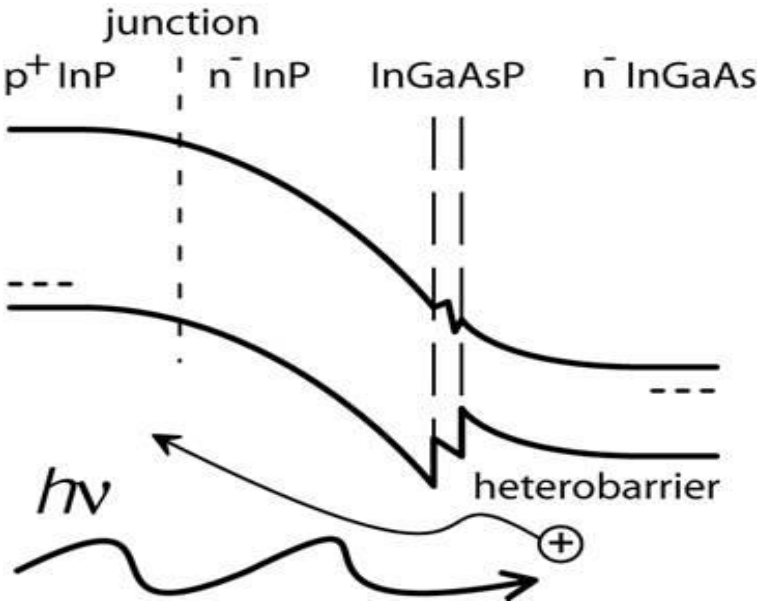


Figure 2.7 Qualitative band-diagram of an InGaAs/InP SPAD with SACM structure and sketch of a free-carrier crossing the hetero-interface [25].

The electric field profile in the multiplication zone must be optimized in order to obtain a compromise between high avalanche triggering probability and low dark count generation (due to trap assisted tunneling and afterpulsing)[10]. The intermediate charge layer serves this purpose: by tuning its width and doping concentration, in fact, the field distribution can be changed[9].

The mixing ratio of Indium and Gallium (i.e.: the index "x" in the formula $In_xGa_{1-x}As$) is chosen to match the lattice constant of the absorption layer with that one of the InP multiplication layer. For this reason, the $In_xGa_{1-x}As$ with x=0.53 is often called "standard InGaAs" The result of this choice is that both materials can be grown on top of each other with a lower amounts of material impurities; this is an important

consequence, since defects (specially for InP) increase the generation of dark carriers and act as trapping centers, which are responsible for afterpulsing.

The fabrication of these diodes involves epitaxial growth of III-V compounds, using different techniques such as liquid-phase epitaxy, vapor-phase epitaxy and molecular beam epitaxy.

During this process of fabrication, it is also necessary to consider another problem which could arise: it is not only necessary to have sufficiently uniform breakdown over the active area, but also a sufficiently high value of the breakdown voltage at room temperature.

Figure 2.8. An example of a commercial InGaAs/InP SPAD: the id210 single photon detection system - near infrared range (900 - 1700nm) [34].

3. Parameters of performance and operating conditions of SPAD.

The most important parameters we have to consider if we want to describe the performances of a SPAD are the detection efficiency η (also named PDE, Photon Detection Efficiency), the dark count probability P_{dc} (or DCR, Dark Count Rate), the afterpulse probability P_{ap} and the timing jitter τ_{jitter}. A primary goal of SPAD device is to optimize the device structure and the profile of the internal electric field in order to minimize DCR and at the same time to maximize PDE.

The noise equivalent power (NEP) of a SPAD is given by

$$\mathbf{NEP} = \mathbf{h\upsilon}\frac{\sqrt{\mathbf{2DCR}}}{\mathbf{PDE}}$$

Where h is the Planck's constant and ν is the optical frequency. The proportionality of NEP to $(DCR)^{1/2}$/PDE provides a performance metric, that is not unique. In fact some applications require different metrics, such as the scaling of the quantum bit error rate (QBER)with DCR/PDE in quantum key distribution.

3.1 Photon Detection Efficiency

The Photon Detection Efficiency is the parameter which gives the ratio of the number of avalanches triggered into the diode and the number of photons that arrive into the diode; it can be decomposed in four different factors, obtaining:

$$\eta = \eta_{coup}\, \eta_{abs}\, \eta_{inj}\, \eta_{aval}$$

The first factor, $\eta_{coup,}$ describe the coupling efficiency at the air-semiconductor boundary, which is determined by the probability of the photon to be coupled to the active area and the probability that it pass the air-semiconductor boundary; to avoid reflection losses an anti-reflection coating is used and to enhance the coupling between the photon and the active area it is necessary to have a proper focusing.

The second factor is related to the second step of photodetection: the absorbing of the photon in the absorption layer; this is a stochastic process regulated by the width of the absorption layer and the absorption coefficient of the material by which it is composed. The probability that a photon is absorbed in a material layer whose width is dz equals αdz, where α is the absorption coefficient; thus, the fraction of photons not yet absorbed decays exponentially (after a distance equal to α^{-1}, the photon flux is reduced by a factor *e*). The absorption coefficient is related to the wavelength of the photons and becomes very small for photons whose energy is smaller than the energy band-gap, because they can only be absorbed by free carriers. Instead, for energy higher than the energy band-gap, a photon can be absorbed by the promotion of an electron from valence to conduction band, thus creating an electron-hole pair [6]. Since InGaAs has a direct band-gap, a phonon is not required to conserve the total momentum.

The photon flux after a distance z from the beginning of the absorption layer is given by:

$$\mathbf{I(z) = I_0 \exp(-\alpha z)},$$

where I_0 is the photon flux at the surface. Thus, the probability that a photon is absorbed in a region of width dz centered around z is:

$$\mathbf{P_{abs}(z)dz = \alpha \exp(-\alpha z)\, dz}.$$

To maximize the absorption efficiency, the absorption layer might be grown as near as possible to the surface of the diode and it should be quite thick, since the fraction of the incident photon flux that will be absorbed into a photodiode whose absorption layer has a thickness of d_{abs}, equals to $1 - \exp(-\alpha d_{abs})$.

The third factor is present due to the fact that, even if a photon is absorbed, generating an electron-hole pair, one of these two carriers (usually the one with the higher ionization coefficient) must be injected into the multiplication layer. Since in InGaAs/InP SPAD's there is a heterobarrier which has to be crossed, the probability of recombination at this interface must be considered, leading to presence of an injection efficiency factor η_{inj}.

The last factor, the avalanche efficiency η_{ava}, is necessary because, even if the carrier has reached the multiplication region, an eventual recombination could still prevent the triggering of an avalanche.

The factors η_{inj} η_{ava}, increase first linearly with the excess bias and then tends to saturate to 1; a higher excess bias is also desirable to reduce the jitter in photon timing, however it should also lead to an higher dark count rate (which increases almost exponentially with excess bias). A trade-off between the two parameters is thus necessary.

Using a calibrated broadband light source followed by a monochromator it is possible to measure the spectral dependency of the detection efficiency; fixing the detection efficiency to 10% at 1550 nm, and scanning the wavelengths, it can be shown that single-photon detection is possible between 1100 and 1650 nm.

3.2.1 Dark counts

The dark counts are caused by avalanches randomly triggered without incident signal at the input of the detector. There are three main mechanisms which induce this self-triggering effect into the diode:

- Thermal excitation in the absorption layer, discussed in the next paragraph
- Band-to-band tunneling processes in the multiplication layer, which will be discussed in the paragraph 3.1.3
- Trap-assisted tunneling after a current is transited into the diode in the multiplication layer, which will be discussed in paragraph 3.3

The first two processes are the only ones which produce avalanches not due to the presence of photons, and they cause dark counts; instead, to see the third process it is necessary to previously have avalanches into the diode (which can be a

consequence of both a laser detection or a dark count), and it lead to afterpulses (which will be discussed in the next paragraph).

The simplest way to reduce dark counts is to cool the detector, so that the number of thermally generated carriers is decreased and dark counts are mostly due to band to band tunneling and, more importantly, trapped charges.

Unfortunately, it is not possible to arbitrary reduce the temperature, because three different problems should rise. First of all, if the temperature of the detector is too low, it could work with worse performances, because the afterpulse probability would be enhanced (see next section). Then, the detector could not function; in fact, when the temperature decreases, the breakdown voltage decreases too, but if it is no larger than the reach-through voltage (the voltage for which the high-field region extends into the InGaAs absorption layer, which is not temperature-dependent) the detector cannot work. Finally, there are practical and economical reasons that prevent to cool down too much the diode: it is better to implement thermoelectric cooling systems based on Peltier cells than liquid nitrogen cooling or other techniques.

The rate of dark counts increases also with the excess bias applied to the diode, because of two effects: the field-assisted enhancement of the emission rate from generation centers and an increase of the avalanche triggering probability [18].

When working in gated mode, the dark counts are usually expressed as the probability to have a dark count per gate or as the dark count probability per nanosecond of gate duration; this is useful when it is necessary to compare results obtained with different gate durations or using the detector in free-running mode.

3.2.2 Thermally generated dark counts

The spontaneous thermal generation of dark counts is dominated by the Shockley-Read-Hall (SRH) effect in the depleted regions, which is mainly due to local defects; these defect, in fact, lead to the presence of trap levels into the band-gap, where the carriers could be trapped, thus splitting their jump from valence to conduction band (or viceversa) into two different steps, making the total process less energetically-expensive. This process will be schematically depicted in the section dedicated to “traps and afterpulsing”, together with electron and hole trapping.

SRH process is relevant for $In_{0.53}Ga_{0.47}As$ avalanche diodes, because of the small band gap (0.752 eV) and it can be reduced improving the lattice quality.[17]

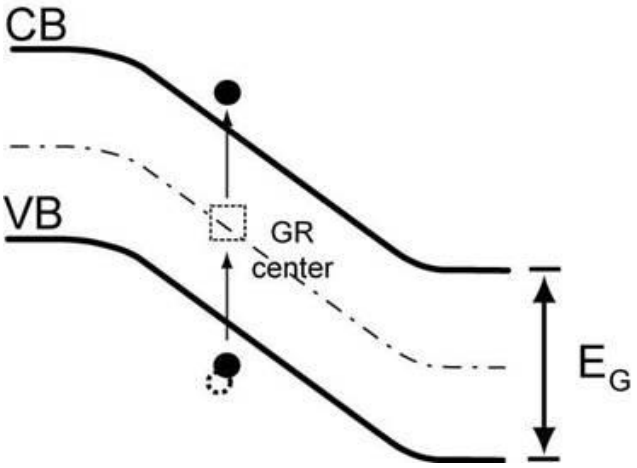


Figure 3.1. Spontaneous thermal generation and SRH effect. A defect generates a trap level into the band gap, thus splitting the jump of a carrier (in this figure, an electron) into two jumps. This is the origin of the so-called afterpulses.

3.3.3 Dark counts generated by band-to-band tunneling

Since the electric field in the multiplication region is very high, electrons can tunnel from valence band to the conduction band; this process generates also a hole in the valence band, therefore it is a process that generates e-h pairs.

In fig. 3.2 is shown what happens to the slopes of the energy bands by changing the electric field into the multiplication layer; due to the increasing of the slope with higher fields, the width of the potential barrier which have to be passed by an electron of the valence band in order to go into the conduction band is reduced. In fact the width of this barrier is $E_g/(qF)$, where E_g is the energy band-gap, q is the charge of an electron and F is the electric field[6].

If the electric field is increased too much, it will become possible for an electron to tunnel from the valence band to the conduction band, thus leaving at the same time a hole in the valence band.

Finally this e-h pair can trigger an avalanche, thus making possible to see a dark count.

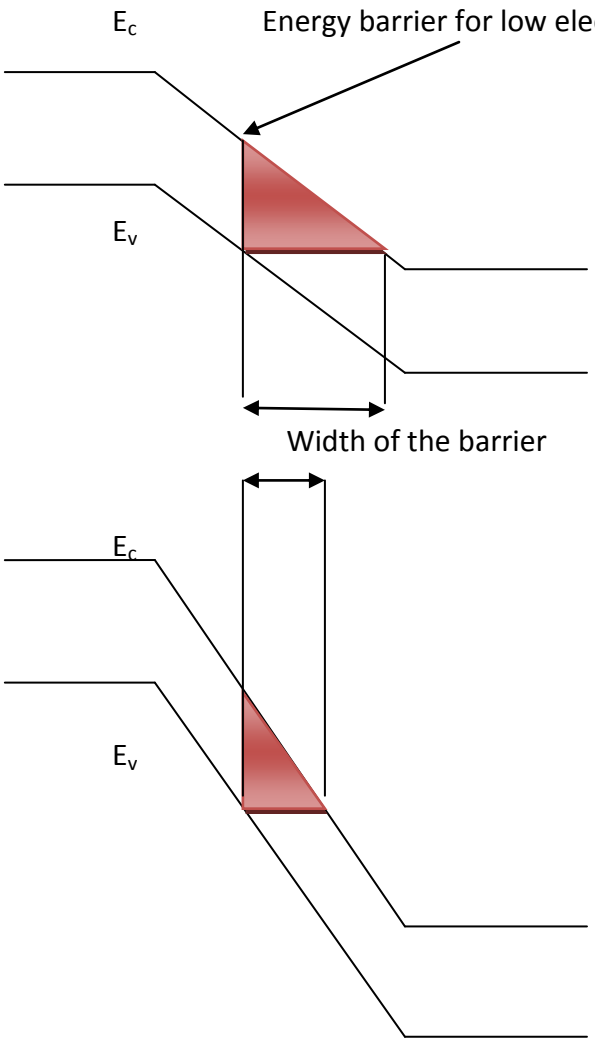


Figure 3.2. Energy bands at low and high electric field: effect on band-to-band tunneling.

3.3 Afterpulsing effects.

Perhaps the major problem limiting the performance of present InGaAs/InP APDs is the enhancement of the dark count probability by so-called afterpulses [14]. These false detection counts are also called “secondary dark pulses”[18]. During an

avalanche many charge carriers flow through the depletion layer of the SPAD and some of them could be captured by trapping centers in the depletion layer; subsequently, they are released with a statistically fluctuating delay, whose mean value depends on the deep levels involved. Finally, these de-trapped carriers can trigger another avalanche, thus generating an afterpulse correlated with a previous avalanche pulse.

The number of trapped carriers during an avalanche pulse increases with the total number of carriers which composes the avalanche that crosses the junction. It is thus evident how much it is important to quench quickly the avalanche (different quenching techniques will be presented in the next chapter) and to reduce the current intensity. However, the current intensity is proportional to the excess bias voltage V_e, and since V_e depends directly on the requested detection efficiency and/or time resolution, the trapped charge per pulse can be limited only by minimizing the quenching delay [18].

If it not possible to reduce the number of trapped charges at a sufficiently low level, it should be useful to implement an additional feature into the quenching circuit, in order to reduce the afterpulses due to the captured carriers at an acceptable level; this feature is the dead time (also named hold-off time in literature), that is the time after an avalanche during which the voltage is maintained under the breakdown value, thus preventing the carriers released from retrigger an avalanche

3.3.1 Traps and afterpulsing.

Trapping centers are defects in the semiconductor lattice, which cause an energy level, E_t, inside the band-gap. They can both capture and emit charge carriers into the conduction and valence bands. The capture and emission processes result in recombination, generation or electron or hole trapping; all these processes are schematically illustrated in figure …, where the arrows show the motion of trap carriers between the energy levels originated by trap centers and the energy bands[6].

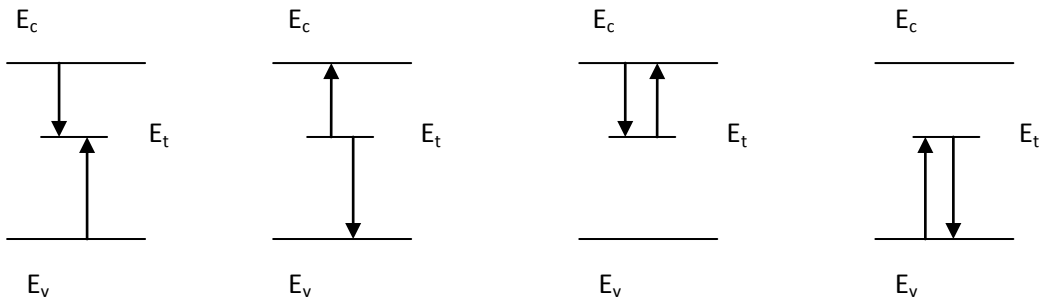


Figure 3.3. Recombination, generation, and electron and hole trapping at centers in the band-gap.

In the first process shown in figure…, the trap captures an electron from the conduction band and then releases it into the valence band: the result of this process is that an electron has left the conduction band and a hole has left the valence band. The name of this process is recombination. This process can also happen reversing the

order of the events: initially an electron could be trapped, then the defect could trap a hole from the valence band and finally it could also trap another electron from conduction band.

The second process shown in figure… is the thermal generation of a couple of carriers: at the beginning an electron from valence band is trapped by a defect, thus leaving a hole. Subsequently this electron is emitted into the conduction band, thus becoming a free carrier. Also in this process the order of the events caould be different

The third and fourth processes are trapping events. In the third one a trap captures an electron from the conduction band and subsequently releases it. The result of this trapping is that the motion of the electron is delayed for a certain amount of time. The fourth process is the same, but the captured carrier is a hole[6].

In practical devices, such as the one tested for this report, many different types of traps may be present; some of them mainly capture electrons, others mainly capture holes. The lifetimes of these traps, that is the time constant of the exponential decay associated with a trap, may differ and may vary with temperature in different ways. Afterpulses due to these trapping processes may therefore be observed within different time intervals, depending on the corresponding lifetime.

3.3.2 A model for afterpulsing probability density in time.

In this model [6][43][44] is discussed the afterpulsing probability due to electrons being captured by traps into the depletion layer. The modeling of hole traps is entirely analogous.

The probability that an electron, which was previously captured by a trap into the depletion layer, located at a certain position z_t, generates an afterpulse when released is equal to $P_{pe}(z_t)$, where P_{pe} is the electron breakdown initiation probability (discussed in [41] and [6]).

The probability that an electron is trapped while traversing a distance dz equals $N_{te}\sigma_{te}dz$, where N_{te} and σ_{te} are respectively the concentration and capture cross section of the electron traps.

Introducing the concept of effective width of the high field region for electron trap afterpulsing[6], here named W_e, we can express the probability that one random electron, generated at a random position during breakdown, is then trapped and afterwards generates an afterpulse as

$$\mathbf{P_{ae} = N_{te}\,\sigma_{te}\,W_e}$$

For holes similar equations can be derived, by introducing the concept of high field region effective width for hole afterpulsing W_h.

To calculate W_e, it necessary to know P_{pe} and the normalized spatial distribution of the ionization events P_g; P_{pe} can be calculated using the method

described in section 2.3.1 of the reference [6], while P_g can be calculated following the method explained in the section 2.7.2 of the same reference [6].

So far, we have obtained the probability that only one of the electrons generated during a breakdown pulse will cause an afterpulse, but when a SPAD breaks down, many electrons are generated.

If *n* electrons are generated, the probability that at least one of them will generate an afterpulse equals to

$$P_{at} = 1 - (1 - P_{ae})^n \approx nP_{ae}$$

And the approximation is good if the product of n and P_{ae} is small, a situation which occurs when the avalanche charge is relatively small.

The probability that k electrons will be trapped and will cause an afterpulse, when n electrons were previously generated in the depletion layer is given by the binomial distribution:

$$P_f(k) = \binom{n}{k} (P_{ae})^k (1 - P_{ae})^{n-k}$$

This equation is only valid if no saturation effects occur, that is when the number of traps is very big; in this case, filling one trap more, will not reduce the concentration of available trapping centers and thus will not reduce the trapping probability.

In the SPADs used for practical applications, the afterpulsing probability should be slow; in this case the product of n and P_{ae} will be small and so the binomial distribution can be approximated by the Poisson distribution, thus obtaining:

$$P_f(k) = \frac{(P_{ae}n)^k}{k!}\exp(-P_{ae}n) \approx \frac{(P_{ae}n)^k}{k!}$$

Now that the probability that a trap will be filled and an aftepulse will occur has been calculated, it remains to calculate when this afterpulse will occur.

The probability that a filled trap will release the captured carrier in a time interval dt is given by dt/τ, in which τ is the lifetime of the trap.

The probability that a trap, which was filled at t=0, will still be filled at time *t*, is given by [43]:

$$P_{tf} = \exp^{\left(-\frac{t}{\tau}\right)}$$

After a photon being detected, the probability for observing an afterpulse within a time interval dt around t equals:

$$P_a(t) = \frac{A\exp\left(-\frac{t}{\tau}\right)}{\tau}dt$$

In which the factor A corresponds, both to the probability that a trap will have been filled during the avalanche triggered by the photon, and to the probability that the captured carrier will start another avalanche after being released from the trap.

The "total" probability for observing an afterpulse after the detection of a photon, named P_{at}, is obtained by integrating $P_a(t)$:

$$P_{at}(t) = \int_0^\infty \frac{A\exp\left(-\frac{t}{\tau}\right)}{\tau}dt = A$$

Thus, since this probability should equal the probability that a trap has been filled during the detection of the photon and also the probability that the captured charge carrier will initiate breakdown, then:

$$\mathbf{A = n\, P_{ae} = n\, N_{te}\, \sigma_{te}\, W_e}$$

If we consider now a practical device, with more than one type of trap, and with traps which can be either electron and hole traps, then the afterpulsing probability density is given by:

$$\mathbf{P_a(t)dt} = \sum_{\mathbf{i}} \frac{\mathbf{A_i}\left(-\frac{\mathbf{t}}{\boldsymbol{\tau}_\mathbf{i}}\right)}{\boldsymbol{\tau}_\mathbf{i}}\mathbf{dt}$$

In which τ_i is the lifetime of the i-th trap located into the SPAD. For every kind of trap there is also a different A_i

$$\mathbf{A = n\, N_{ti}\, \sigma_{ti}\, W_i}$$

In which N_{ti} and σ_{ti} are respectively the concentration and capture cross section of the i-th kind of electron trap, and W_i equals could be equal to W_e or W_h depending on whether the trap is for electons or for holes.

3.3.3 Application of the afterpulsing model for the analysis of measurement data

The model explained in the previous chapter can be applied in order to interpret the measurement data of afterpulsing.

In fact, it is possible to obtain the afterpulsing probability density in time experimentally, as explained in the next chapter. Then the A_i and τ_i values can be obtained by fitting on the measurement results (see chapter…) and can be used to determine the product of N_{ti} and σ_{ti}.

Later it is possible to compute W_n an W_p as explained in [6] and the number of carriers generated in the depleted region during the avalanche process can be calculated considering that the breakdown current reduces the bias over the SPAD from $V_{br} + V_{ex}$ to V_{br}, obtaining

$$\mathbf{n} = \frac{C_p V_{ex}}{q}$$

where C_p is the total parasitic capacitance across the SPAD (see section…), V_{ex} is the excess bias voltage and q is the electronic charge[6].

Finally, the the lifetimes of the traps, τ_i, depend on temperature following the law[43][44]:

$$\tau_i = C \exp\left(-\frac{E_{A_i}}{kT}\right)$$

In which E_{A_i} are the activation energies of the different traps. If we consider an electron trap, for example, this value represents the difference between the energy

level of the trap and the bottom of the conduction band. For a hole trap it is, instead, the difference between the energy level of the trap and the top of the valence band.

In conclusion, it is possible to obtain the activation energies of the different traps from afterpulsing measurements, by taking data at different temperatures.

3.4 Time resolution.

Time resolution, also referred as jitter, is the precision with which the arrival of a photon can be measured. The delay between the detection of an avalanche and a signal photon, in fact, depends on several stochastic processes inside the diode, such as the depth of absorption, charge drift or avalanche built-up.

All these uncertainties lead to a broadened distribution of the time of detection, which can be measured using a TCSPC (time correlated single photon counting) setup[8].

4 Quenching techniques for Single Photon Avalanche Diodes.

As previously mentioned, a Single Photon Avalanche Diodes operates in the so-called Geiger mode; it means that, if it is biased beyond its breakdown voltage V_{br}, once triggered the avalanche current keeps on flowing, thus putting the device into a non-sensitive state. In order to put the SPAD into a sensitive state again, the avalanche process must therefore be stopped and later the device must be brought back into its original quiescent state: this “re-arming” process is made by a quenching circuits.[8][22]

The interruption of the avalanche multiplication process is called “quenching”. The time elapsed from onset until the avalanche is quenched is called “quenching time”. The procedure used to bring the SPAD back into its original quiescent state is called “reset” and the time needed to do this is called “reset time”[22].

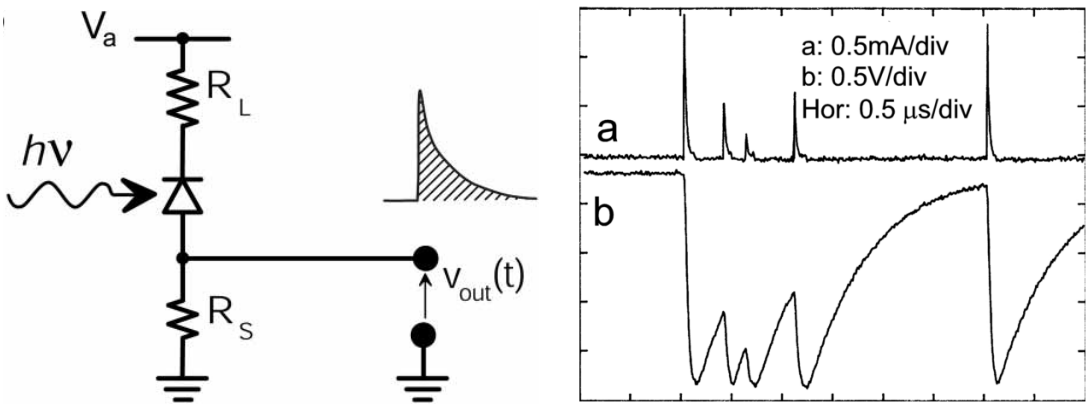
It is very important to choose the most suitable quenching circuit to the desired application, because it directly affects the performances of the device.

In this section we describe three commonly applied quenching techniques: passive quenching, active quenching and gated-mode operation (with a short introduction on rapid gating, a recent developed technique)

4.1 Passive Quenching.

As previously stated, once an avalanche has been triggered, the avalanche current must be quenched; this can be achieved by lowering the SPAD's bias voltage below breakdown, thus stopping the avalanche process into the diode[22].

In the early studies on avalanche breakdown in junctions [23] the avalanche current quenched itself simply by developing a voltage drop on a high impedance load which is in series with the diode. These simple circuits, illustrated in figure… , are still currently employed and have been called passive quenching circuits (PQC's)[24].



(a) (b)

Figure 4.1. (a): Schematic circuit diagram of a SPAD with a passive quenching circuit; Typical values for the resistors are: $R_L \approx 500$ kΩ and $R_S \approx 50$ Ω. When an avalanche current is triggered, a voltage drop is originated over the ballast resistor R_L, which lowers the voltage over the diode and finally quenches the avalanche. (b): Current (upper graph) and voltage (lower graph) evolution in the diode, which shows the retriggering process of a SPAD in a PQC; the speed of recovery depends on the RC constant of the circuit. Extracted from [25].

4.1.1 Operation of a PQC for a SPAD

The SPAD is reverse biased through a high-value ballast resistor R_L (some hundreds of kΩ), C_d is the junction capacitance of the diode (typically ≈ 1 pF), C_s is the stray capacitance, that is the capacitance to ground due to the diode terminal connected to R_L (typically few picofarads).

The diode resistance R_d is the sum of two in-series resistances : the space-charge resistance of the avalanche junctions and the ohmic resistance due to neutral semiconductor which must be crossed by the current [18]. The R_d value depends on the semiconductor device structure. It could be lower than 500 Ω for SPADs with wide area and thick depletion layer and from a few hundred ohms to various kiloohms for devices with a small area and a thin depletion layer[18].

A small resistor R_s (usually 50 Ω, which is the proper value for adapting the impedance with connecting coaxial cables) is connected to the other terminal to observe the current pulse.

In figure… the equivalent circuit of a PQC in current-mode output configuration is shown. The diode is biased with an applied voltage V_A. When an avalanche is triggered, the corresponding action in the equivalent circuit is closing the switch.

In the same figure, the typical waveforms of diode current I_d and diode voltage V_d, or of the transient excess voltage $V_{ex} = V_d - V_{br}$ are shown:

$$\mathbf{Id\ (t) = \frac{Vd(t) - Vbr}{Rd}}$$

Immediately after the triggering, the avalanche current rises to its peak value , given by the excess bias V_{ex} voltage divided by SPAD's resistance R_d[22]. Then it discharges the parasitic capacitances C_d and C_s , so that both I_d and V_d exponentially decrease toward their asymptotic steady-state values of V_f and I_f :

$$\mathbf{I_f = \frac{VA - Vbr}{Rd + RL} \approx \frac{Vex}{RL}}$$

$$\mathbf{V_f = V_{br} + R_d I_f}$$

, where the approximation is due to the fact that it must be $R_d >> R_L$ as shown in the following[18].

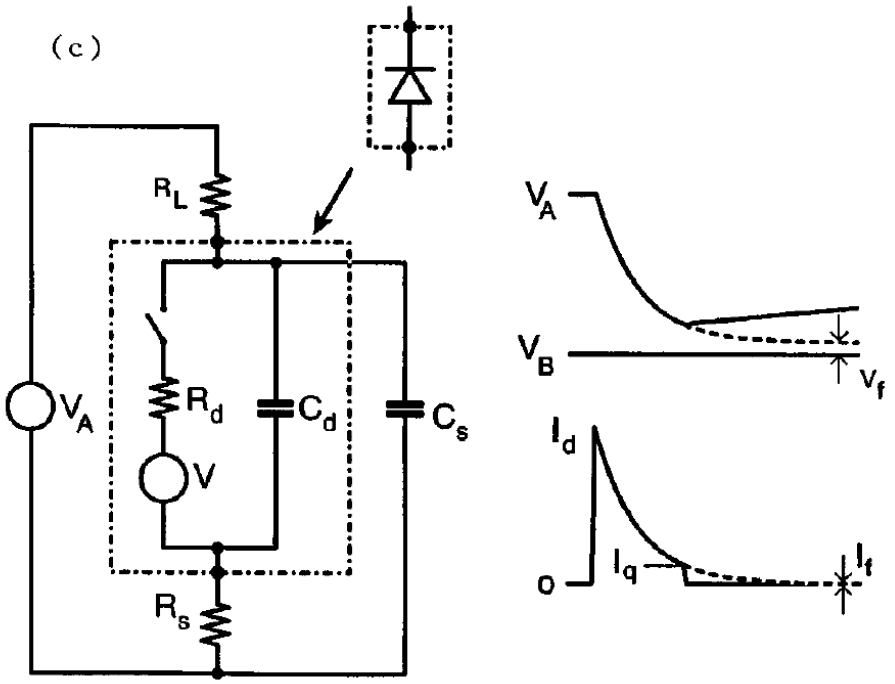


Figure 4.2. Equivalent circuit of a PQC in the current-output mode configuration[18]. The avalanche signal is subsequently sensed by a comparator, which generates a standard output useful for counting the avalanches triggered. On the right the behaviors of diode current I_d and diode voltage V_d is shown. The diode current rises quickly to its peak value and then it decreases exponentially towards I_f and meanwhile the catode voltage also decreases exponentially; when the avalanche reaches I_Q, it is quenched to zero and thus the voltage starts to slowly return to its initial value.

4.1.2 The quenching time constant.

The quenching time constant τ_q is set by the total parasitic capacitance $C_d + C_s$ and by the equivalent parallel resistance of R_d and R_L (if we neglect the resistor R_s, we

can consider the two capacitances and the two resistances in parallel), which is in practice simply R_d:

$$\tau_q = (C_d + C_s)\frac{RdRL}{Rd+RL} \approx (C_d + C_s)R_d$$

Since the diode voltage V_d never falls below the breakdown voltage V_{br}, the avalanche should not be quenched and a current equal to I_f should continue to flow through the device. If this final value is high enough, that is true, and in the diode there will be a self-sustaining avalanche, since enough carriers are present at any one time in the space charge region[22].

Nevertheless, if I_f is very small, V_f is very near V_{br}. At some point the intensity of $I_d(t)$ becomes very low and the corresponding number of carriers that traverse the avalanche region is so small that, since the avalanche process is statistical, it can happen that none of these carriers may impact ionize and the avalanche self-quenches.

The boundary between these two behaviors is not sharply defined, although a value of about 100 μA is often used [18][22], while sometimes a lower value of 20 μA [25]. This imposes a minimum value of the ballast resistor R_L in order to suitably reduce the final current.

As previously said, the avalanche process is statistical and the quenching time is statistical, too. However, a first order estimation, useful for comparisons among different quenching circuit, can be obtained [22].

If we consider the quenching time as the instant the exponentially-decaying current crosses the “quenching threshold” (that is the 100 µA mentioned earlier), therefore the quenching time can be calculated as

$$t_Q = \tau \ln\left(\frac{I0 - If}{IS - If}\right)$$

where $I_0 = V_{ex} / R_S$ is the peak initial value of the avalanche current and I_s is the quenching threshold.

Looking at the last formula, we can notice how it is important to reduce the value of τ in order to reduce the quenching time; hence it is important to minimize the parasitic capacitance of the diode (this is one of the benefits of using the Negative Feedback Avalanche Diodes, which will be introduced in the chapter….); for more recent circuits, with SPADs directly integrated with a PQC, C_P could be even less than 1 pF, with a passive quenching time of some tens of nanoseconds[22].

The ultimate limit in reducing parasitic capacitance of a diode is due to the intrinsic junction capacitance of the detector itself: using a depleted region 1 µm thick and a 50 µm-diameter SPAD (the capacitance is proportional to the square of the diameter), the smallest achievable parasitic capacitance is in the order of 200 fF [22], thus leading to a passive quenching time of some nanoseconds.

4.1.3 The avalanche charge.

The total charge generated during an avalanche process Q_{pc} is an important parameter for evaluating the trapping effects which lead to after pulses; it is given by the integral of the current which flows into the diode from avalanche triggering until quenching (that is the area underlying Id (t) in the figure…) and it depends on the parasitic capacitance, the excess bias voltage and the quenching time[22].

If we consider an avalanche-quenching current $I_{Q,}$ then the quenching voltage V_Q (the minimum voltage reached over the SPAD at the quenching time t_Q) will be

$$\mathbf{V_Q = V_B + I_Q R_d}$$

And we can use this value in order to evaluate the total charge Q_{pc}, obtaining

$$\mathbf{Q_{pc} = (V_A - V_Q)(C_d + C_s) \approx V_E(C_d + C_s) \approx I_f t_r}$$

Where $t_r = R_L (C_d + C_s)$ is the characteristic time constant of the voltage recovery [18]. We can note that, since $R_L >> R_d$, then $t_r >> t_Q$ and the total time between the trigger of an avalanche and the recovering of the initial voltage across the diode is largely due to the recovery time.

Another way to reduce the avalanche charge is to reduce the quenching time by increasing the value of the ballast resistor. In fact, even if the quenching time constant remains almost the same, since $I_f \approx V_{EX}/R_B$ the final value of the current is smaller, then the threshold I_S is crossed earlier[22].

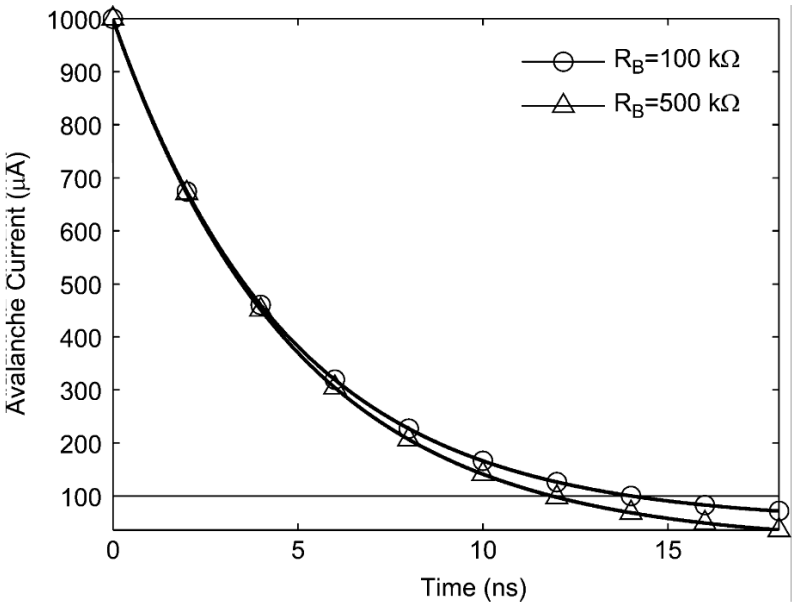


Figure 4.3 Diode avalanche current for two different R_B values[22].

A more accurate estimation of the overall avalanche current can be found in [22], where it is calculated as:

$$Q = V_{EX}\,(C_d + C_s)\left[1 + \frac{I_f \ln\left(\frac{I0-If}{Is-If}\right) - Is}{I_0}\right] = V_{EX}\,(C_d + C_s)\,\alpha$$

Where $V_{EX}(C_d + C_s)$ is the theoretical avalanche charge with infinite ballast resistance R_L, and α is a factor which depends on I_f (when the diode and the excess bias voltage are chosen). Looking at the figure …we can notice that, if the final value I_f is reduced to the 30% or 50% of I_s, the approximation $\alpha \approx 1$ is justified and the avalanche charge will be small. On the contrary, if the final current is large ($I_f \approx I_s$), the avalanche charge will be much larger than the theoretical value.

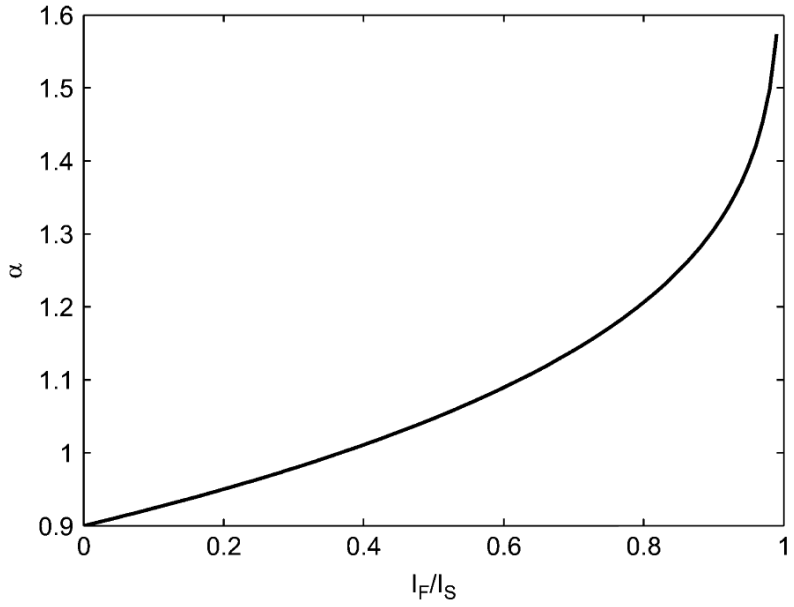


Figure 4.4 Dependency of α on I_f [22].

Moreover, the avalanche charge is directly proportional to the total parasitic capacitance ($C_d + C_s$), so it is possible to reduce the charge by proportionally reducing this capacitance. In particular, the stray capacitance C_s will be smaller if the PQC is directly integrated with the detector.

4.2 Active quenching.

A solution which permits to avoid the drawbacks of passive quenching circuits, mainly slow voltage recovery and not well defined dead time is to use a different kind of quenching circuits: the Active Quenching Circuits (AQCs). They were firstly introduced in 1975 [35], then they were reported as discrete component circuits [36] and finally became available as NIM modules [37]. Recently, integration of the quenching electronics for InGaAs/InP APDs to an application-specific integrated circuit (ASIC) has been implemented [39].

4.2.1 Operation of an active quenched SPAD.

The principle of AQC is simple: to sense the rise of the avalanche pulse and react back on the SPAD, forcing with a controlled bias voltage source the quenching and reset transitions in short and well defined times[18][22][25].

The avalanche is sensed through a low impedance, and both quenching and reset are obtained thanks to active components, such as pulse generators or fast active switches, after which these circuits are named.

A basic diagram in which the active circuitry is represented by a pulse generator is shown below, in fig. 5.5 .

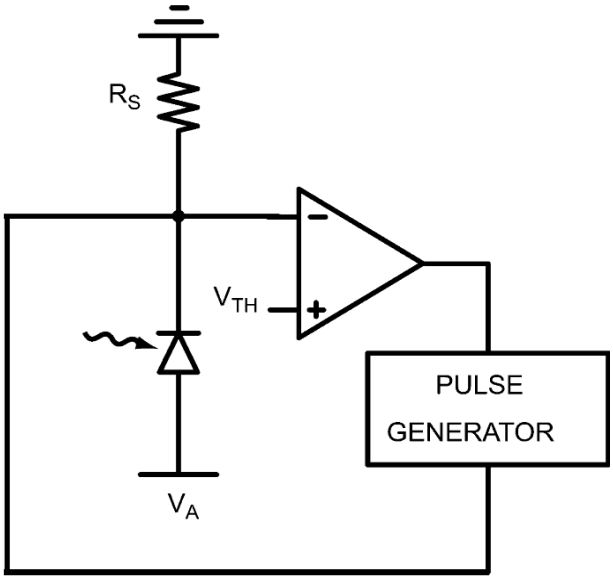


Figure 4.5 Basic diagram of an AQC[22].

In AQCs the bias voltage is lowered below breakdown when quenching an avalanche, so the quenching time is accurately controlled because it is not affected by fluctuations due to the statistical process of avalanche multiplication. After the avalanche quenching the bias is maintained below breakdown for a predetermined period (the hold-off time) and then brought back to its quiescent value.

4.2.2 Active quenching application specific integrated circuit (ASIC)

Recently, integration of the quenching electronics for InGaAs/InP SPADs to an ASIC has been implemented[54]. Results on some key parameters of APDs have demonstrated that active-quenching ASICs can efficiently improve the noise-efficiency performance and it has been shown that these APDs can work in free-running

mode[54][39]. An interesting comparison with conventional electronics, showing the improvements of ASICs can be found in [54]. These advantages, in particular low afterpulsing and noise are mostly attributed to the state-of-the-art ASIC[54].

4.3 Gated Mode operation and rapid gating.

A diode, both passively or actively quenched, can operates both in free-running mode (it can detect a photon at any time, except for the time necessary to reset the avalanche) or in Gated mode.

In gated-mode operation, the SPAD is photosensitive only during the gate-on window,ice versa, during the gate-off time interval, the SPAD is disabled and absorbed photons do not trigger the avalanche.

This operation mode is obtained by initially operating the device slightly lower than the breakdown voltage (e.g. 0.5 V lower), at a polarisation voltage V_{POL}. During the gate off time interval (T_{OFF}), the avalanche can not be triggered since the electric field is not sufficient to self-sustain the avalanche process. When a gate voltage applied, V_{GATE}, to the device switches to the high level, it raises the SPAD voltage (i.e. $V_{POL} + V_{GATE}$) above breakdown. The excess bias voltage V_{EX} is the difference between the SPAD on voltage and the breakdown voltage V_{BD}. In this conditions the SPAD can detect single-photons[62].The end of the gate defines the time of quenching.

Additionally, it is possible to introduce a hold-off time (or dead time) to ensure that the system is come back to its initial condition and it can detect photons with a low afterpulsing probability.

This operation mode is particularly suited when the signal which must be detected has a fixed frequency too, like for instance in QKD systems, because it can largely suppress dark counts, since the detector is active only during the photon arrival time[8]. It still remains the problem of afterpulsing, since the avalanches generated at the beginning of the gate generate a lot of trapped charges ; in order to solve this problem, one could try to decrease the gate width, but this is possible only to ≈ 1 ns, since under this value the capacitive response of the diode should hide the avalanche signal.

Recently, this problem has been solved by using very short gates in the so-called rapid gating[8][22][63]. In this case, gating frequencies of ≈ 1 GHz are used, with effective gate widht between 100 ps and 200 ps. Since with these ultrashort gating times, the avalanches are far from saturation, the output signal of the diode is weak ; thus, while the afterpulsing effect can be significantly suppressed, another technical challenge emerges : discriminate the small avalanches signals (few millivolts) from the background signal due to the strong capacitive response of the diode.

Two different approaches have been successfully introduced : sine-gating with filtering(SG) [63][64][65] and square-wave gating with self-differencing (SD) [66]. The first solution introduces a sine-wave gating signal and band-stop filters (BSFs) to filter out the background frequency response and its harmonics, while the second one proposes to use square waves to gate the SPAD and a differencing circuit to subtract

the output signals during two consecutive clocks, in order to acquire the weak avalanche signal[63]. In Fig. 5.6 a realization of the first solution is illustrated.

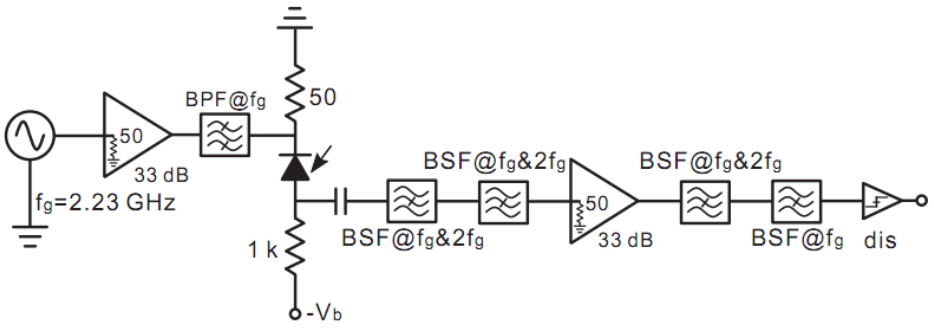


Figure 4.6. Rapid gating detector set-up using the sine-gating plus filtering technique[64]. BPF : Band pass filter, BSF : Band stop filter

5. Report of the tests on the Negative-Feedback Avalanche Diode (NFAD) made by Princeton Lightwave.

The diode under test is an InGaAs/InP single photon avalanche diode (SPAD) operating in the third telecom window (1550 nm) and at a temperature of 223.15 K.

The SPAD is biased above the breakdown voltage, thus it can be used in Geiger mode for the detection of single photons: a single photo-excited carrier can trigger an avalanche that generates a macroscopic current pulse, which will be recorded using an electronic circuitry.

After the avalanche, a negative feedback circuit using a load resistor is used to drop the bias voltage under the breakdown voltage.

The goal of these experiments is to develop a SPAD exploiting the innovative characteristics of the NFAD[20], which will be deeply explained in the next paragraph. At this point, we can say that the most important merit of NFADs is that its integrated resistor allows to quench the avalanche very quickly, thus reducing the number of trapped carriers, which should be responsible of afterpulsing effect. Therefore it is possible to use the diode in free-running mode, applying a very short deadtime.

For these reasons, firstly we have tested the diode in order to see its behavior in terms of efficiency and dark counts rate, both in gating and free-running mode (chapter 5.4), and then in terms of afterpulsing effect (chapter 5.5).

Then we have introduced an innovative technique, which use a program running on an FPGA, that is able to characterize automatically and quickly the behavior of afterpulses; this program has also a lot of interesting functions, which are useful for

understanding the effect of high order afterpulsing (see chapter 3.3 for an introduction), or in order to apply a dead-time and see its effect in reducing afterpulses (chapter 5.6).

Finally (chapter 5.7), we have tested the diode to characterize its jitter.

5.1 InP based Negative Feedback Avalanche Diodes (NFADs).

The limitations of the performances of SPADs, such as low photon counting rate and the absence of photon number resolution, can be traced back to the positive feedback inherent in the impact ionization-driven avalanche process. With the goal of radically improving upon these structures, Itzler et al. [20], have added monolithically integrated negative feedback elements (a suitable thin film resistor) in series with best-in-class SPAD structures, the resulting diode, called negative feedback diode (NFAD), is characterized by limited-in-current avalanches keeping high deterministic gain values.

An optimized implementation of this kind of device can lead to very small avalanches (e.g., ~ 10^5 - 10^6 carriers, a number which must be compared with the typical dimension of SPAD avalanches that is ~ 10^7 - 10^8) and extremely tight distribution (the excess noise factor $F(M) = \langle M^2 \rangle / \langle M \rangle^2 \rightarrow 1$ for linear mode operations, where M is the multiplication gain). The diode used in the subsequent experiment is of the same kind [20].

Having a smaller avalanche size, it is possible to reduce the carrier trapping thus obtaining a lower probability of afterpulsing; this allows re-arming the NFAD just after ≈1 μs, enabling detection rates up to ~ 1 MHz.

The introduction of negative feedback using a load resistor is conceptually equivalent to passive quenching, but it presents lower parasitic effects. In fact, with an ideal monolithic integration, the amount of charge Q required prior to avalanche quenching is simply the product of the avalanche diode depletion capacitance C_d and the excess bias V_{ex} (i.e. $Q = C_d * V_{ex}$); Q can thus be reduced through appropriate scaling of C_d (the magnitude of the excess bias V_{ex} is dictated by the desired value of PDE, because PDE increases with larger V_{ex}).

Moreover, the NFAD has a very simple design and operation mode: with just a fixed dc bias voltage corresponding to the sum of the diode avalanche breakdown V_{br} and the desired excess bias V_{ex}, the NFAD will independently execute the entire arm, avalanche, quench, and re-arm cycle and generate an output pulse every time an avalanche event is induced[20].

It is possible to realize NFADs with different amounts of feedback resistance. The device under test should be optimal for the various trade-offs that need to be managed: if the feedback is much larger, then quenching will be even faster, but recharging time constants will be that much longer; and if the feedback is smaller, recharging can be faster, but the feedback will not be adequate to quench the avalanche for higher excess bias operation [40].

Even if the NFAD technique is a young one, it shows some interesting properties, and for this reason we have decided to test a device based on NFAD, with the aim of using it in free-running mode.

5.2 Measures of Dark Counts, Efficiency and Afterpulses:experimental set-up.

The set-up used for characterizing the efficiency and the dark count behavior as a function of the bias applied to the NFAD is schematically illustrated in Fig. 6.1 on the next page.

An arbitrary function generator (Tektronix AFG3102) triggers both a digital delay pulse generator (SRS) (Dg 535, Stanford Research System, Inc.) and a 1550-nm Laser Diode (LD), at a frequency of 10 kHz (unless otherwise mentioned). It also produces gate pulses 3 V high and 150 ns long, which are superimposed over the dc bias in order to have a total applied voltage sufficient to switch the detector from gated off to gated on.

The Laser Diode produces short optical pulses with ≈200 ps full-width at half-maximum (FWHM), that then are split into two parts by a 10/90 asymmetric fiber beam splitter (BS): the first one (90% of the power) goes to a power meter (IQ 1100, EXFo Company) and it is used to regulate the variable attenuator (Var. ATT. IQ3100, Exfo Co.) in order to obtain one photon per pulse (unless otherwise mentioned). In this attenuator there is also a shutter that is used to switch on and off the illumination of the detector.

The attenuated light is guided to the NFADIn the same chip of the diode other elements are also fixed: first, there is a 50-Ω resistor used to obtain an optimal coupling with the resistor that is “seen” in the output; second, there is a 1 kΩ resistor that serves in case of undesired high-values of bias in order to limit the current flowing

into the APD; then, there is a ferrite bead, a passive electric component used to suppress high frequency noise in order to have a stable bias voltage; finally, there are two capacitor of 10 nF and 100 nF, used to accumulate carriers and as a bypass for high-frequency currents. A photograph of the box containing the NFAD is presented after the experimental set-up.

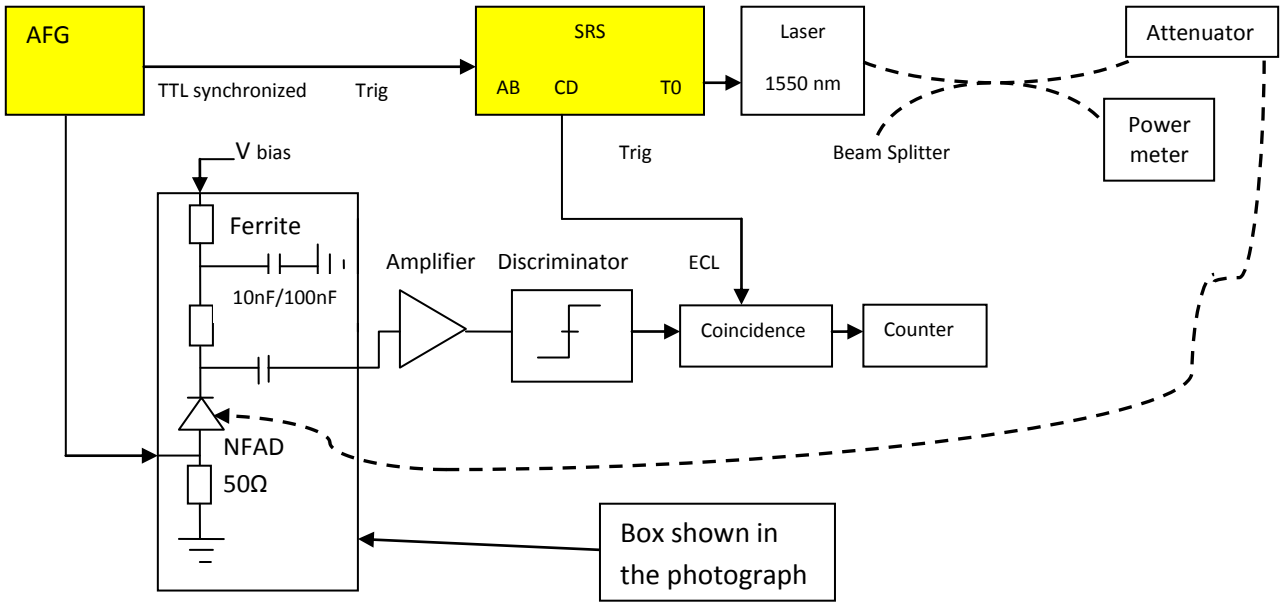


Figure 5.1. The experimental set-up used for measures of dark counts, efficiency and afterpulses (with double-gated method)

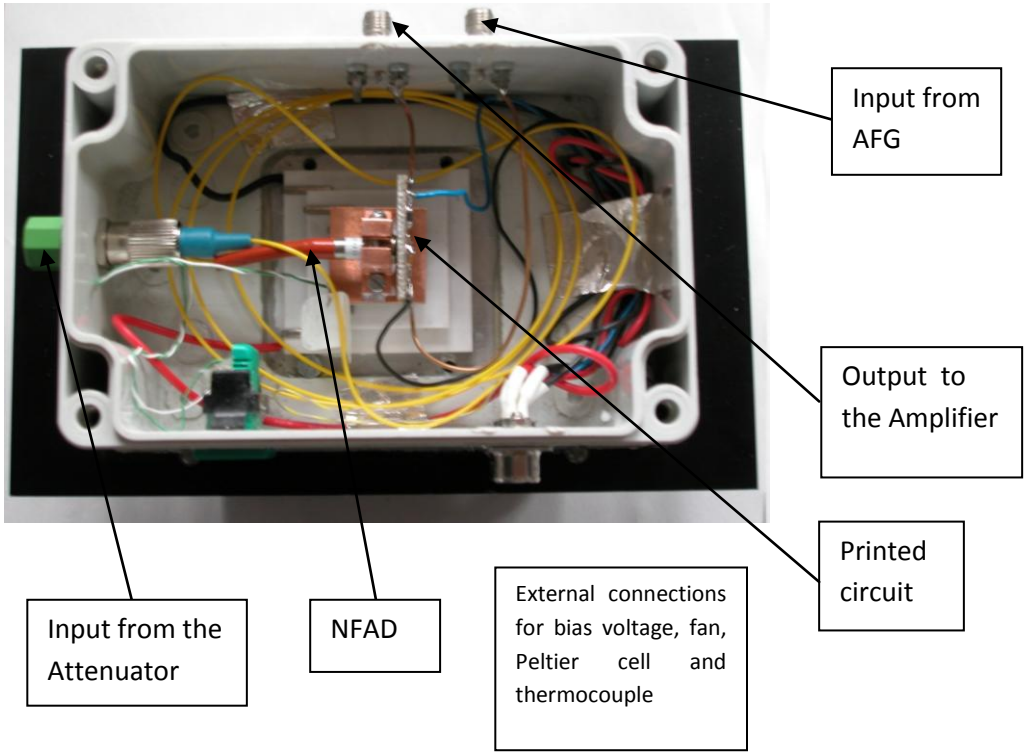


Figure 5.2 Photograph of the box represented in Fig. 6.3.1, with descriptions of the external connections; the NFAD is fixed over the Peltier cell.

The diode is dc biased with a precise and stable voltage generator (Keithley 2400) and its output is ac-coupled via a 1 nF capacitor with an amplifier (ZHL-42W, MiniCircuits).

The output of the amplifier is directed into a discriminator, whose logical signals are sent into a coincidence system together with a gate produced by the SRS. The delays of the AFG and of the SRS are set so that the photons arrive on the APD when the voltage gate is present; the coincidence system allows rejecting the transients signals generated by the discriminator due to the rising and falling edges of the gate pulse.

Finally the NIM output of the coincidence system is connected with a counter.

When the diode is operated in the so-called free-running mode, the set-up is simpler: the AFG is not necessary, and the SRS triggers directly the LD; at the end, the NIM output of the discriminator is sent directly to the counter, without any coincidence system.

5.3 Detection efficiency and dark counts measures.

For a given bias voltage the detection efficiency and the dark counts are measured. The experiment is repeated many times, changing the bias, for two cases: first in free-running mode, second in the gated mode; in the last case the width of the voltage gate is 100 ns.

The integration time is chosen in order to have a statistical uncertainty on the counts less than 5%, except for the lowest bias voltage, when it is anyway less than 10%.

In the free-running mode the dark count rate is simply the ratio of the number of counts over the integration time. The effective efficiency is then:

$$\eta = \frac{S-N}{f*\mu}$$

where S is the detection rate with the laser, N is the detection frequency without the laser, due to the noise, μ indicates the mean number of photons per laser pulse and f is the repetition frequency of the laser.

In the gated mode, instead, the dark count probability per ns is:

$$\mathbf{P_{dc}} = \frac{\boldsymbol{Cdc}}{\boldsymbol{f} * \boldsymbol{\tau ab}}$$

Where C_{dc} is the observed dark count rate, τ_{ab} is the effective width of detection gate in nanosecond and f is the repetition frequency of laser pulses. Then P_{dc} is multiplied per 10^9, thus obtaining the normalized dark count rate per second; this value is now comparable with the dark count rate in the free-running mode.

The detection efficiency η is calculated as

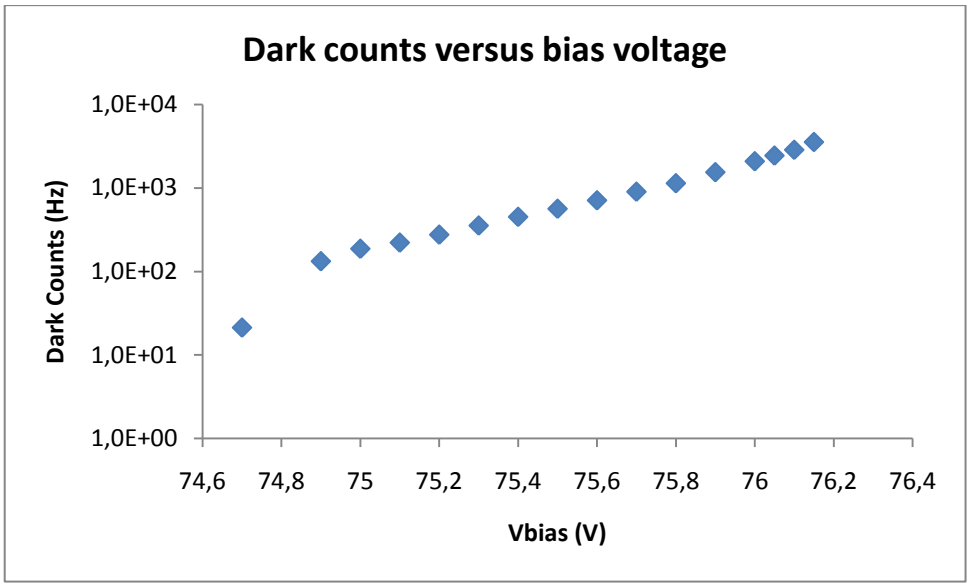
$$\boldsymbol{\eta} = \frac{\mathbf{1}}{\boldsymbol{\mu}} \mathbf{ln} \frac{(\mathbf{1} - \boldsymbol{Cdc/f})}{(\mathbf{1} - \boldsymbol{Cde/f})}$$

where μ is the mean photon number per laser pulse and C_{de} is the observed detection rate per laser pulse.

In the free-running mode the data obtained from the experiment are reported in the appendix (table1).

From these data two figures are extracted: the first one shows the evolution of dark counts versus the bias voltage applied on the detector and the second one shows the evolution of effective efficiency, again as a function of the bias voltage.

(a)



(b)

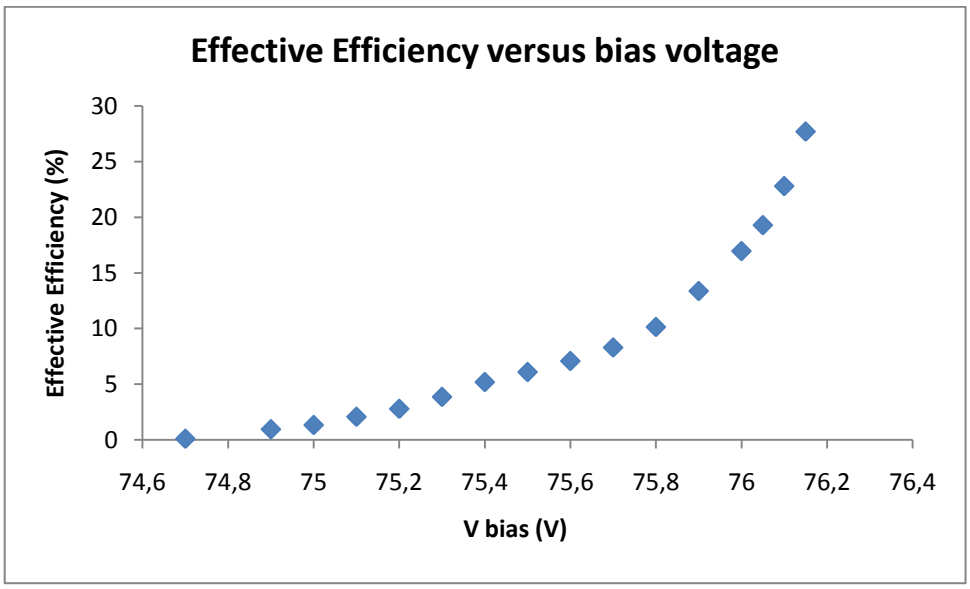
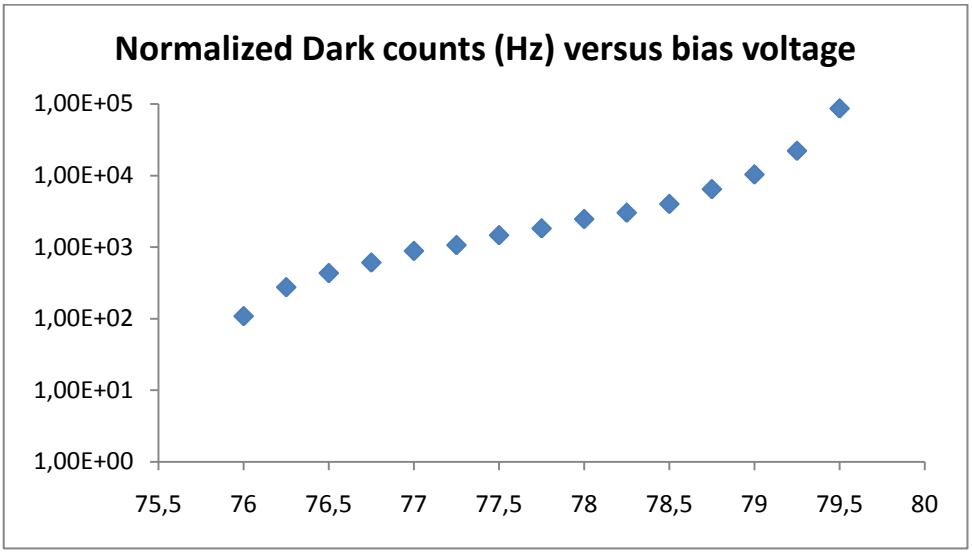


Figure5.3 Free-running mode. (a) dark counts and (b) effective efficiency as a function of the bias voltage applied on the NFAD.

Then, the same experiment is repeated for gated-mode, and a similar table reports the data obtained (Appendix, table 2).

Also in gated-mode, two figures are extracted from these data: the first one shows the evolution of the normalized dark counts per second, while the second one shows the evolution of the efficiency again versus dc bias.

(a)



(b)

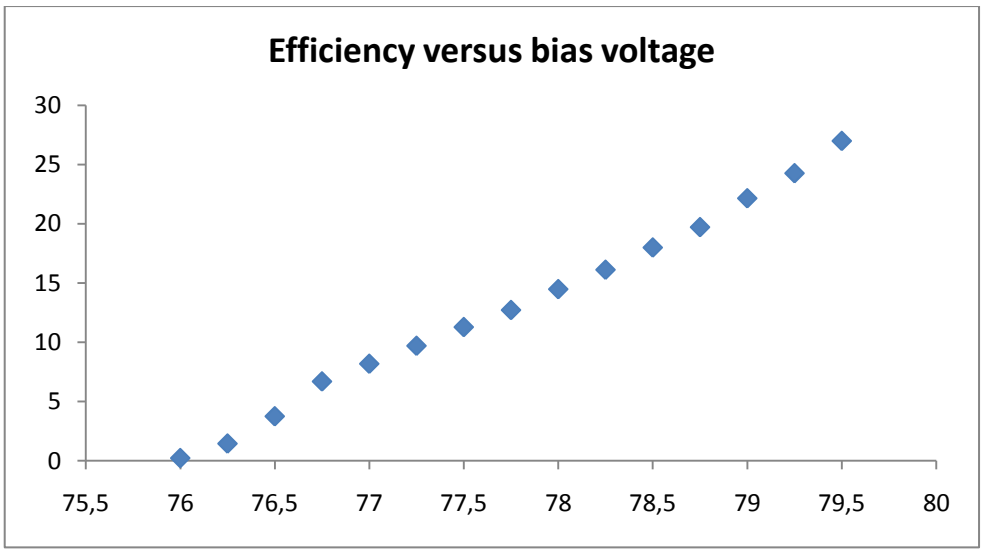


Figure5.4 Gated mode: (a) Normalized dark counts (per second) and (b) efficiency as a function of the bias voltage applied on the NFAD.

Comparing the four graphs previously shown, we can note that, while the evolution of dark counts is exponential in both regimes, the evolution of the efficiency, in gated-mode, and the evolution of the effective efficiency, in free-running mode, are different: in fact, while in gated-running mode the tendency is almost linear (the slight oscillation is probably due to small variations of temperature), in free-running mode the tendency is almost exponential.

Moreover it must be underlined that, in order to reach a certain efficiency (for example 25%), the applied bias is different in the two cases: in the gated mode this efficiency is reached at ~79,4V, while in the free-running mode it is enough to apply a much lower bias of ~76.1 V. Also this difference among the two cases is due to afterpulses, since in the free-running mode (without an external applied hold-off time) many more of the measured counts are originated from afterpulses, compared with the gated measurements [40].

These different behaviors rebound on the evolution of dark counts rate versus efficiency, which are shown in the next figures.

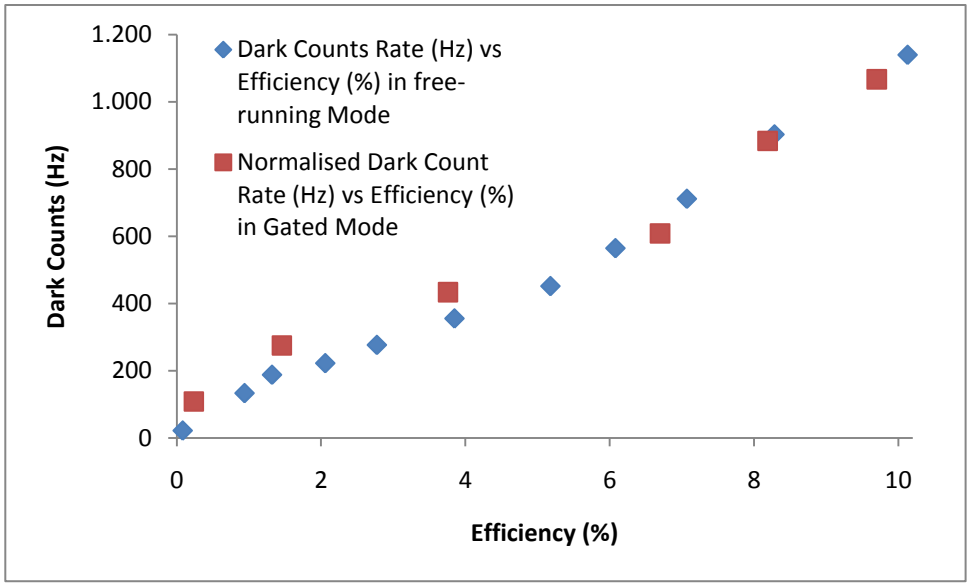


Figure 5.5. Behavior of the noise as a function of the efficiency. For the case of the Free-running Mode the noise is the "raw" Dark Count rate, while for the Gated Mode it is the Normalized Dark Counts Rate (considering only the time during which the diode is active).

For both the regimes the graphs of D.C. rate versus efficiency are reported. We can note that, for an efficiency lower than 10%, the behaviors in the two cases are similar, even if the evolution is linear in free-running mode and slightly exponential in gated mode. We report approximately 10^3 Dark Counts per second (10^{-6} ns^{-1}) when $\eta \approx$ 10% and $5*10^2$ dark counts per second when $\eta \approx 5\%$..

After that region the two behaviors diverge: in the free-running mode the evolution continues to be linear even for high values of η; in the gated mode, instead, the dark counts rate increases more than linearly.

In conclusion we can underline that the different behavior of the Dark Count Rate as a function of the Efficiency in free-running mode and in Gated-mode, is a consequence of the different behavior of the Efficiency, which rises exponentially with

the applied bias voltage in the Free-running mode and linearly in the Gated-mode. This difference is due to afterpulsing effect, which affects more the diode in free-running mode then in the gated mode.

5.4 Afterpulses measures with the double-gated method.

The last section is dedicated to afterpulses. The measures are taken using two different set-ups: the double-gated method and a process governed by an FPGA.

While the double-gated method is a well-known technique, the second technique is a new one and it has some interesting features.

First of all it is simpler and faster compared with the double-gated method, because when the parameters in the software are fixed it is not necessary to scan the time after a detection in order to measure the afterpulses, but it is sufficient only to fix the resolution and the temporal extension of the measure at the desired values.

With the help of the FPGA it is also possible to discriminate the high order afterpulses from the first order afterpulses (see chapter 3.3), thus permitting a more accurate analysis of the afterpulsing behavior.

As we have previously explained, when an avalanche is triggered and a current flows through the diode, some of the charge carriers could be trapped by trapping centers in the multiplication layer and subsequently released at random times. Afterwards these carriers could trigger other avalanches (afterpulses) which introduce correletions betweens consecutives pulses, thus limiting the maximum repetition rate.

First of all, the double-gated method is used: two coincidence windows are sent into the diode, with delays suitably regulated in order to detect first detection and dark counts (in the first gate), and then afterpulses due to them and dark counts (in the second gate).

This method is also known as time-correlated carrier counting method [43] and it is a type of "pump-probe" measurement.

The values fixed for the parameters of the experiment are: η≈10% (which corresponds to approximately a V_{ex} of 3 V), repetition frequency = 10 kHz, leading and trailing times of the two gates = 5 ns, 1 photon per pulse and gate width of 100 ns.

For every value of the delay of second gate, the probability of afterpulses per ns is calculated simply as

$$\mathbf{P_{ap}} = \frac{Cap}{Cde * \tau cd}$$

Where C_{ap} is the observed afterpulse rate (obtained subtracting the dark counts rate from the total counts rate in the second gate), C_{de} is the observed detection rate (obtaining subtracting the dark counts rate from the total counts rate in the first gate) and τ_{cd} is the width of the second coincidence gate. The results are shown in figure 5.7.

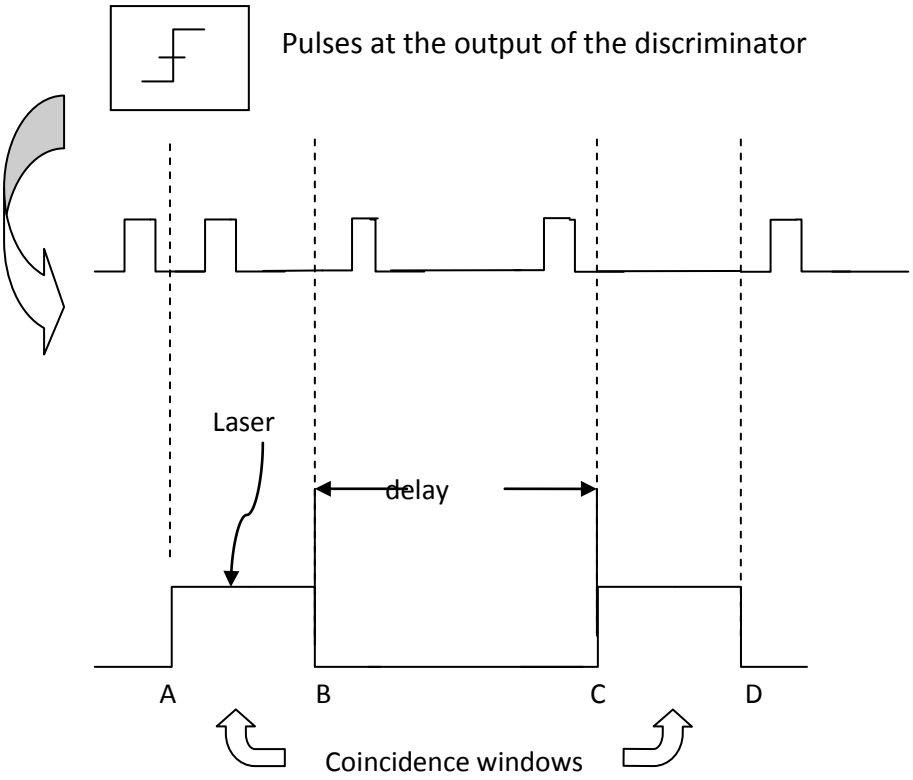


Figure 5.6. The double-gated method. The SPAD is biased under its breakdown voltage, thus being normally in a non active state; an arbitrary function generator produces square biasing gates with a full width at half maximum of 150 ns, an amplitude of 3 V and a repetition frequency of 10 kHz. When a biasing gate is applied, the diode is biased over the breakdown voltage, so an avalanche could be triggered, filling the traps and thus generating subsequent afterpulses; these afterpulses are measured during a second bias gate situated at varying delays after the first one. Coincidence windows (shorter than biasing windows) are necessary in order to discriminate, at the output of the discriminator, the pulses due to detection and the derivative pulses originated by the square biasing gates as a consequence of the ac coupling.

The afterpulsing data obtained with the double-gated method are gathered in the tables shown in the appendix (table 3). The integration time was 30 s for dark counts and 10 s for detection counts (with the laser). The dark count rates were obtained simply dividing the total dark counts by the integration time and the same for the detection counts (with the laser). The afterpulsing effect is evaluated in terms of probability per ns, thus obtaining a normalized result for different gate widths.

The probability of afterpulsing per ns is plotted versus the delay of the second gate in Fig 6.7.

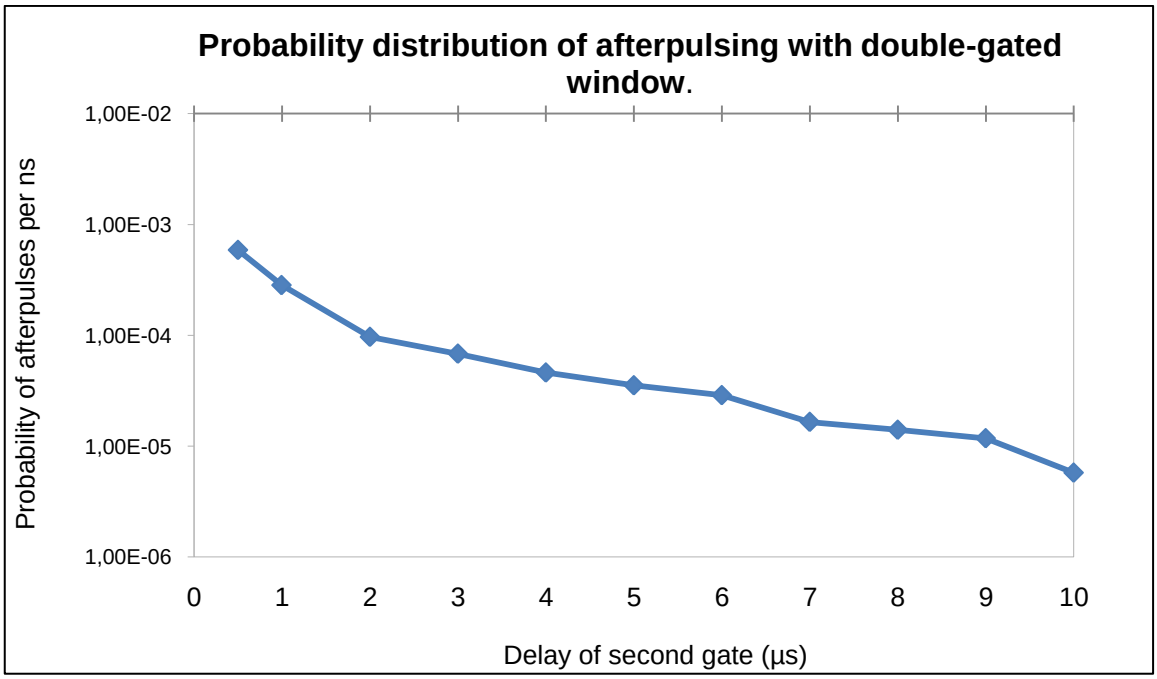
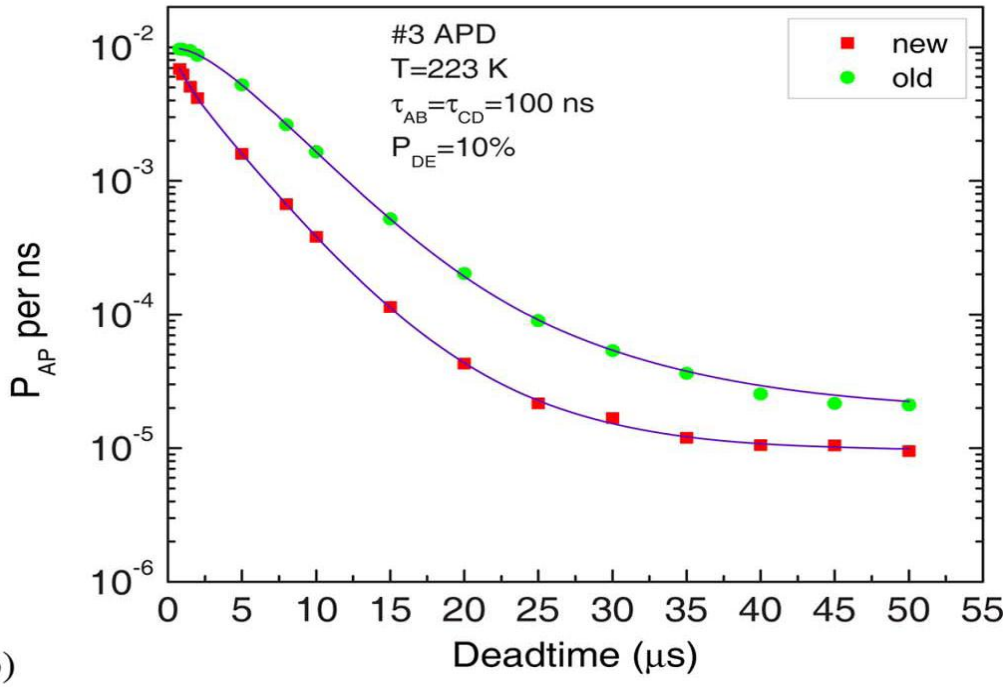


Figure 5.7. Afterpulse probability per ns versus delay of the second gate

Then, in figure 6.8, there are the corresponding results previously obtained using an InGaAs-InP avalanche photodiode at 1550 nm with an active quenching ASIC (application-specific integrated circuit).

It is possible to say that the results obtained with the negative-feedback diode are better than those obtained with the active quenching ASIC diode[54]; in fact, we report a P_{ap} of less than 10^{-5} after 10 μs, that is about 40 times less. In the Fig. 6.8



b)

Figure 5.8 Afterpulse probability per ns versus deadtime. Data from : J.Zhang, R. Thew,J.-D. Gautier, N. Gisin, and H. Zbinden, " Comprehensive Characterization of InGaAs-InP Avalanche Photodiodes at 1550 nm With an Active Quenching ASIC", IEEE J. of Quantum Electron., vol. 45, NO. 7, July 2009 [54].

5.5 Afterpulsing measures with an FPGA.

5.5.1 Experimental set-up and behavior of the FPGA.

Finally, an FPGA (Spartan 3A)is programmed in order to measure the afterpulsing distribution as a function of time. The aim of the procedure is to develop an innovative program running on an FPGA, which will be able to characterize automatically and quickly the afterpulsing behavior; with this new method, external devices like Arbitrary Functions Generators used to change the delay are no longer needed. Moreover, the software which will be presented in this chapter has many other options that allow to study deeply the afterpulsing behavior: it is possible to analyze the importance of high order afterpulsing and characterize the curve of afterpulses starting from a known initial condition.

The FPGA is triggered by a pulse generator (HP 8133A pulse generator) at a frequency of 50 MHz; both of them can generate signals that, after being passed through an AND-gate, trigger the laser. The result of this set-up (which is shown in figure 7) is that laser pulses are sent into the diode following the commands of the FPGA, which is driven by a software, whose settings are chosen by the user.

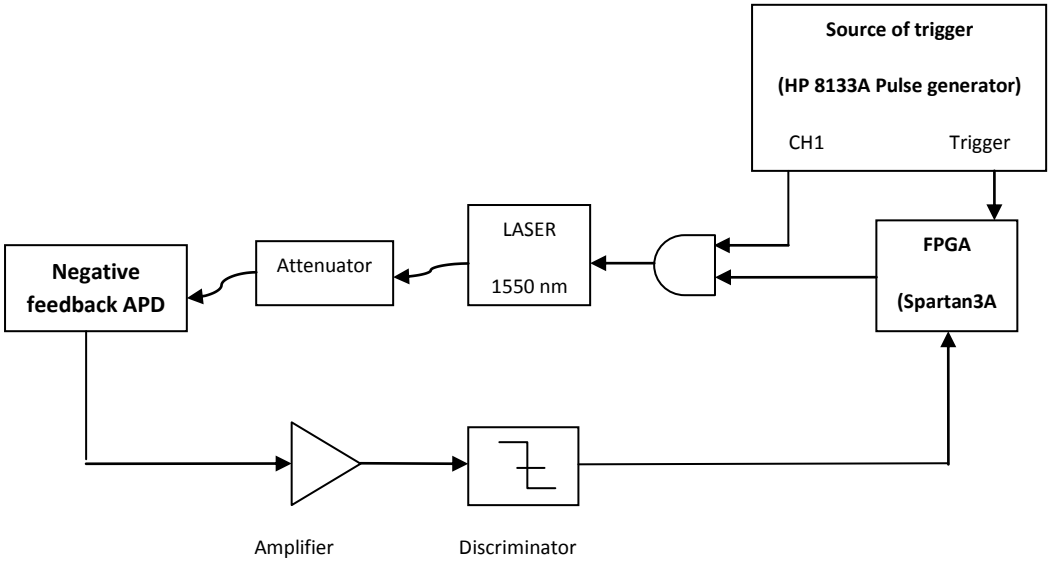


Figure 5.9. Set-up used for the measures of afterpulses using an FPGA.

It is important to note that the selection of the triggering frequency involves also the choice of the width of the time-bin (the temporal interval used to discretize the data in a histogram) of the histogram, thus 50 MHz correspond to a time-bin of 20 ns. If a lower resolution was acceptable for an experiment it should be thus possible to change the triggering frequency, thus obtaining also a longer histogram.

The logical behavior of the FPGA is represented in figure 8. At the beginning, it waits until there are no detections (due to dark counts and to their afterpulses) for a fixed time (we used 10 μs for these measures, which correspond to 500 waiting cycles)(the time-bins during which the FPGA waits for no detections), in order to put the diode every time in a well-known regime, in which the effects of its previous behavior are negligible. It is important to remember that, between two different laser pulses for which afterpulses are registered, there is a delay which is at least the sum of

the waiting-for-no-detection time and the length of the histogram; so, choosing for example a waiting-for-no-detection time of 10 μs and an histogram of 50 μs, the total delay should be 60 μs, thus being long enough to avoid the influence of afterpulses due to the preceding laser pulse.

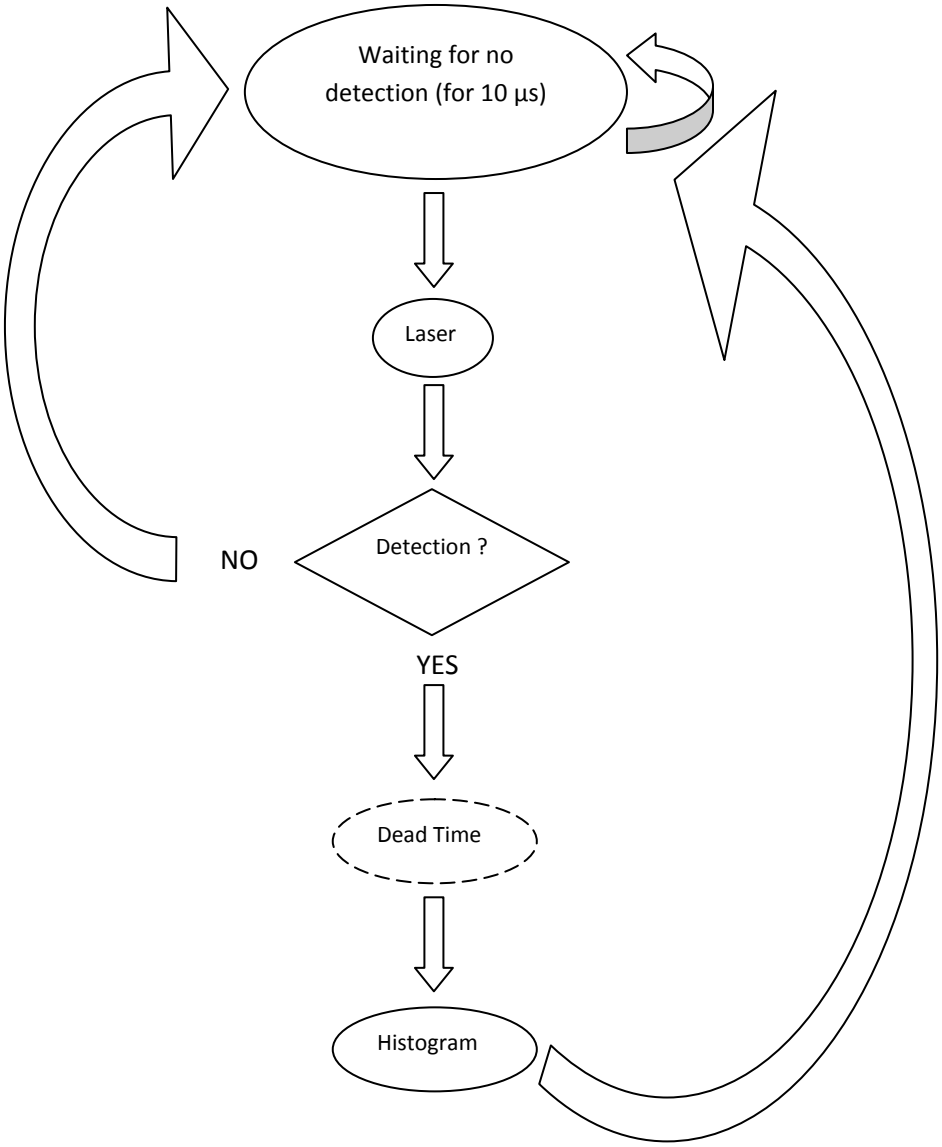


Figure 5.10 Logical behavior of the FPGA. During this experiment the dead time was not already implemented, thus every time there was a detection the FPGA went directly into the histogram mode.

Then the FPGA sent a pulse to the AND-gate, where it arrives together with the synchronized pulse provided by the pulse generator, triggering the laser. Once a laser pulse is suitably attenuated (in order to have the desired average number of photons per pulse) and sent into the diode, it could trigger or not an avalanche: therefore, after an appropriate delay and for one time-bin, the FPGA controls if there is a detection of the avalanche generated by the laser pulse. If there is a detection, the FPGA starts to waiting for one or more detections (due to afterpulses or dark counts), registering them into a 250-bin-histogram. Instead, if there are no detections into the appropriate time-bin, the FPGA returns at the first step of its cyclic behavior.

The result of this measure is a histogram, which shows the evolution of the detections, due to dark counts or afterpulses, after a laser-detection.

Aside from the histogram showing the afterpulsing evolution with time, the software provides also other measures, such as:

-the number of laser pulses sent to the detector

-the number of laser pulses which are detected into the appropriate time-bin and the related laser detection efficiency

-the total number of all detections occurred during a measure and the related total detection efficiency (obtained dividing this number by the number of laser pulses sent to the detector)

-the afterpulsing probability (during the histogram time) that is the number of detection occurred into the histogram divided by the number of laser pulses which were previously detected. This measure is not corrected for dark counts, because it is

not possible to discriminate them from afterpulses, but it is possible to subtract dark counts through a fit as it is shown later.

-parameters of control, such as: the number of detections during the waiting-for-no-detection time, the number of time bins which compose this waiting time and the number of detection between the sending of the laser pulse and the time bin during which it is expected to detect it (if there is a detection during this period, the software considers it as an error; anyway it waits until the time bin during which it expects to detect the laser pulse and considers the possible detection in the calculation of laser detection efficiency, but later it returns at the first step of the cycle).

The software allows the user also to fix some parameters before every experiment, such as:

- The number of waiting cycles, from 1 (corresponding to 20 ns) to 9999 (corresponding to 199.98 μs). Since a slightly decrease of laser detection efficiency was observed for less than 100 waiting cycles (2 μs), it is better to choose a number larger than this. This problem was thoroughly investigated, but it is not already well-understood; however, the behavior of laser detection efficiency versus the number of waiting cycles is stable after 100, allowing us to consider this as the correct value.

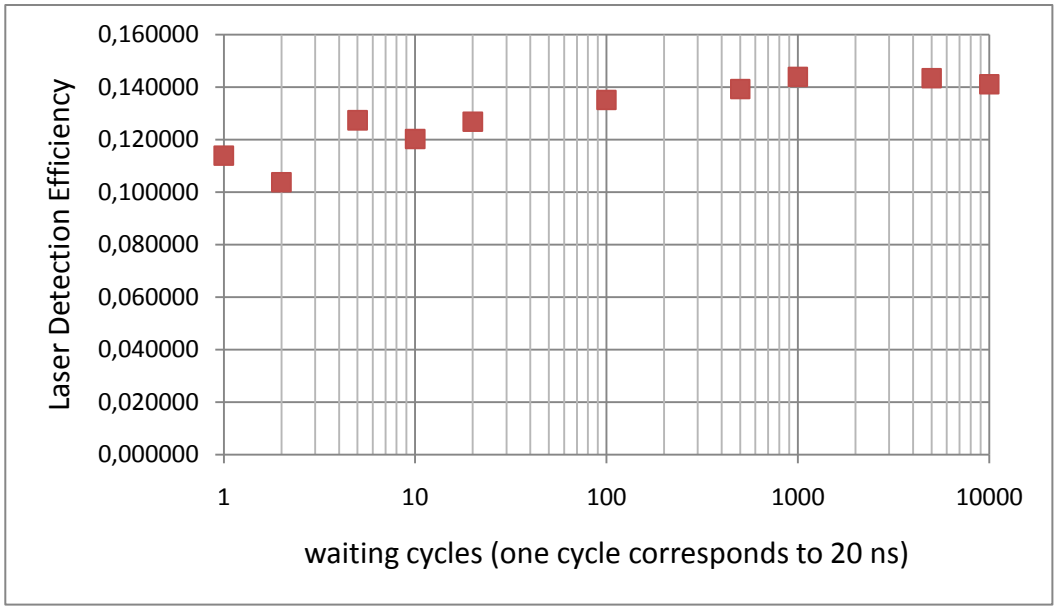


Figure 5.11. Laser Detection Efficiency versus the number of waiting cycles. The measures were registered at -50 C, with a bias voltage corresponding to ~14% of efficiency at 500 W.C.

- Number of cycles between laser sending and detection, due to the time length of the path travelled firstly by the laser pulse and then by the avalanche current pulse generated

- Dead time; it is provided by an external circuit, which generates a gate of 3 V of amplitude and a minimum length of the dead time of 20 ns

- Width of the histogram. Since the FPGA program allows only registering a 250-time-bin histogram, this is possible to be changed only by growing the width of a time bin. This can be done by reducing the resolution of the histogram. For example, some measures were taken by including 10 cycles of 20 ns each for every time-bin, thus obtaining 250 time bins of 200 ns and a histogram width of 50 µs .

5.5.2 Measures of afterpulses with the FPGA set-up and their statistical analysis.

In Fig. 5.12 it is shown the evolution of the probability of afterpulses per ns versus time, at a repetition frequency of laser pulses of about 97 kHz, at a temperature of 223.15 K, at an efficiency of approximately 10% and with 0.1 photon per pulse.

After the graph, there are also the parameters obtained by fitting the results with a curve of a double exponential decay:

$$y = y_0 + A_1 * \exp\{-t/\tau_1\} + A_2 * \exp\{-t/\tau_2\}$$

The result of this fit with leads to lifetimes of the traps which are around 512 ns for τ_1 and 3132 ns for τ_2, with associated probabilities which are around $3*10^{-4}$ for A_1 and $1.2*10^{-4}$ for A_2.

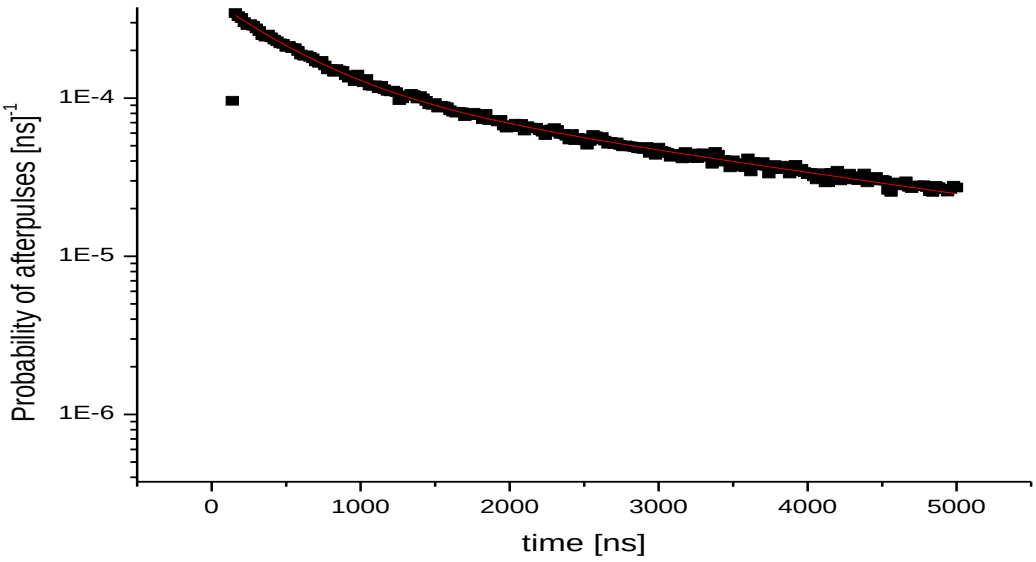


Figure 5.12 Probability of afterpulses per ns versus time at 223 K, $f_r \approx 97$ kHz, $\eta \approx 10\%$ and 0.1 photons per Pulse. The results of the fit are: $y_0 = 10^{-6}$ (this parameter was fixed before the fit, and was previously obtained with a linear fit of the curve starting from 100 μs after the detection of the laser pulse), $A_1 = 3*10^{-4}$, $\tau_1 = 512.9$ (±9.8), $A_2 = 1.2*10^{-4}$ ($\pm 3.2*10^{-6}$), $\tau_2 = 3132$ (± 78.1), $\chi^2 = 8.4*10^{-12}$, $R^2 = 0.999$

Fig. 5.13 shows the total probability of afterpulsing at a detection efficiency of 10%. We note that for long delays the measured values tend to the dark count probability, which is about $6 \cdot 10^{-6}$/ns for the tested device. As a comparison, a curve recorded with a "standard" APD and active quenching is added [54]. We see that the afterpulsing probabilities are reduced by almost two orders of magnitude.

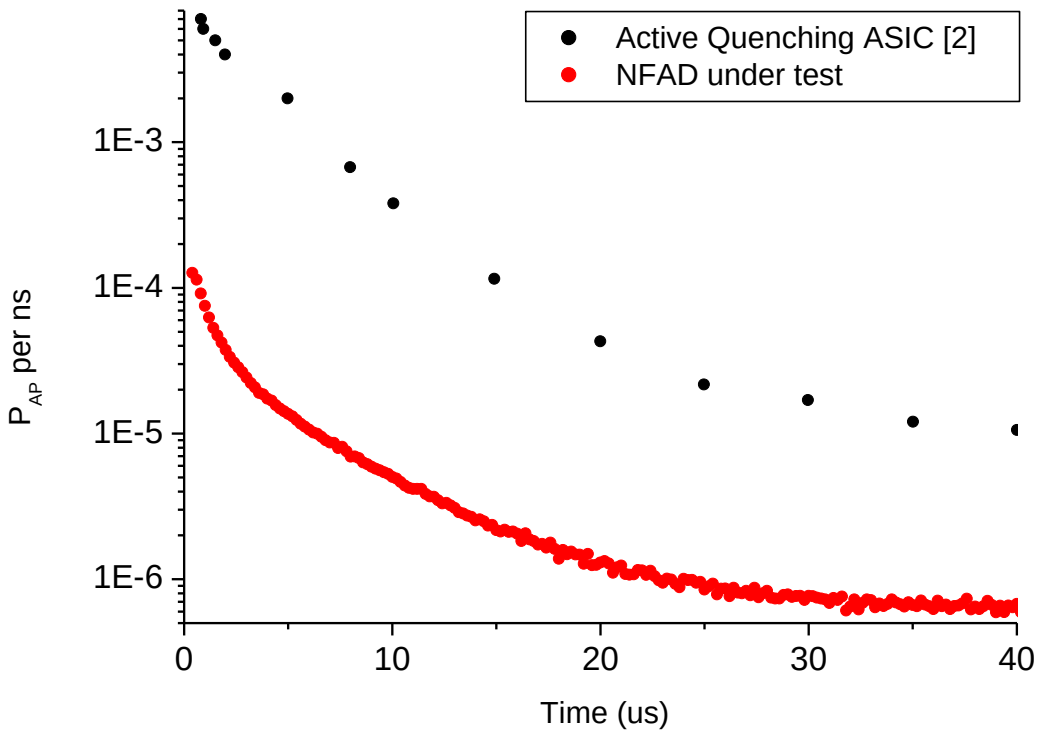


Fig. 5.13. Afterpulse probabilities as a function of time after an avalanche.

5.6 Hold-off time measures.

Besides reducing the avalanche charge (see chapter4), another way to reduce afterpulsing intensity is to keep the diode biased below breakdown for a period of time after avalanche quenching. This period of time is called hold-off time (or often dead

time) and during this period the trapped charges could be released without triggering avalanches, thus emptying the trapping centers without adding any correlations between consecutive avalanches [22].

When the hold-off time is concluded, the diode is biased again over the breakdown voltage and it can emits again current pulses due to laser detections, dark counts or afterpulses with a trapping time higher than the hold-off time.

Unfortunately, the addition of an hold-off time reduces the maximum repetition rate to a frequency $f_{max} = 1/\tau_{hold\text{-}off}$, so a trade-off must be reached; for example, with a hold-off time of 1 μs, the maximum repetition frequency is 1 MHz.

We test the diode using different hold-off configurations, in order to choose the best one; the goal, in fact, is to obtain a well-functioning and optimized diode, which will be used in other experimental set-ups, in particular for testing QKD systems.

5.6.1 Set-up and circuitry for hold-off time.

The NFAD diode used for the experiments here reported is used also for the hold-off measures, even if the output capacitive coupling was changed for reasons which will be clear later. The other elements contained in the box of the diode are the same used for the experiments presented in the previous chapters.

In the figure 5.13 the set-up used for the hold-off measures is shown.

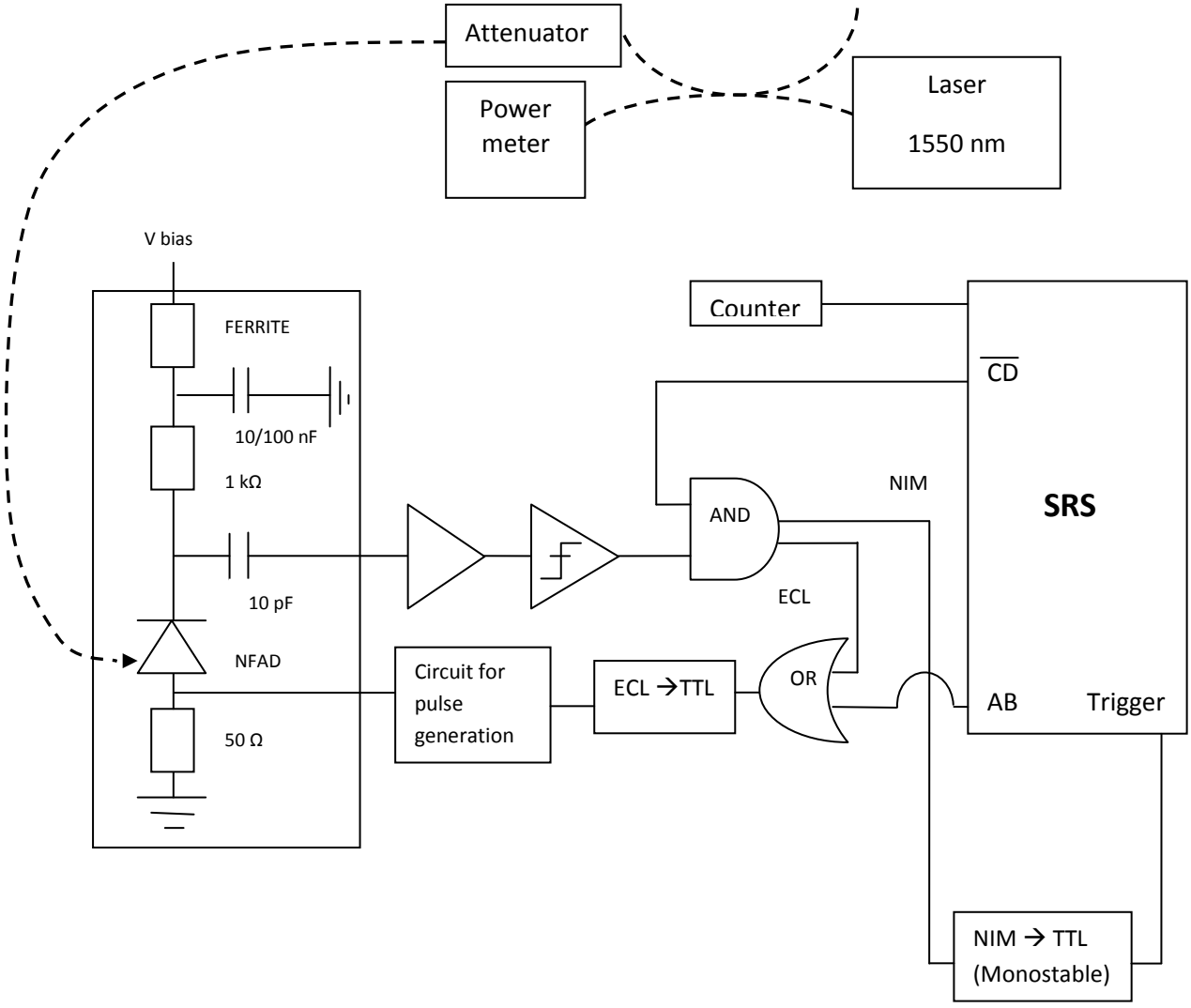


Figure 5.14. Set-up used for testing the NFAD with an external-applied hold-off time.

A 1550-nm Laser Diode (LD) is operated in continuous wave (CW) mode; it produces a stream of photons which are separated by random intervals. Subsequently the optical

path of these photons is the same explained in the paragraph 5.3 for the previous set-up

The output of the attenuator is directed into a chip containing the NFAD described in Sec. 6.3. Also the other elements of the chip are unchanged, except for the ac-coupling of the output, which is now reduced to 10 pF, for reasons which will be soon explained. The diode is biased using a voltage generator (Agilent E3612A).

The output of the chip is then connected with an 30dB-amplifier (Wenteq microwave) and the output of the amplifier is directed into a discriminator.

This discriminator is then connected to an AND-gate; the aim of this gate is to filter the spurious signal generated by the AC coupling at the output of the diode, when the gate used to obtain the hold-off time is applied. In fact, the second input of the AND gate is synchronized with the arriving time of this derivative spurious signal and it is set to zero only for an interval of 5 ns centered around this time.

After being passed through the AND-gate, the ECL signal generated by the avalanche goes to an OR-gate and then into an ECL -> TTL converter with an internal monostable; this device generates a TTL signal having an amplitude of only 2 V, which is not sufficient if we want to reduce the bias under the breakdown voltage, when the applied excess bias is more than 2 V.

An external analogic circuit was used in order to solve this problem: it receives the TTL signal with 2 V of amplitude as input and then returns a TTL signal of ≈ 4 V of amplitude as output. This is the signal which is finally used to apply the desired hold-off time to the diode.

A feature of this circuit is that, if we want as output a “regular” rectangular signal of a certain voltage (for example the 4 V desired), without any significant ripple

at the beginning or at the end of it, we must generate a pick-to-pick signal of ≈ 4 V with a ≈ 0 V off-set. It means that, when there is not a hold-off signal applied, the diode is biased with a negative voltage of ≈ 2 V; thus in the subsequent tables showing the results obtained, the bias voltage is the total bias voltage applied over the NFAD both with the voltage source and with the hold-off time circuit.

The sum of all the delays, from the onset of the avalanche to the onset of the signal used in order to have an hold-off time, is about 20 ns; this time is sufficiently small to avoid any other detection after the avalanche, before the start of the hold-off time.

To synchronize the SRS with the arrival of the photon I employed the output signal of the AND gate to trigger the SRS.

Once the SRS is triggered, it can provide two signals: the AB gate, which is an ECL signal that goes directly into the OR gate thus defining the hold-off time, and the CD gate, which is a reversed ECL signal (it is always active at one-level except between the time C and D, when it is maintained at the zero-level in order to filter the noise due to the derivative signal).

A photograph of the experimental set-up is shown in the next figure.

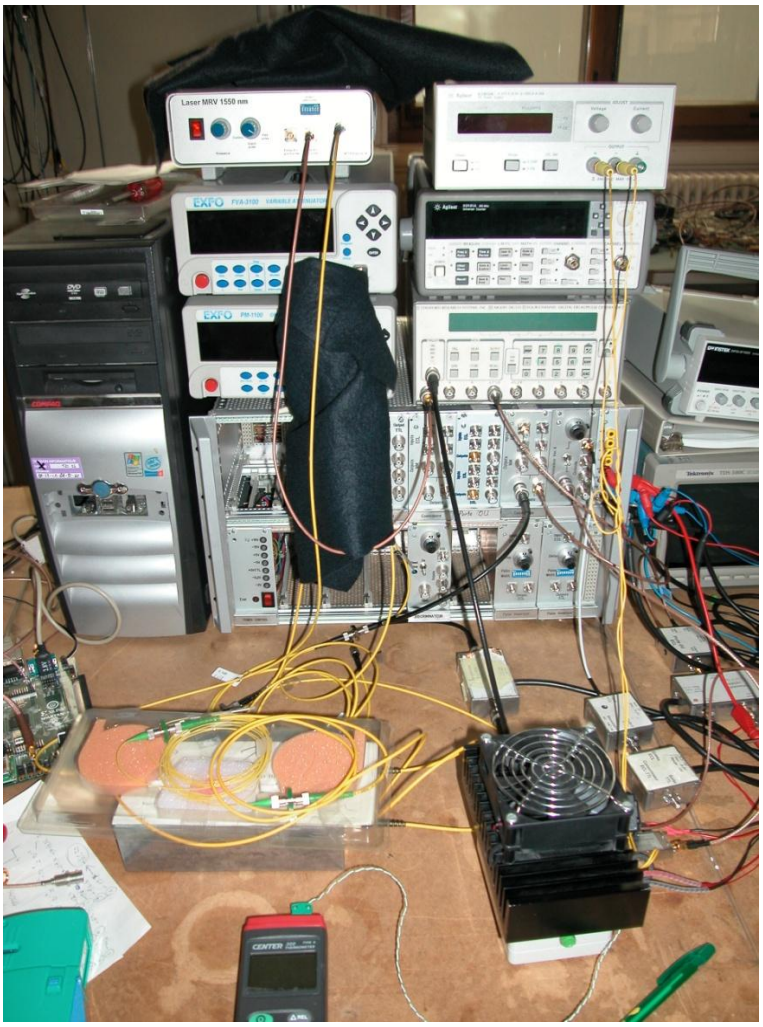


Figure 5.15 Photograph of the experimental set-up used for the hold-off time characterization.

5.6.2 Quantum (or corrected) and effective efficiencies.

For every experiment explained in this section, the temperature is set to 223 K, and for every value of the hold-off time the bias voltage on the NFAD is varied and the counts corresponding to all detections are registered. To register the dark counts, an optical shutter is used to remove the laser signal.

In this free-running mode scheme with an hold-off time applied to the detectors after every avalanche, two different efficiencies are determined: the quantum efficiency and the effective efficiency.

The quantum efficiency (also named corrected efficiency) is calculated in the following way:

$$\eta_Q = \frac{\left[\frac{S}{(1-S\tau_h)} - \frac{N}{(1-N\tau_h)}\right]}{n}$$

Where the efficiency is corrected for the hold-off time, τ_h, applied to the NFAD. Then, S represents the signal detected when n photons (per second) are sent to the detector and N denotes the noise due to the dark counts. The afterpulses of noise and signal detections are thus included both in S and N.

This efficiency gives the probability of detecting a photon that is incident on the detector while it is active[50]. However, it is more interesting to know the effective efficiency of the detector, where no constraints are placed on the detector being ready

for a photon that is there, as should be the case for a free-running detector. To calculate this, no corrections for hold-off time are imposed and it is defined as

$$\eta_{eff} = \frac{(S-N)}{n}$$

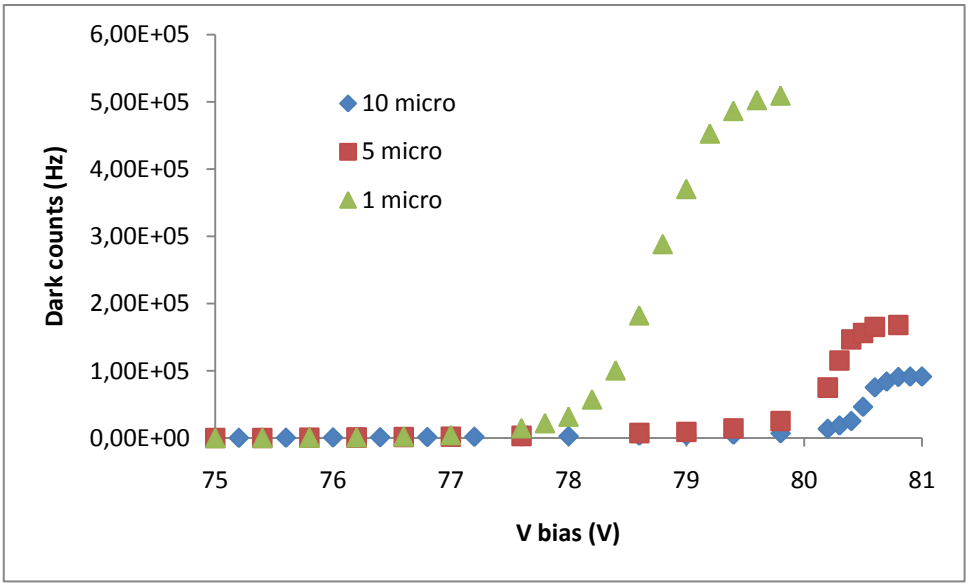
The effective efficiency saturates at lower applied bias compared with the quantum efficiency.

5.6.3 Results and discussion.

First the diode is tested with an average number of 10^4 photon per second, changing the applied hold-off time, from a minimum of 1 μs to a maximum of 10 μs. The data thus obtained with using different hold-off times (10, 5 and 1 μs) are shown in the table4 in the appendix.

In the Fig.5.15 the evolution of the dark counts versus the applied bias voltage is shown. A zoom of the first part of the graph is later depicted.

(a)



(b)

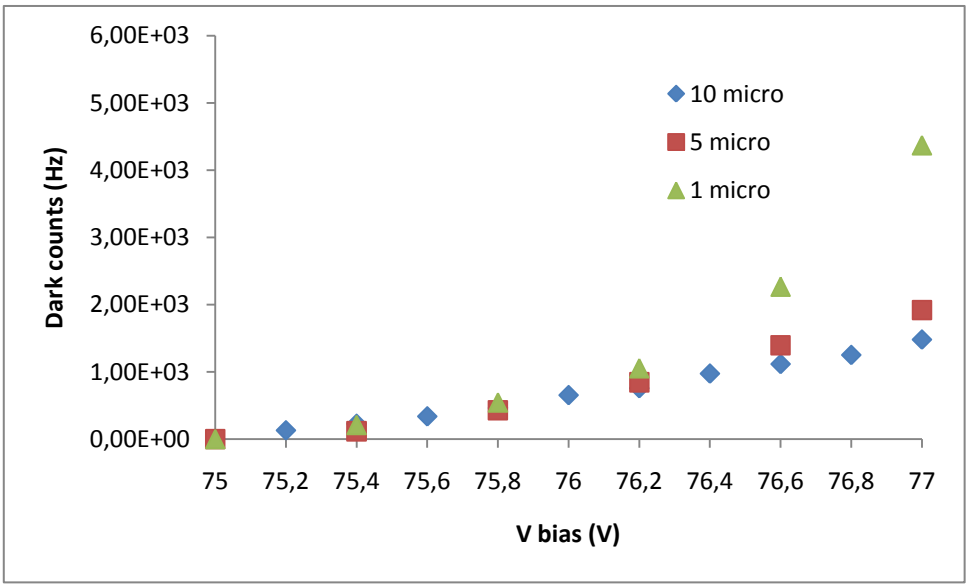


Figure 5.16. (a)Noise characteristic of the NFAD tested in free-running mode with 10^4 photons per second for three different values of the hold-off time. (b)Zoom of the first part of graph (a).

Looking at the data up to high excess bias values we can note that the dark counts reach a saturation value which is determined by the fact that, at these biasing values, after the end of each hold-off time another photon is immediately detected;

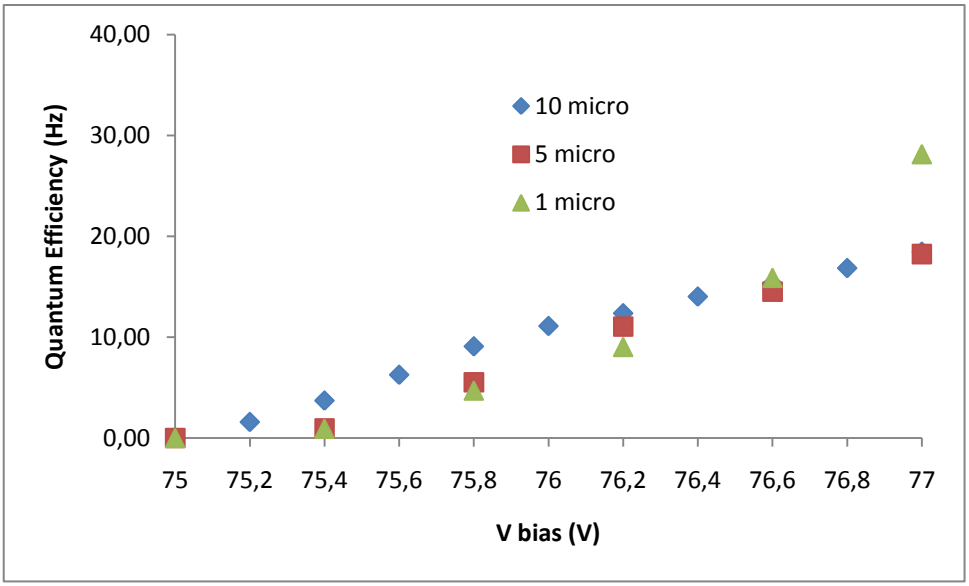
thus the maximum dark counts frequency which can be detected should be the reverse of the hold-off time, but the experimental results are different. In fact, the saturation frequencies measured are respectively 90 kHz, 166 kHz and 500 kHz, and it seems like there was an additional hold-off time of 1 μs or that the diode was not active for this interval of time after the end of the hold-off time.

The same behavior is present also at higher frequency (the data and graphs are collected in the appendix, respectively in table 5 and 6 for 10^5 and 10^6 photons per second) and similar analysis and considerations can be obtained, leading again to an additional hold-off time of 1 μs.

The reason of this fact is still unclear, but we hope that, when an integrated circuit for generating the hold-off time will be done, it disappears.

Then are shown the evolutions of the quantum (or corrected) and effective efficiency versus the applied bias voltage.

(a)



(b)

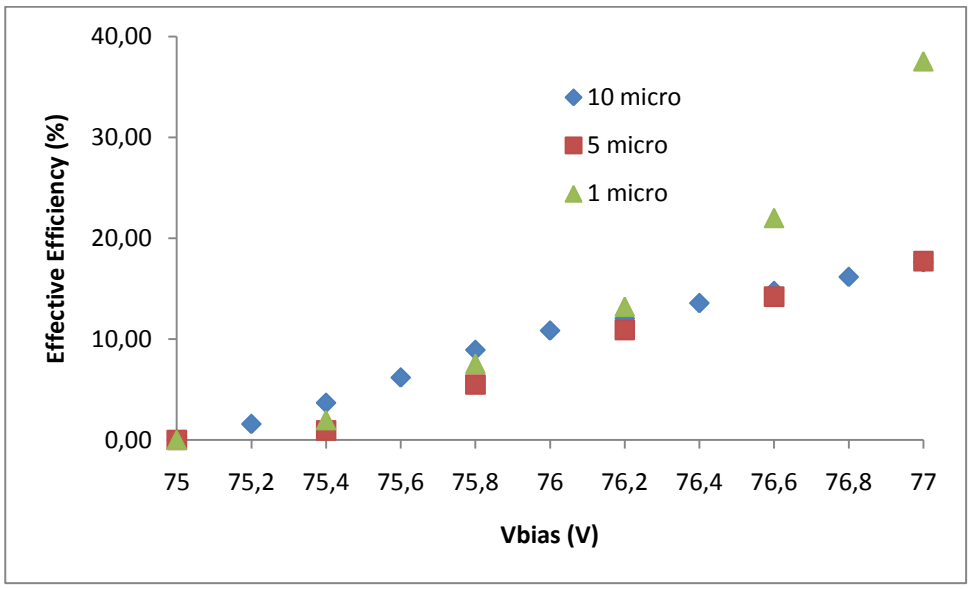


Figure 5.17. Evolution of quantum, (a), and effective, (b), efficiencies of the NFAD tested in free-running mode with 10^4 photons per second for three different values of the hold-off time.

Looking at the graphs in Fig. 5.16, we can note than the behaviors of the quantum efficiency (corrected for hold-off time) and of the effective efficiency are similar for low bias voltages. They are both equal to ≈ 5% at 75,5 V and to ≈ 10 % at 76 V with an almost linear evolution. After ≈ 76,5 V the two behaviors start to diverge: the quantum efficiency is again quite linear up to 77 V for each value of hold-off time, while in the effective efficiency case the behavior with 1-μs-hold-off-time is far from linear and diverges from the behavior with 5 and 10 μs hold-off-times.

The observed exponential increasing of the efficiency for high bias and low hold-off is due to the afterpulsing effect. At lower bias the fact that the data measured for different hold-off are similar, suggests that for these values of bias the afterpulsing is negligible. On the contrary, when the behavior of the efficiency starts to increase exponentially, we cannot exclude effects connected with afterpulsing. Currently, deeper studies are necessary to correct the formula of the efficiency in order to account afterpulsing effect.

Finally we can see in figure 5.17 the evolution of the dark counts frequency versus the effective efficiency, again for the three values of hold-off time for which the NFAD was tested. The dark counts frequency of the detector at 10% effective efficiency is lower than 10^3. This value is the lowest one obtained using passive-quenched detectors and it is comparable with the performances of gated-mode detectors; in comparison with them or with active-quenched detectors, however, passive-quenched detectors are easier to realize

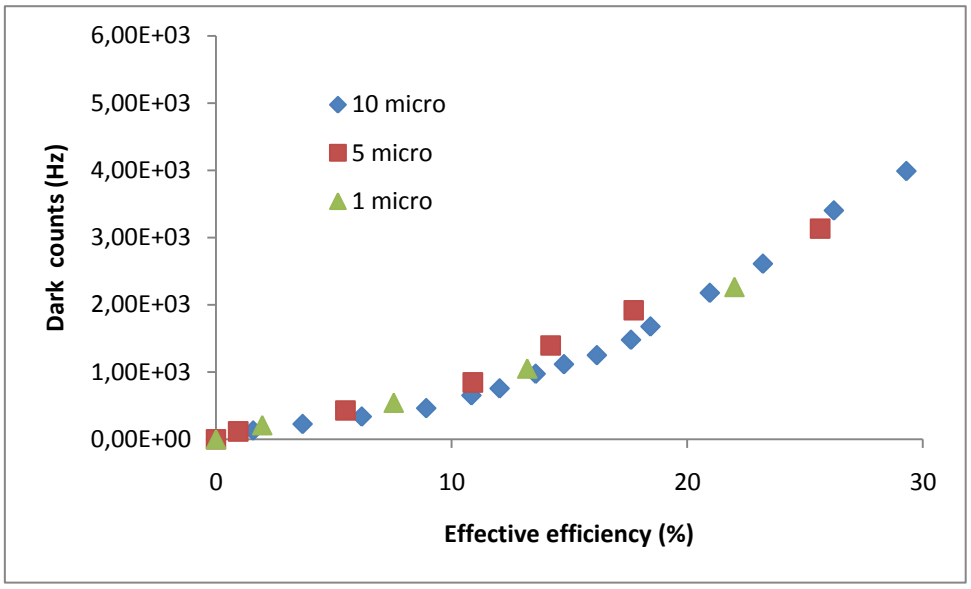


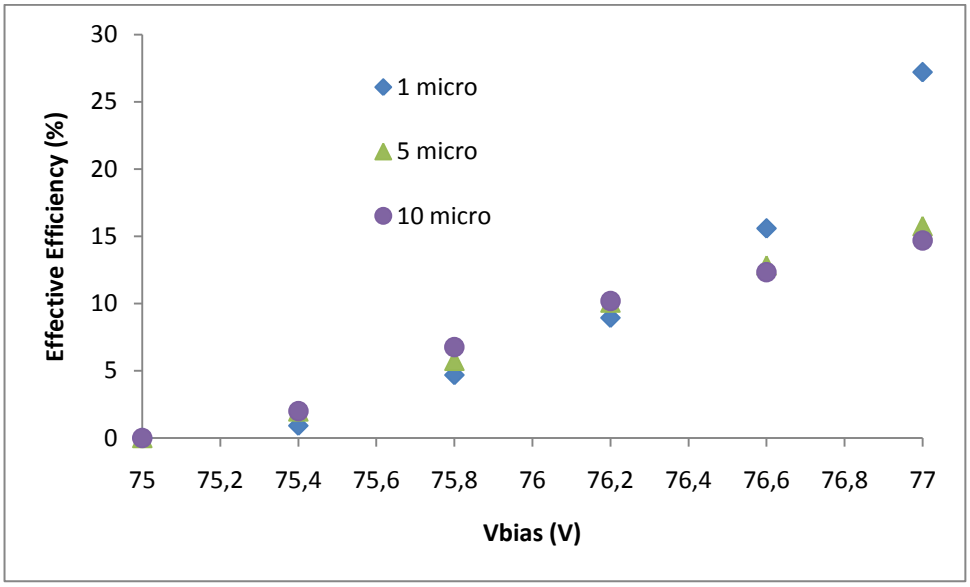
Figure 5.18. Free-running mode; dark counts frequency versus effective efficiency

Then, the diode is tested firstly with an average number of 10^5 photons per second, changing the applied hold-off time, from a minimum of 1 µs to a maximum of 10 µs. The data thus obtained are shown in the table 5 in the appendix, together with the figures obtained from these data.

Finally, the NFAD is tested with an average number of 10^6 photons per second, changing the applied hold-off time, which now takes the values of 1µs, 0,5µs and 0,2 µs. The data thus obtained are shown in the appendix (table 6), again with the associated graphs.

In order to compare the behavior of the detector at a higher average number of photons per second, we show in figure 5.18 the effective efficiency as a function of the applied bias, for 10^5 and 10^6 photons per second.

(a)



(b)

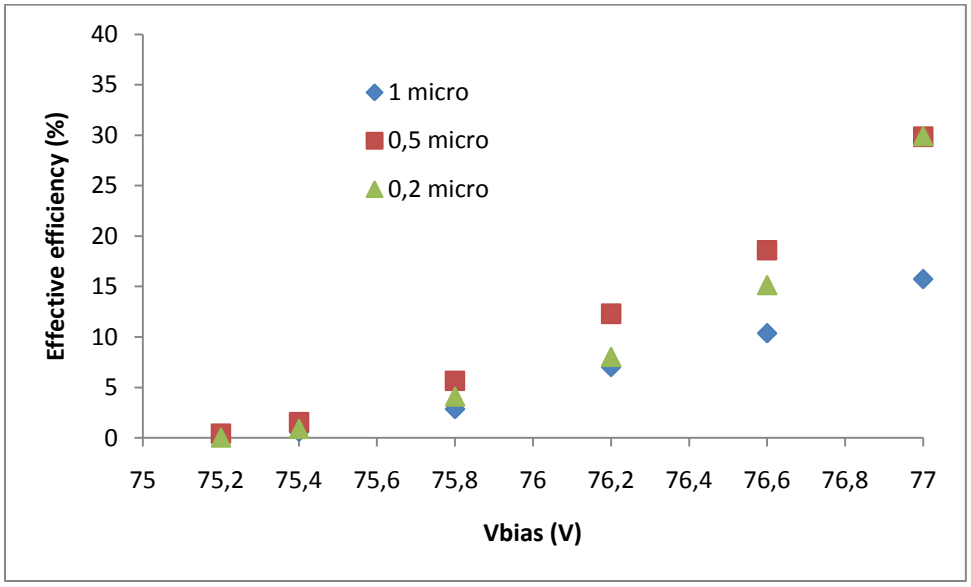


Figure 5.19 Effective efficiency as a function of applied bias (a) for 10^5 and (b) 10^6 photons per second.

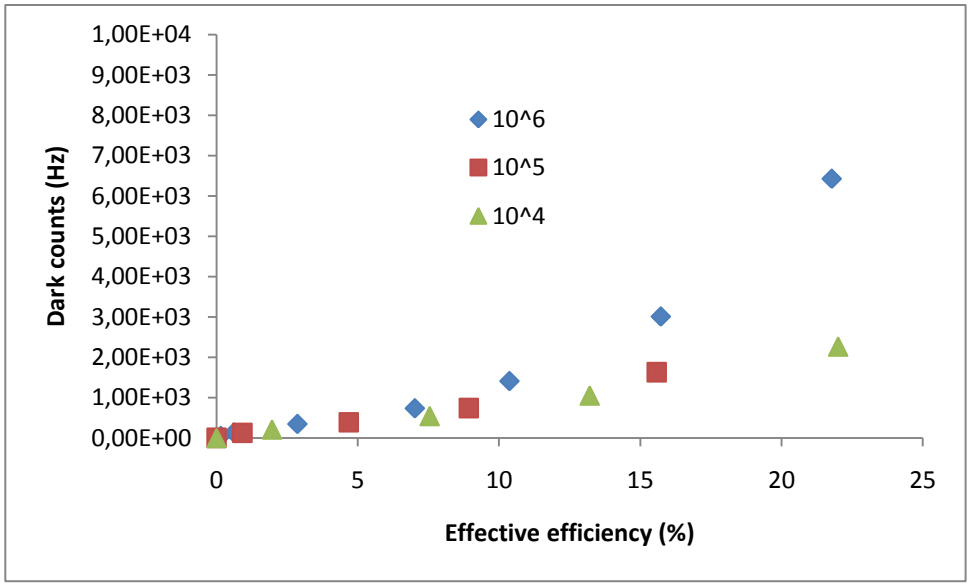


Figure 5.20. Dark counts as a function of effective efficiency for an hold-off time of 1 µs at different average numbers of photons per second.

In figure 5.19, the behavior of the detector with an applied hold-off time of 1 µs is shown, changing the average numbers of photons per second. Up to 10% effective efficiency, the results are similar in any case, and a dark counts rate of less then 10^3 [Hz] is confirmed ; at higher efficiency values, the behavior of the detector with 10^6 photons per second diverges, due to the higher afterpulsing effect.

5.7 Jitter measures.

Timing resolution, also referred to as jitter, is the uncertainty in the detection time of an avalanche when incident photons arrive at a fixed time.

Performing this experiment we have measured the total overall jitter of the system, due to the laser pulses, the electronic devices used for reading the signal and to the intrinsic jitter of the NFAD.

In order to calculate the jitter we have used a TCSPC board with a resolution of ≈ 12 ps and we have performed the experiment at 223.15 K.

The experimental set-up is thus organised: a digital delay pulse generator (SRS) (Dg 535, Stanford Research System, Inc.) triggers a 1550-nm Laser Diode (LD) which produces short optical pulses with ≈200 ps full-width at half-maximum (FWHM) at a repetition frequency of 100 kHz and with an intensity of 1 photon per pulse. The SRS gives also the starting signal to the TCSPC (time correlated single-photon counting) board, while the stop signal is provided by the NIM output of the discriminator which is put after the amplifier. The SRS estabilish also a precise delay (≈1 ps of precision) between the two signals.

The jitter is measured for different values of the bias voltage and, as shown in fig..., it decreases with increasing bias. A diagram of the set-up is shown in figure...

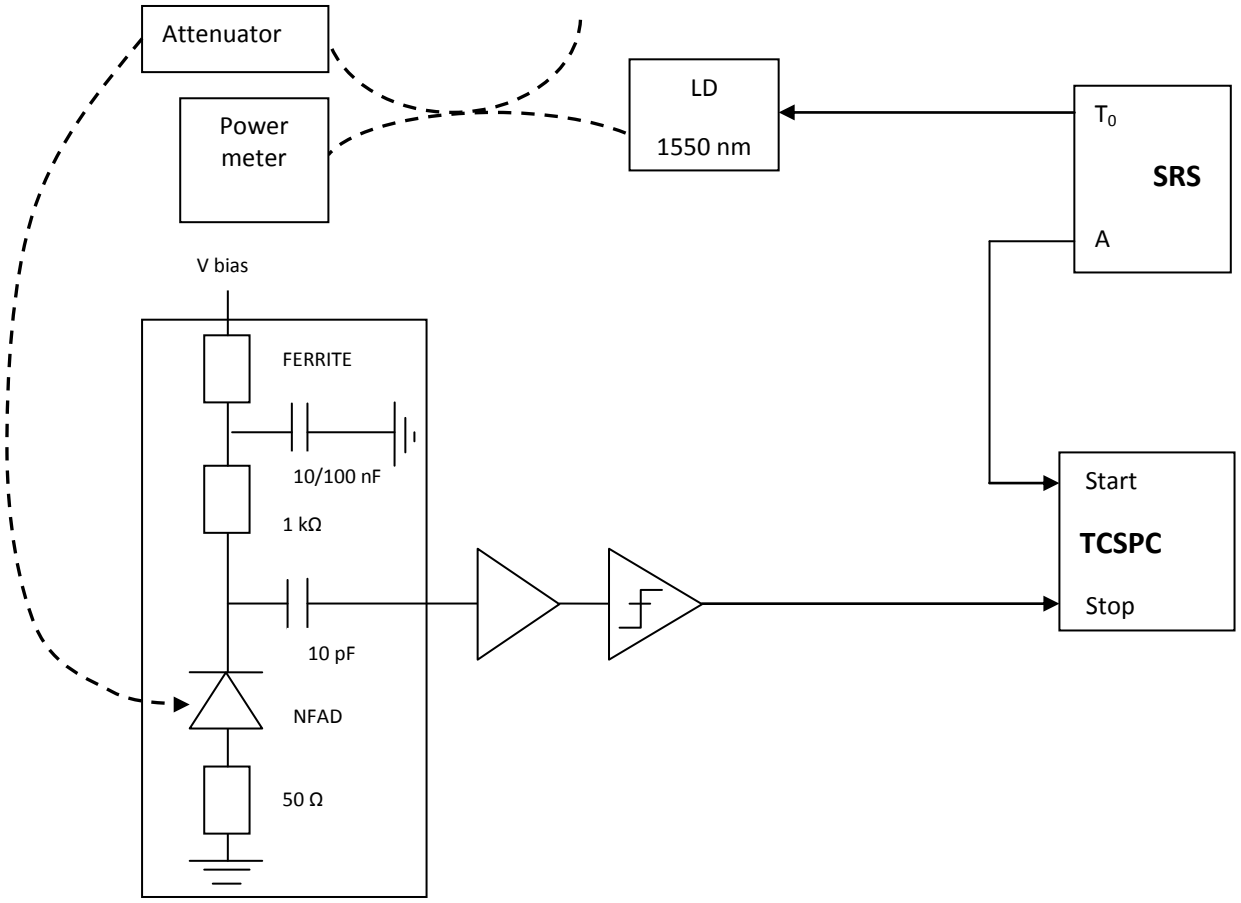


Figure 5.21 Experimental set-up used for the jitter measures. A digital delay pulse generator (SRS) triggers both the Laser Diode (LD) and a TCSPC board; the stop signal is provided by the output of the discriminator.

The results obtained are recoiled in the next table.

Vbias	detections	noise	Eff(%)	Max counts	Time max	Time half1	Time half 2	jitter
75.6	4468	2399	2.069	242	36.8041	36.3891	37.085	0.6959
75.8	8781	4499	4.282	481	36.7065	36.4258	36.9385	0.5127
76	12996	6685	6.311	690	36.5478	36.3037	36.731	0.4273
76.2	22319	11825	10.494	874	36.3769	36.2183	36.5844	0.3661
76.4	61947	36200	25.747	1079	36.2670	36.1329	36.4563	0.3234

In order to calculate the overall jitter of the system we look for the time for which the TCSPC has provided the maximum number of detections and we take this number; then we look for the two times for which the TCSPC has taken the half of the maximum detection counts. If, as usual, there is not a column of the histogram with exactly this number of counts, we interpolate between the two columns with the nearest number and we obtain the so-called “Time half 1” and “Time half 2” in the table; finally, the jitter is considered the difference between these two numbers.

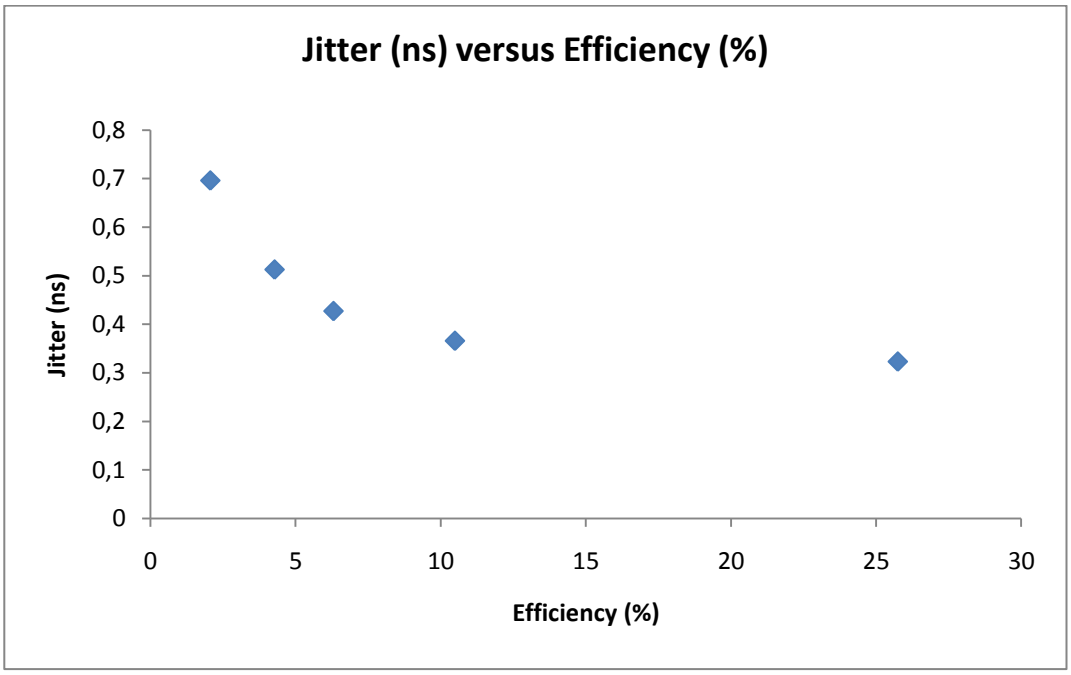


Figure 5.22 Jitter versus efficiency measured at 223 K and with 10^5 pulses per second with 1 photon per pulse. The FWHM of the laser pulses is ≈ 200 ps.

As shown in Fig. 5.21 the jitter decreases with increasing efficiency; the best value, obtained for efficiencies greater than ≈25% is around 300 ps. Since the FWHM

of laser pulses is ≈200 ps, this means that the minimum jitter of all the system, except for the laser, ≈223 ps. In fact the square of the total jitter is the sum of the square of the FWHM of the laser and the square of the jitter of all the system (except the laser).

Using a laser with a littler pulse width, it should be possible to achieve a total jitter near to this value.

6. Potentiality of the NFAD under test for its application in a QKD system based on the COW protocol. Based on an internal report made by Nino Walenta.

In the evaluation of the performances of the hold-off technique applied on a NFAD, it is important to characterize not only the reduction of the afterpulsing probability, but also the effects of this technique on the raw key creation rate R_{raw}, before error corrections and privacy amplification, which is a quantity usually presented to characterize the performance of a QKD system [51][52].

6.1 QKD simulations.

In the previous chapter a free-running NFAD was tested for different values of bias voltage and hold-off time. Based on these results we want to estimate its potential for a QKD system based on the coherent one-way (COW) protocol[52][53].

First, the detection rate without illumination, R_{noise}, was measured. This is effectively a measure of dark counts, R_{dc}, and their afterpulses, R_{ap} [53]:

$$R_{noise} = \frac{R_{dc} + R_{ap}}{1 + \tau_h (R_{dc} + R_{ap})}$$

This quantity, R_{noise}, was usually referred as dark count rate, or R_{dc}, in the previous chapters, but since now our aim is to discriminate between the "true" dark counts and their afterpulses, this new nomenclature is introduced.

For a sufficient long hold-off time τ_h, we can assume the afterpulses-of-dark-counts rate R_{ap} equal to zero. Hence, the real dark count rate R_{dc} I obtained by fitting an exponential decay to the $R_{noise}(\tau_h)$ curve and calculating

$$\boldsymbol{R_{dc} = \frac{R_{noise}}{1 - \tau_h R_{noise}}} \textbf{ = 631.21 Hz}$$

Then, the number of detections due to afterpulses of dark counts is

$$\mathbf{R_{dc} = \frac{R_{noise}}{1 - \tau_h R_{noise}} - R_{dc}}$$

If for the moment we assume that afterpulses stem only from dark counts (neglecting high order afterpulsing), then the afterpulse probability is [53]:

$$\mathbf{P_{ap}} = \frac{R_{ap}}{R_{dc}} = \frac{R_{noise}}{R_{dc} - \tau_h R_{noise} R_{dc}} - 1 = \frac{R_{noise,corr} - R_{dc}}{R_{dc}}$$

where

$$\mathbf{R_{noise,corr} = \frac{R_{noise}}{1 - \tau_h R_{noise}} = R_{dc} + R_{ap}}$$

When the detection rate with illumination, R_{phot}, was measured, what was effectively measured is:

$$\mathbf{R_{phot}} = \frac{R_Q + R_{dc} + R_{ap}}{1 + \tau_h (R_Q + R_{dc} + R_{ap})} = \frac{(R_Q + R_{dc})(1 + p_{ap})}{1 + \tau_h (R_Q + R_{dc})(1 + p_{ap})}$$

Where $R_Q = \mu f \eta_Q$ is the detection rate due to photons. Note that now we assume that afterpulses can stem from dark counts or photon detections, but not from previous afterpulses. Knowing the dark count rate and the afterpulse probability we can calculate the detection rate [53]:

$$\mathbf{R_Q = \frac{R_{phot}}{(1 - \tau_h R_{phot})(1 + p_{ap})} - R_{dc}}$$

And finally

$$\boldsymbol{\eta = \frac{R_Q}{\mu f}}$$

Where μ is the average number of photon per pulse ans f is the repetition frequency of the pulses (in free-running mode we consider only the product of this two quantities).

The next figure shows the secret key rate as function of fiber length for this detector (with different hold-off time applied) and in comparison with an SSPD ($\eta = 0.07$, R_{dc} = 50 Hz, τ_h = 20 ns) and a sinus gated detector ($\eta = 0.1$, R_{dc} = 5000 Hz, τ_h = 10 ns, p_{ap}= 0.04).

Looking at the simulations obtained for the detector based on NFAD we can note that, at a fixed fiber length, the secure key rate for the COW protocol approaches the secure key rates obtained with sinus gating for almost every length up to ≈ 170 km, where both detectors start to give very low rates. Instead, making the same comparison with the simulations obtained with an SSPD detector, we cannot approach its secure key rate only after ≈ 170 km, even if for these values of fiber length the secure key rates are very low (less than 100 Hz).

This result is very encouraging if we want to develop a QKD system based on a NFAD.

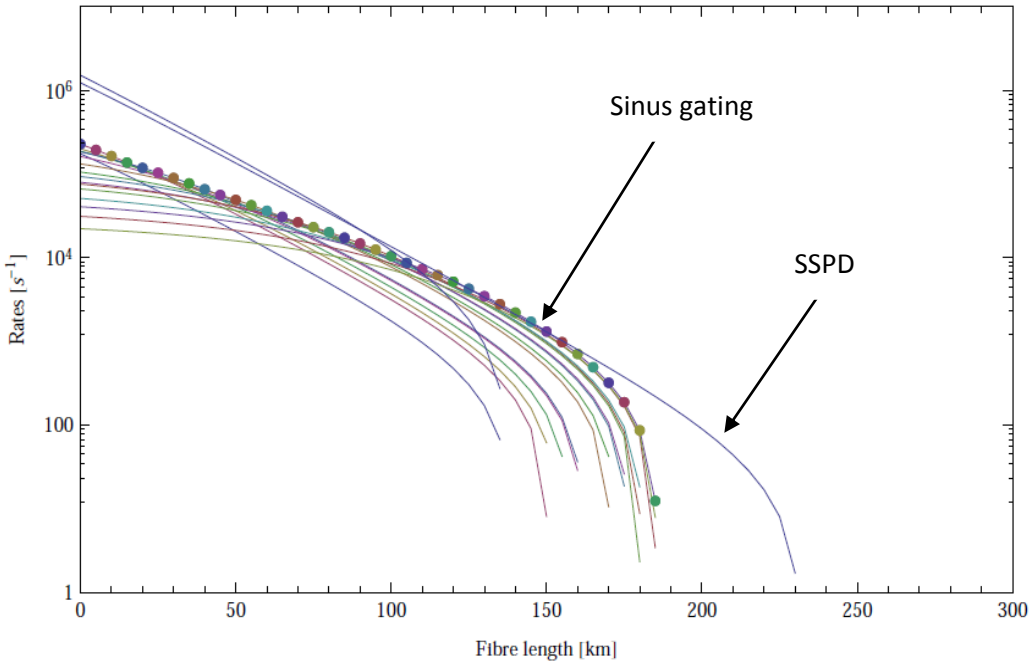


Figure 6.1. The simulations of secure key rate for the COW protocol versus fiber distance, using the NFAD with different hold-off times applied (coloured lines), the sinus gating detector with 10 ns of hold-off time (line with superimposed points) and the SSPD detector with 20 ns of hold-off time (blue line shown by the arrow)[53].

We have made another simulation, with the aim of finding the optimal hold-off time for each fiber length. The results are shown below.

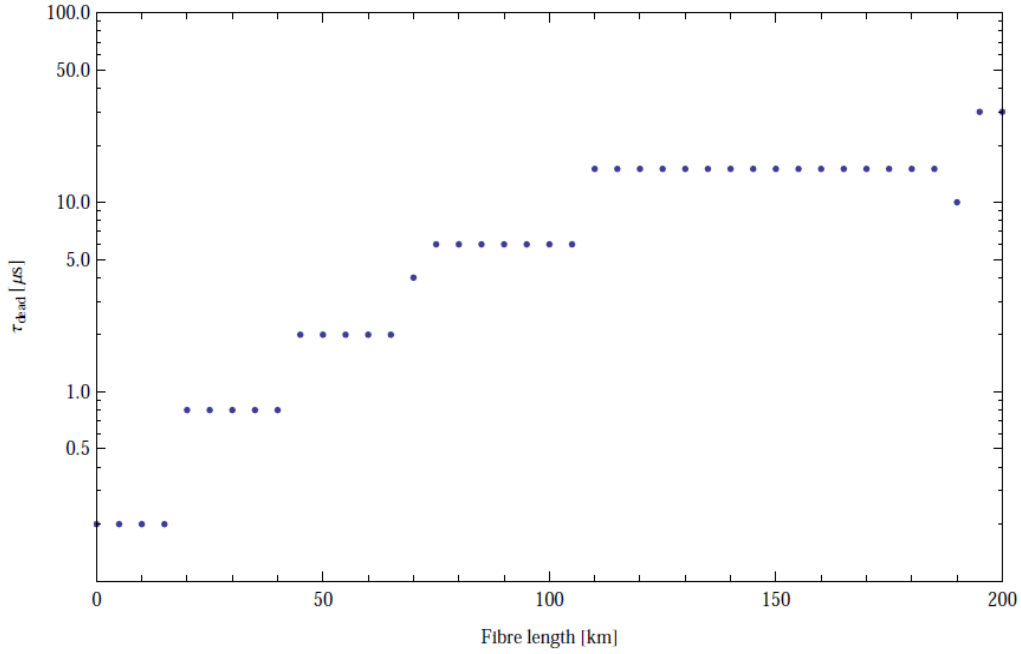


Figure 6.2. Optimal hold-off time versus fiber length[53].

This result is very important, because it suggests us how to vary the hold-off time imposed to the diode, depending of the fiber length used in the QKD system. The suggested hold-off time, as we can see in fig. 6.2, varies from 0.1 to 20 µs, for a fiber length ranging from 0 to 200 km.

7. Conclusions

In conclusion, we have presented the results obtained with a InP Negative Feedback Avalanche Diode (NFAD), which seems interesting for free running operation; exploiting its monolithically-integrated resistance, in fact, it ensures a rapid

quenching of the avalanches generated in the APD, thus reducing the number of carriers trapped and consequently the afterpulse probability. Therefore it is possible to use the diode in free-running mode, applying a very short deadtime in order to empty the trapped charge.

For these reasons, we have tested the diode in order to see its behavior in terms of efficiency and dark counts rate, both in gating and free-running mode; obtaining less than 1 kHz @ 10% efficiency. Then, in order to determine the correct dead time, we have characterized the afterpulse probabilities using a new algorithm, implemented by means of an FPGA hardware. The algorithm is able to record the first afterpulse occurring after the detection one (useful for the characterization of the afterpulse phenomena) or all the afterpulse occurring in the entire range of measure (useful to characterize the detector in operating condition).

Then, we have developed a circuit which provides a hold-off time for the detector : everytime an avalanche voltage signal is detected, this circuit triggers a variable monostable in order to apply the desidered hold-off time. We have thus reported a DCR of ≈ 0,8 kHz @ 10 % efficiency, using a hold-off time of 1 μs and sending 10^4 photons per second into the detector and ≈ 1 kHz increasing the average number of photons per second up to 10^6.

Finally, we have measured the timing resolution, obtaining a jitter of less then 400 ps @10% efficiency.

Appendix

Table 1: Detections with and without the laser in the free-running mode.

Vbias (V)	Dark Counts (over 30s)	Dark Counts Rate (Hz)	Detections with laser (over 30 s)	Det Rate (Hz)	Effective efficiency
74.700	639	21.3000	891	29.7000	0.0840
74.900	3982	132.7333	6805	226.8333	0.9410
75.000	5621	187.3667	9586	319.5333	1.3217
75.100	6657	221.9000	12836	427.8667	2.0597
75.200	8286	276.2000	16610	553.6667	2.7747
75.300	10648	354.9333	22198	739.9333	3.8500
75.400	13538	451.2667	29077	969.2333	5.1797
75.500	16924	564.1333	35165	1172.1667	6.0803
75.600	21335	711.1667	42546	1418.2000	7.0703
75.700	27081	902.7000	51935	1731.1667	8.2847
75.800	34188	1139.6000	64575	2152.5000	10.1290
75.900	46363	1545.4333	86436	2881.2000	13.3577
76.000	62689	2089.6333	113535	3784.5000	16.9487
76.050	73347	2444.9000	131181	4372.7000	19.2780
76.100	85910	2863.6667	154269	5142.3000	22.7863
76.150	106538	3551.2667	189595	6319.8333	27.6857

Table 1.1 Data from the experiment: changing the bias, the evolution of efficiency and dark counts were reported. Dark counts were measured for 30 s and total detections (with the laser) were measured for 10 s; both values were chosen in order to have a good accuracy.

Table 2: Detections with and without the laser in Gated mode.

Vbias (V)	Dark Counts	DC Rate (Hz)	Dark counts per ns	Dark Counts per second	Detection Counts	Det Rate (Hz)	Efficiency (%)
76	13	0.1083	1.0833E-07	108.333333	709	23.6333	0.2355
76.25	33	0.2750	2.7500E-07	275	4349	144.9667	1.4575
76.5	52	0.4333	4.3333E-07	433.333333	11079	369.3000	3.7586
76.75	73	0.6083	6.0833E-07	608.333333	19454	648.4667	6.6984
77	106	0.8833	8.8333E-07	883.333333	23613	787.1000	8.1892
77.25	128	1.0667	1.0667E-06	1066.66667	27771	925.7000	9.7032
77.5	176	1.4667	1.4667E-06	1466.66667	32051	1068.3667	11.2839
77.75	219	1.8250	1.8250E-06	1825	35886	1196.2000	12.7219
78	296	2.4667	2.4667E-06	2466.66667	40511	1350.3667	14.4821
78.25	363	3.0250	3.0250E-06	3025	44737	1491.2333	16.1186
78.5	481	4.0083	4.0083E-06	4008.33333	49505	1650.1667	17.9943
78.75	775	6.4583	6.4583E-06	6458.33333	53834	1794.4667	19.7130
79	1244	10.3667	1.0367E-05	10366.6667	59846	1994.8667	22.1465
79.25	2661	22.1750	2.2175E-05	22175	65148	2171.6000	24.2607
79.5	10399	86.6583	8.6658E-05	86658.3333	72935	2431.1667	26.9843

Table 2.1 Data obtained in the gated-mode experiment: changing the bias, the evolution of efficiency and dark counts were reported. The step in dc bias was set at 0.25 V. Dark counts were measured for 120 s and total detections (with the laser) were measured for 30 s; both values were chosen in order to have a good accuracy. The DC Rate is obtained simply dividing the Dark counts per 30, while the Dark Counts per second are obtained considering only the time during which the diode is active.

Tables 3. Afterpulsing measures with double-gated Method.

DC (AB+CD)(Hz)	DC (AB) (Hz)	Det Rate (AB+CD) (Hz)	Det Rate (AB) (Hz)	P per ns	Efficiency (%)	Delay of the second gate (µs)
2.8333	1.26667	948.5000	894.4333	0.000588	9.9371	**0.5**
2.5000	1.26667	985.4000	957.1000	0.000283	10.3490	**0.995**
2.9333	1.33333	946.6667	936.0333	0.000097	9.9159	**2**
3.6333	1.76667	919.7667	911.7000	0.000068	9.6122	3
2.9333	1.50000	926.4333	920.7667	0.000046	9.6926	4
2.4667	1.10000	930.3000	925.6667	0.000035	9.7399	5
2.7333	1.40000	896.6333	892.7333	0.000029	9.3667	6
2.8000	1.43333	892.0667	889.2333	0.000017	9.3159	7
2.5333	1.23333	907.1000	904.5333	0.000014	9.4838	8
2.6667	1.40000	910.2667	907.9333	0.000012	9.5173	9
2.7667	1.33333	929.9000	927.9333	0.000006	9.7325	10

Table 3.1: Gate width of 100 ns and laser pulse at the beginning

DC (AB+CD)(Hz)	DC (AB) (Hz)	Det Rate (AB+CD) (Hz)	Det Rate (AB) (Hz)	P per ns	Efficiency (%)	Delay of the second gate (µs)
2.5667	1.30000	1107.9333	1052.4667	0.000516	11.7169	0.5
2.4667	1.40000	1086.8667	1059.9333	0.000244	11.4813	1
2.3667	1.10000	1085.6333	1073.2000	0.000104	11.4684	2
2.4000	1.23333	1078.5000	1070.1000	0.000068	11.3881	3
2.1000	1.26667	1077.3333	1071.8667	0.000043	11.3780	4
2.5333	1.50000	1082.0667	1077.6667	0.000031	11.4267	5
2.2667	1.26667	1084.2667	1081.1667	0.000019	11.4541	6
1.9667	1.06667	1075.0667	1072.1000	0.000019	11.3540	7
2.1667	1.26667	1082.2333	1079.8333	0.000014	11.4323	8
2.1667	1.06667	1070.5333	1068.5000	0.000009	11.3012	9
2.1000	1.23333	1079.1000	1077.2667	0.000009	11.3978	10

Table 3.2: Gate width of 100 and laser at the end.

DC (AB+CD)(Hz)	DC (AB) (Hz)	Det Rate (AB+CD) (Hz)	Det Rate (AB) (Hz)	P per ns	Efficiency (%)	Delay of the second gate (µs)
1.6167	0.78333	1020.8000	956.8667	0.001320	10.7513	0.2
1.5167	0.71667	986.6333	959.2333	0.000555	10.3725	0.5
1.3167	0.66667	972.4667	959.7667	0.000251	10.2174	1
1.5167	0.70000	968.2667	962.7667	0.000097	10.1689	2
1.5500	0.71667	961.8333	958.1667	0.000059	10.0974	3
1.5833	0.73333	965.4333	962.9000	0.000035	10.1369	4
1.6167	0.70000	965.6667	963.5333	0.000025	10.1391	5
1.5167	0.71667	977.7667	976.1000	0.000018	10.2742	6
1.4667	0.68333	974.6000	973.0667	0.000015	10.2396	7

Table 3.3: Gate width of 50 ns and 1 photon per pulse.

DC (AB+CD)(Hz)	DC (AB) (Hz)	Det Rate (AB+CD) (Hz)	Det Rate (AB) (Hz)	P per ns	Efficiency (%)	Delay of the second gate (µs)
1.3000	0.60000	6599.9000	6252.3000	0.001110	10.7865	0.2
1.0333	0.51667	6499.5333	6248.2333	0.000803	10.4959	0.3
1.3333	0.60000	6460.9000	6254.0667	0.000659	10.3858	0.4
1.3333	0.51667	6446.0000	6269.1667	0.000562	10.3438	0.5
1.2000	0.51667	6370.5667	6286.7333	0.000265	10.1339	1
1.2667	0.63333	6334.9667	6300.3000	0.000108	10.0362	2
1.3333	0.71667	6306.8667	6285.2667	0.000067	9.9598	3
1.3000	0.63333	6303.6000	6290.0667	0.000041	9.9510	4
1.3833	0.61667	6307.6000	6296.3333	0.000033	9.9617	5
1.2667	0.60000	6276.0000	6268.9667	0.000020	9.8766	6
1.2500	0.68333	6324.3000	6318.5000	0.000017	10.0072	7
1.3333	0.61667	6306.6333	6302.3333	0.000011	9.9591	8
1.3667	0.68333	6321.6000	6317.8000	0.000010	9.9997	9
0.0000	0.58333	6321.6000	6318.3667	0.000012	10.0011	10

Table 3.4: Gate width of 50 ns and 10 photons per pulse.

Tables 4: Hold-off time measures for 10^4 photons per second using different hold-off times.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75	0	0	0,00	0,00
75,2	286,3	128,7	1,58	1,58
75,4	593,7	226,1	3,68	3,71
75,6	955,2	336,6	6,19	6,27
75,8	1354,3	462	8,92	9,09
76	1736,6	652,4	10,84	11,11
76,2	1959,2	755,4	12,04	12,37
76,4	2329,9	973	13,57	14,03
76,6	2592,4	1115,4	14,77	15,33
76,8	2866,5	1249,8	16,17	16,85
77	3240,1	1478,5	17,62	18,48
77,2	3520,6	1676,3	18,44	19,44
77,6	4273	2176,8	20,96	22,38
78	4931,3	2609,8	23,22	25,07
78,6	6024,9	3402,2	26,23	28,89
79	6918,7	3988,2	29,31	32,79
79,4	8199,2	4930,5	32,69	37,45
79,8	10920,9	6905,5	40,15	48,42
80,2	19985,5	13573,6	64,12	92,72
80,3	25938,4	18552,2	73,86	122,45
80,4	33613	25148,2	84,65	170,35
80,5	55169	46304,2	88,65	368,26
80,6	80685,6	75269,4	54,16	1133,91
80,7	86171,3	83861,6	23,10	1034,94
80,8	90833,9	90671,2	1,63	190,27
80,9	91299,4	91252,1	0,47	62,15
81	91387,9	91376,9	0,11	14,81

Table 4.1: Data obtained with a hold-off time of 10 μs.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	quantum efficiency (%)
75	0	0	0,00	0,00
75,4	211,4	117	0,94	0,95
75,8	979,8	430,2	5,50	5,53
76,2	1935,5	845,1	10,90	11,06
76,6	2817,2	1396,5	14,21	14,51
77	3691,7	1918,8	17,73	18,24
77,6	5696,6	3131,7	25,65	26,82
78,6	11522,6	7414	41,09	45,28
79	14245,2	9146,7	50,99	57,53
79,4	21456,2	14301,5	71,55	86,32
79,8	35937,8	25338,7	105,99	147,95
80,2	89935,9	75124,9	148,11	431,05
80,3	125710	115186,6	105,23	668,07
80,4	151379,1	146780,2	45,99	710,92
80,5	158726,6	156093,7	26,33	581,16
80,6	165846,4	165373,3	4,73	160,02
80,8	168165,9	168151,8	0,14	5,56

Table 4.2: Data obtained with a hold-off time of 5 μs.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75	0	0	0	0,00
75,4	405,9	209,6	1,963	1,96
75,8	1298,2	543,7	7,545	7,56
76,2	2373,3	1052,1	13,212	13,26
76,6	4466,7	2265,8	22,009	22,16
77	8122	4367,6	37,544	38,02
77,6	24226,3	14569,3	96,57	100,43
77,8	35911,2	22002,3	139,089	147,52
78	51396,3	31732,3	196,64	214,09
78,2	88378,1	57300,8	310,773	361,62
78,4	145698,4	100498,2	452,002	588,20
78,6	238624,2	182290,7	563,335	904,83
78,8	335934,2	288270,1	476,641	1008,47
79	400195,9	370190	300,059	794,31
79,2	465014,2	452670,9	123,433	123,43
79,4	489468,6	486238,6	32,3	123,15
79,6	503186	502435,1	7,509	7,51
79,8	509257,8	509089,2	1,686	7,00

Table 4.3: data obtained with a hold-off time of 1 μs.

Tables 5: Hold-off time measures for 10^5 photons per second using different hold-off times.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75	0	0	0	0,00
75,4	2168,6	171,2	1,9974	2,05
75,8	7166,1	396,1	6,77	7,32
76,2	10842,2	658,7	10,1835	11,50
76,6	13319,8	982,5	12,3373	14,37
77	16028,3	1342	14,6863	17,73
77,4	18068,7	1648,6	16,4201	20,38
77,8	20314,8	2083,7	18,2311	23,37
78,2	22869,9	2713,2	20,1567	26,86
78,6	24571,5	3197,9	21,3736	29,27
79	27144,8	3842,4	23,3024	33,26
79,4	29778,3	4662,5	25,1158	37,52
79,6	32451	5575,2	26,8758	42,14
79,8	35577,8	6530,8	29,047	48,24
80	41910,6	8692,1	33,2185	62,63
80,2	53810,9	14768,9	39,042	99,17
80,4	75360,5	37236,5	38,124	246,52
80,6	83790,4	58876,5	24,9139	373,75
80,7	89118,9	80001,5	9,1174	418,99
80,8	91020,8	89650,5	1,3703	147,45
81	91441,1	91426,6	0,0145	1,98

Table 5.1: Data obtained with a hold-off time of 10 µs.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75	0	0	0	0,00
75,4	2154,3	179,5	1,9748	2,00
75,8	6107	367,3	5,7397	5,93
76,2	10799,5	730,3	10,0692	10,68
76,6	13947,4	1126,1	12,8213	13,86
77	17329,5	1580,6	15,7489	17,38
77,4	21629,6	2201,9	19,4277	22,03

77,8	26568,6	3061,9	23,5067	27,53
78,2	32227,4	4209,8	28,0176	34,12
78,6	39182,7	5756,5	33,4262	42,80
79	47001	7812,7	39,1883	53,31
79,4	62248,4	12552,7	49,6957	76,98
79,6	73242,7	16865,5	56,3772	97,14
79,8	94676,1	28423,8	66,2523	146,65
80	122893,3	55414,1	67,4792	242,11
80,2	149664,7	108111,7	41,553	359,36
80,3	156294	127815	28,479	361,07
80,4	166306,3	162717,4	3,5889	114,28
80,5	166847,3	164469,4	2,3779	80,75
80,6	168068,4	167673,6	0,3948	15,30

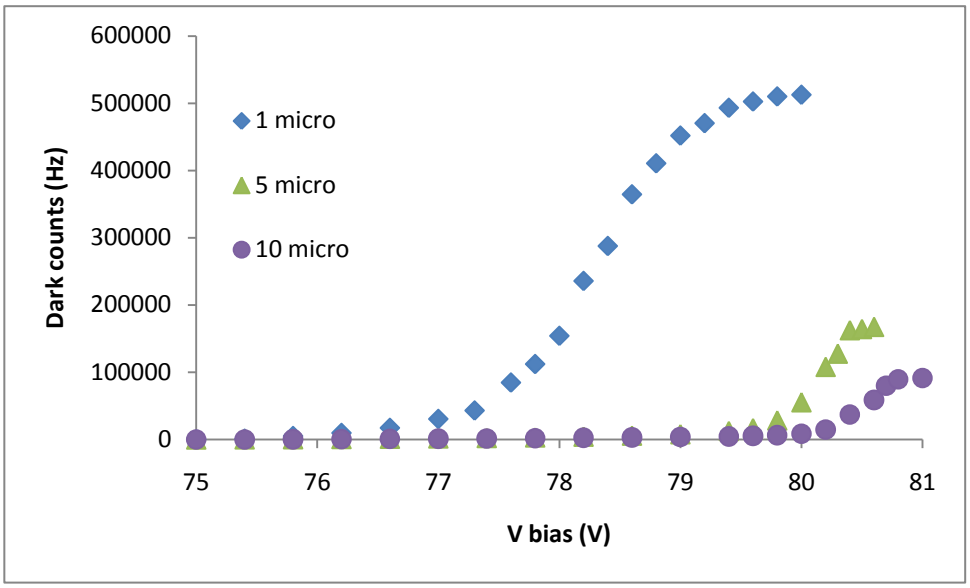
Table 5.2: data obtained with a hold-off time of 5 μs.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75	0	0	0	0,00
75,4	1037,5	124,2	0,9133	0,91
75,8	5067,5	387,7	4,6798	4,71
76,2	9672,3	739,5	8,9328	9,03
76,6	17207,7	1627,6	15,5801	15,88
77	30497,8	3301,8	27,196	28,14
77,3	43024,3	6018,4	37,0059	38,90
77,6	84612,9	12030,3	72,5826	80,26
77,8	112152,6	18849,7	93,3029	107,11
78	154126,2	31241,2	122,885	149,96
78,2	235791,7	64097,1	171,6946	240,06
78,4	287787,7	100408,1	187,3796	292,46
78,6	364609,9	187764,1	176,8458	342,67
78,8	410743,3	270419,9	140,3234	326,40
79	451998,3	374549,4	77,4489	225,96
79,2	470485,1	430310,8	40,1743	133,18
79,4	493351,5	482082,5	11,269	42,95
79,6	502697,6	499580,2	3,1174	12,53
79,8	510234,1	509934	0,3001	1,25
80	512765,3	512665,3	0,1	0,42

Table 5.3: Data obtained with a hold-off time of 1 μs are presented.

Now the figures extracted from the data just presented are shown, starting with the evolution of the dark counts versus the applied bias voltage. A zoom of the first part of the graph is later depicted.

(a)



(b)

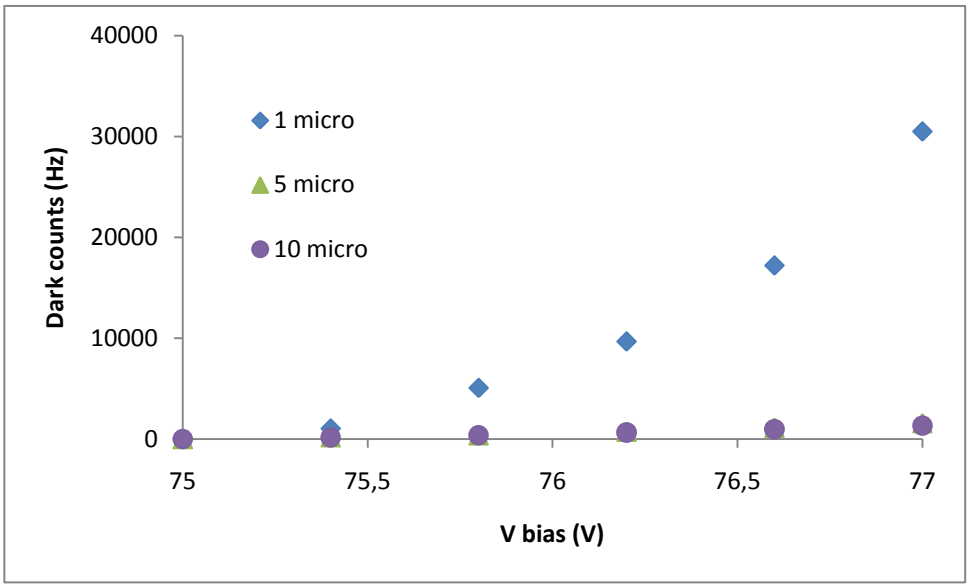
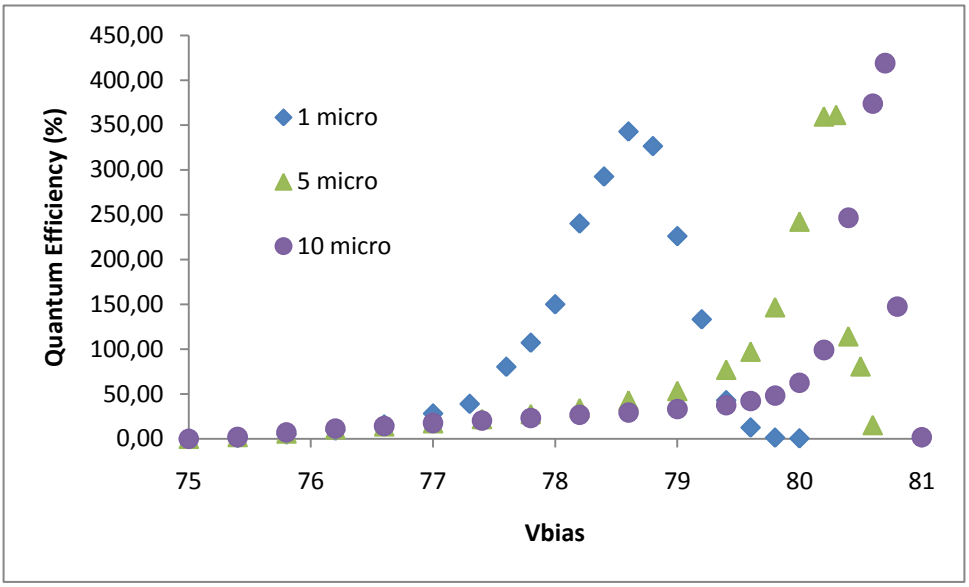
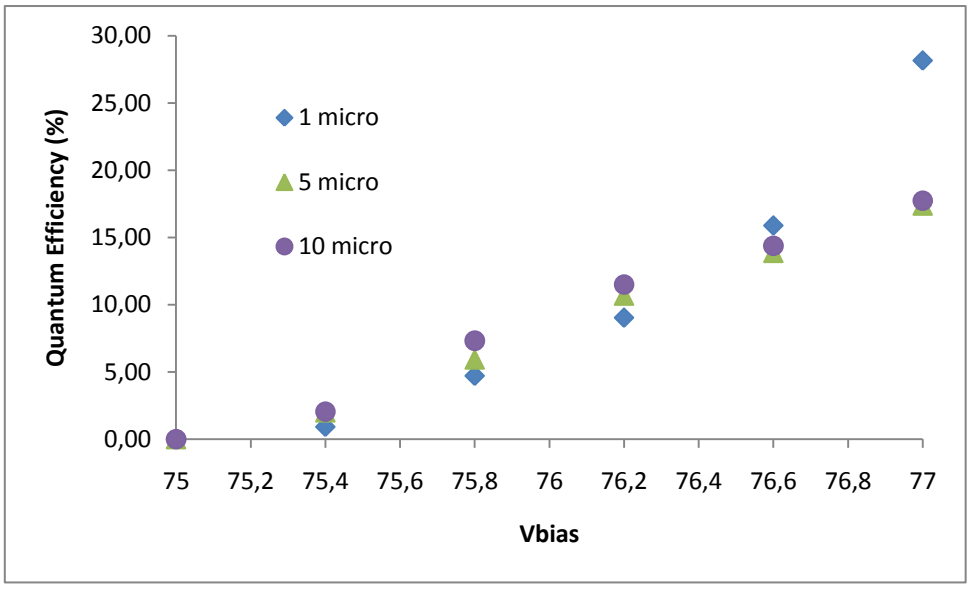


Fig A5.1 Noise characteristic of the NFAD tested in free-running mode with 10^5 photons per second for three different values of the hold-off time, (a), and zoom of the first part of the graph ,(b).

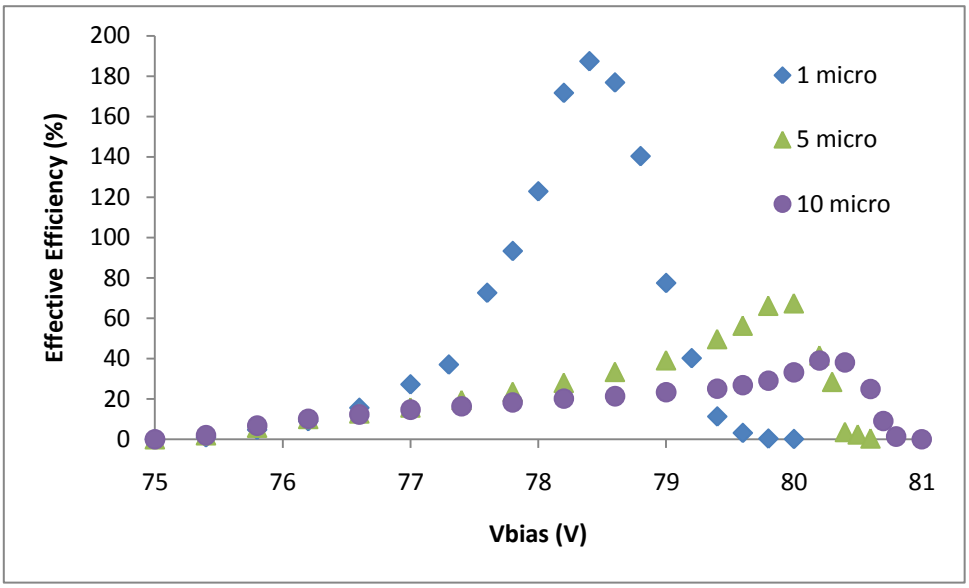
(a)


450,00
400,00
350,00
300,00
250,00
200,00
150,00
100,00
50,00
0,00
Quantum Efficiency (%)
1 micro
5 micro
10 micro
75
76
77
78
79
80
81
Vbias

(b)


30,00
25,00
20,00
15,00
10,00
5,00
0,00
Quantum Efficiency (%)
1 micro
5 micro
10 micro
75
75,2
75,4
75,6
75,8
76
76,2
76,4
76,6
76,8
77
Vbias

(c)



(d)

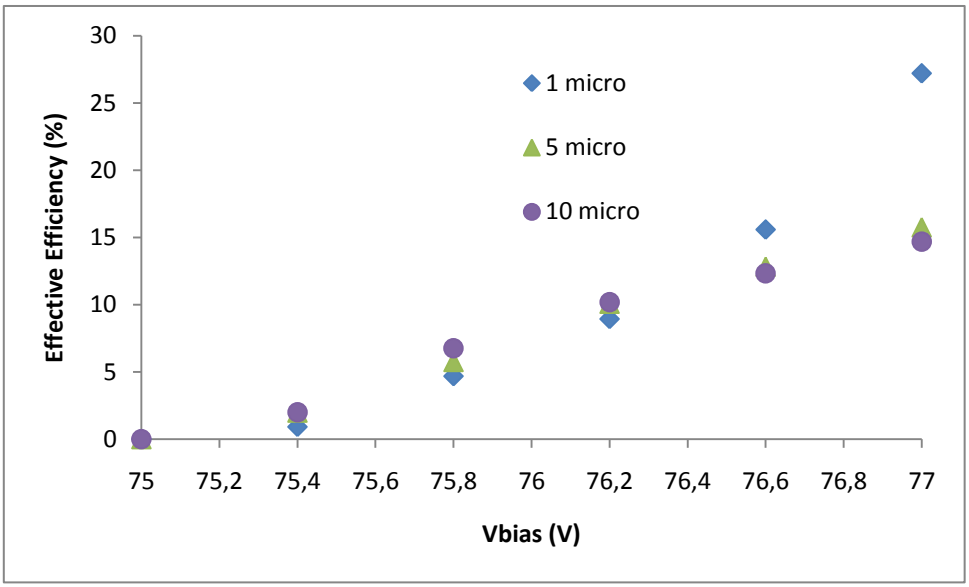
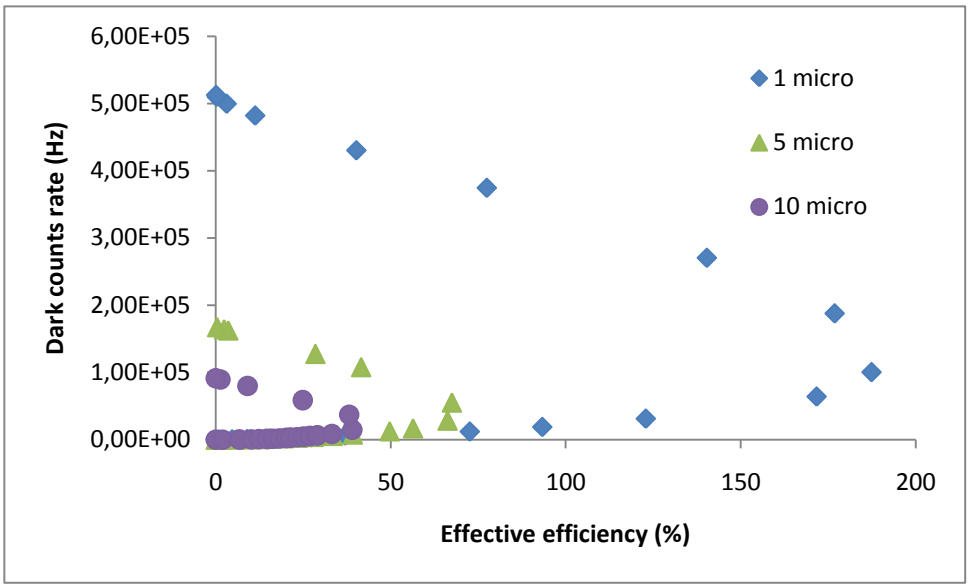


Figure A5.2. Evolution of quantum, (a) and (b), and effective, (c) and (d), efficiencies of the NFAD tested in free-running mode with 10^5 photons per second for four different values of the hold-off time.

Finally we can see in figure A5.3 the evolution of the dark counts frequency versus the effective efficiency, again for the four values of hold-off time for which the NFAD was tested and with a zoom of the first part of the graph.

(a)



(b)

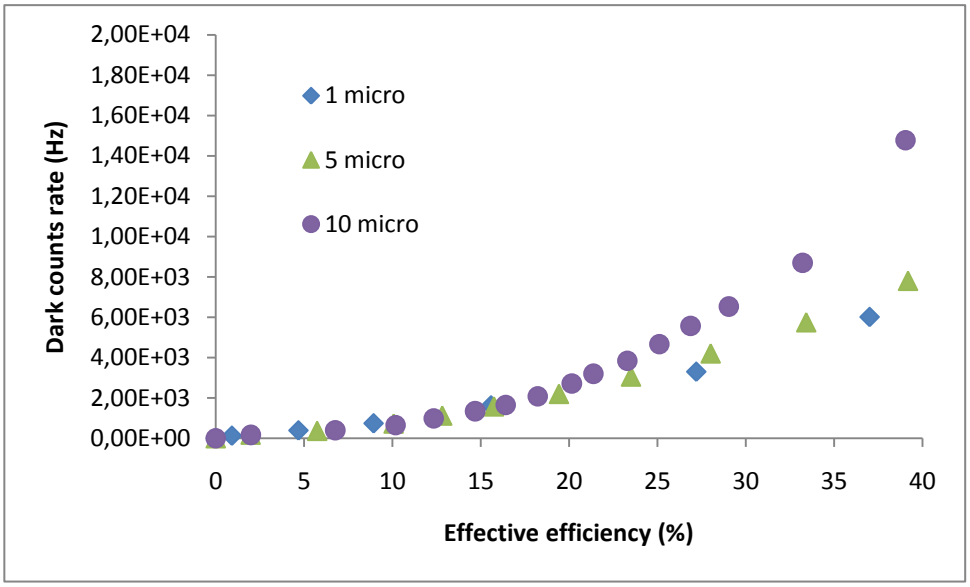


Figure A5.3. Free-running mode; dark counts frequency versus effective efficiency (a), with a zoom of the first part of the graph (b), when 10^5 photons per second are sent to the NFAD.

Tables 6: Hold-off time measures for 10^6 photons per second using different hold-off times.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75,2	1573,5	42	0,15315	0,15
75,4	6323,1	127,6	0,61955	0,62
75,8	28998	341,1	2,86569	2,95
76,2	70961,8	729,5	7,02323	7,57
76,6	105124,2	1405	10,37192	11,61
77	160319,8	3007,2	15,73126	18,79
77,4	224241	6426,5	21,78145	28,26
77,8	304782,3	17634,3	28,7148	42,04
78,2	386318,9	57107,9	32,9211	56,89
78,6	441119,6	171196,2	26,99234	58,27
78,8	459777,1	257724,8	20,20523	50,39
79	478996,8	374333,2	10,46636	32,11
79,2	494962,1	462080,8	3,28813	12,10
79,4	501935,7	489948,4	1,19873	4,72
80	507630,7	503821,6	0,38091	1,56

Table 6.1: Data obtained with a hold-off time of 1 µs.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75,2	4347,3	87,8	0,42595	0,43
75,4	15634,5	214,1	1,54204	1,55
75,8	57024,9	542,1	5,64828	5,82
76,2	124457,7	1397,4	12,30603	13,13
76,6	188736	2812,5	18,59235	20,56
77	306157,8	7729,9	29,84279	35,37
77,4	432617,5	24409,5	40,8208	52,73
77,8	543754	100806,4	44,29476	64,06
78	583714,1	210735,4	37,29787	58,87
78,2	621589,3	358482,7	26,31066	46,51
78,4	642462,9	487254,2	15,52087	30,23
78,6	660982,3	595500,6	6,54817	13,93
78,7	669988,5	638434,6	3,15539	6,97
78,8	675438,9	655241,2	2,01977	4,54
79	682764,2	674525,2	0,8239	1,89
79,2	690224,9	688342,2	0,18827	0,44
79,4	0	0	0	0,00

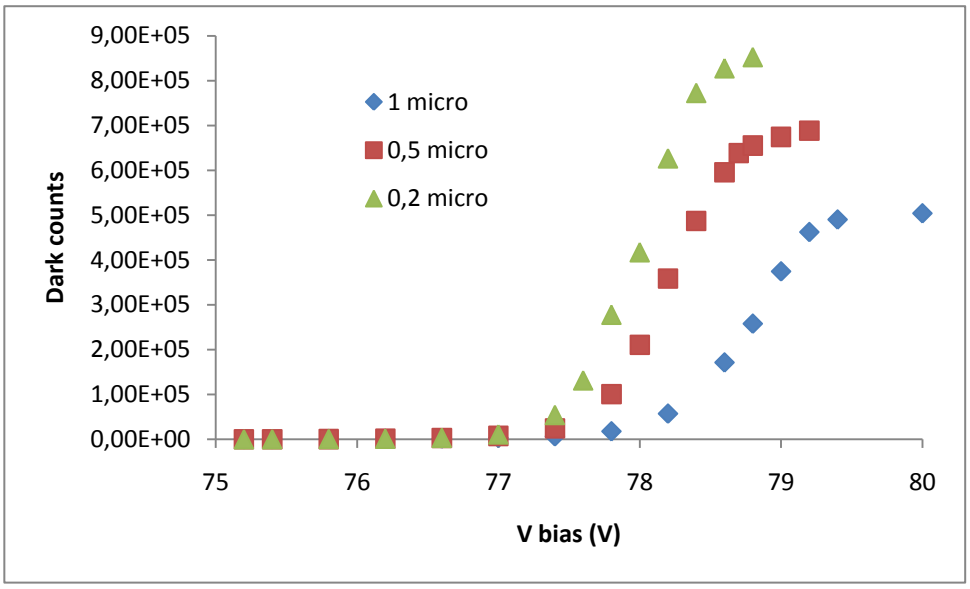
Table 6.2: data collected with a hold-off time of 0,5 µs.

Vbias (V)	Det. Freq(Hz)	Noise freq.(Hz)	Effective Efficiency(%)	Quantum Efficiency (%)
75,2	519,9	24,5	0,04954	0,05
75,4	9318,7	222	0,90967	0,91
75,8	41737	617,2	4,11198	4,15
76,2	81644,8	1223,2	8,04216	8,18
76,6	154773,9	3171,4	15,16025	15,65
77	310129,1	10838,9	29,92902	31,98
77,4	533376,9	54147,4	47,92295	54,23
77,6	648203,6	131339,5	51,68641	60,99
77,8	718837	277809,8	44,10272	54,54
78	756450,8	417161,3	33,92895	43,62
78,2	797328,5	626609,6	17,07189	23,22
78,4	827890,8	772610,1	5,52807	7,84
78,6	845413,9	827606,3	1,78076	2,57
78,8	857439,2	852536,3	0,49029	0,71
79	0	0	0	0,00

Table 6.3: data obtained with a hold-off time of 0,2 μs are presented.

Also from these series of data the figures extracted are shown, starting with the evolution of the dark counts versus the applied bias voltage. A zoom of the first part of the graph is later depicted.

(a)



(b)

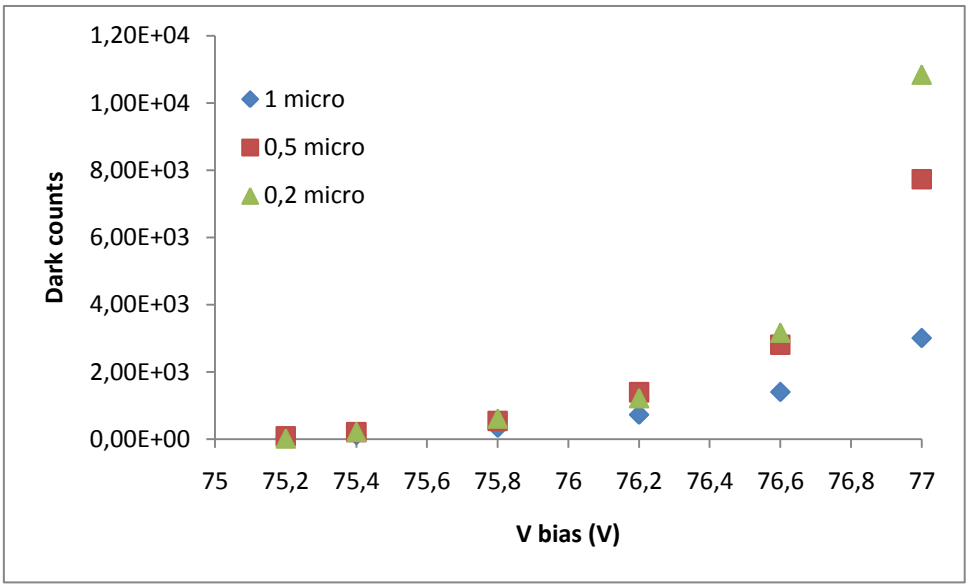
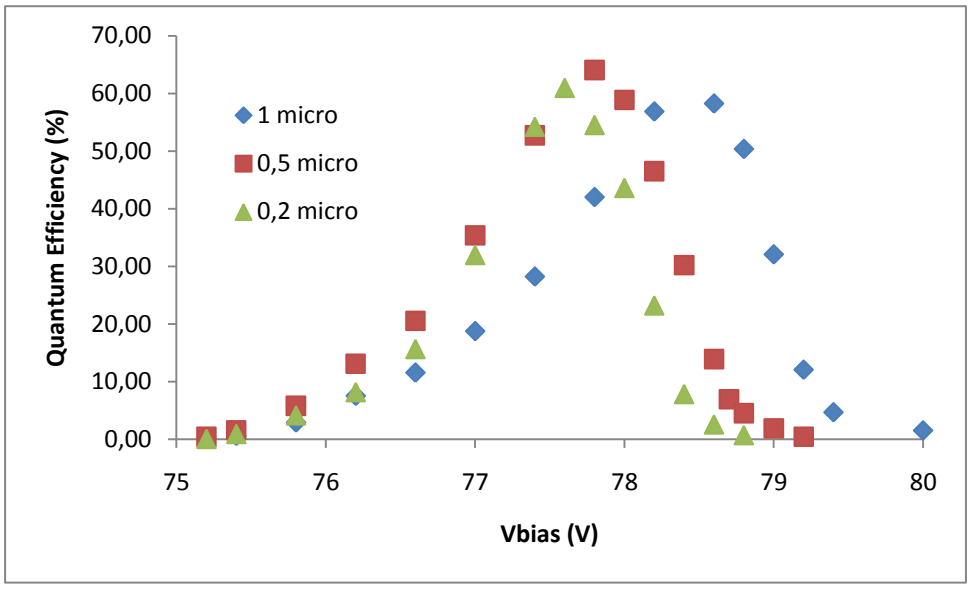
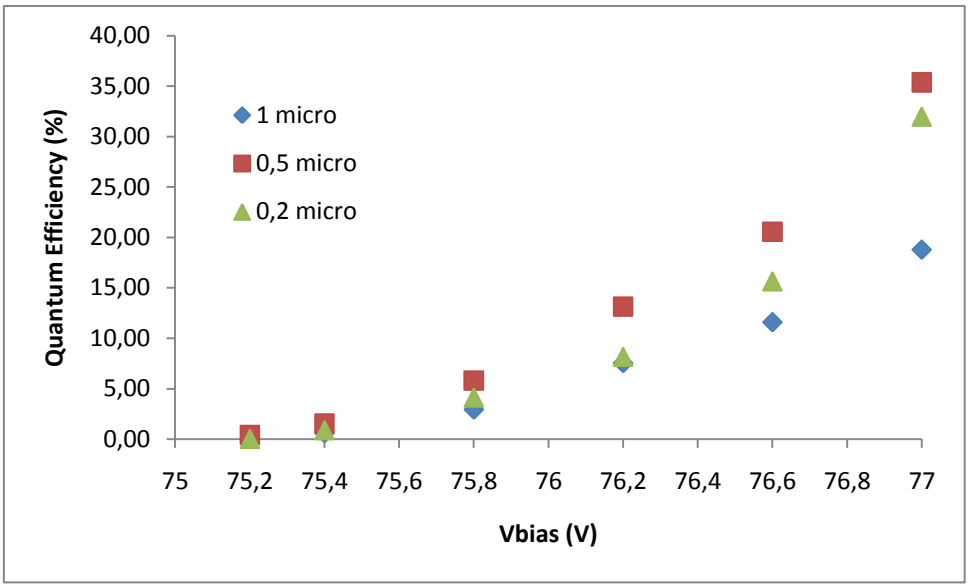


Figure A6.1. Noise characteristic of the NFAD tested in free-running mode with 10^4 photons per second for three different values of the hold-off time.

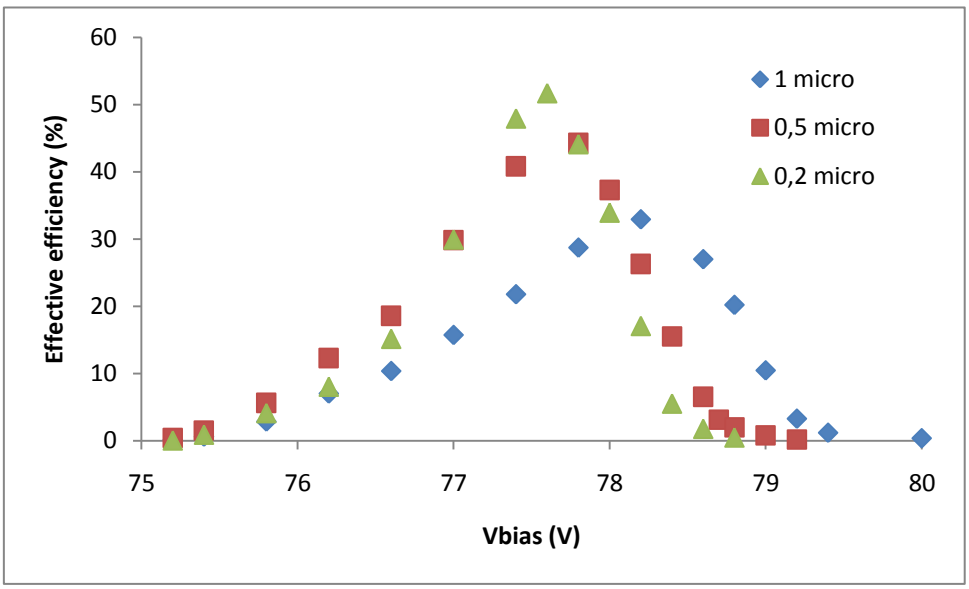
(a)


70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
Quantum Efficiency (%)
1 micro
0,5 micro
0,2 micro
75
76
77
78
79
80
Vbias (V)

(b)


40,00
35,00
30,00
25,00
20,00
15,00
10,00
5,00
0,00
Quantum Efficiency (%)
1 micro
0,5 micro
0,2 micro
75
75,2
75,4
75,6
75,8
76
76,2
76,4
76,6
76,8
77
Vbias (V)

(c)


60
50
40
30
20
10
0
Effective efficiency (%)
1 micro
0,5 micro
0,2 micro
75
76
77
78
79
80
Vbias (V)

(d)

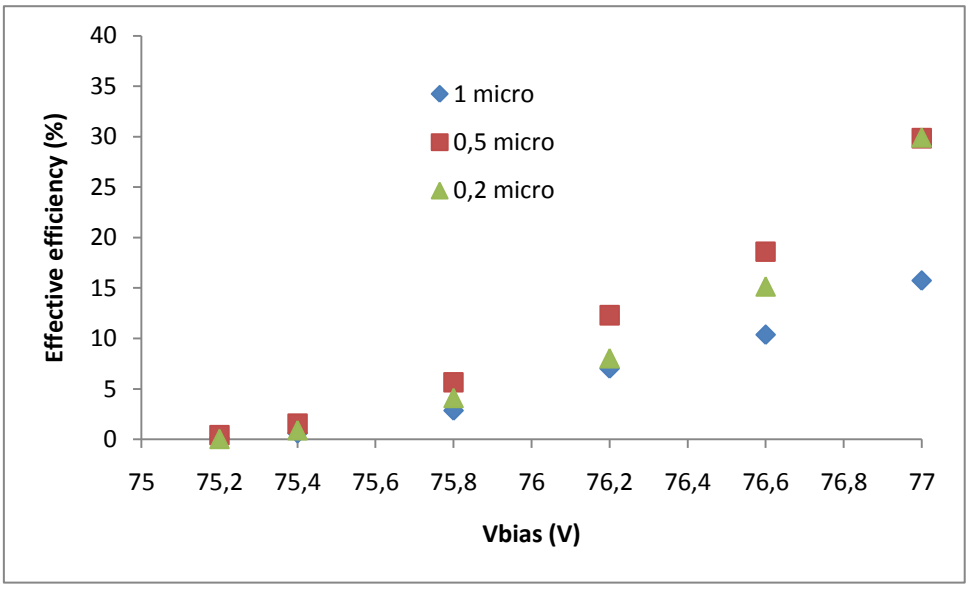
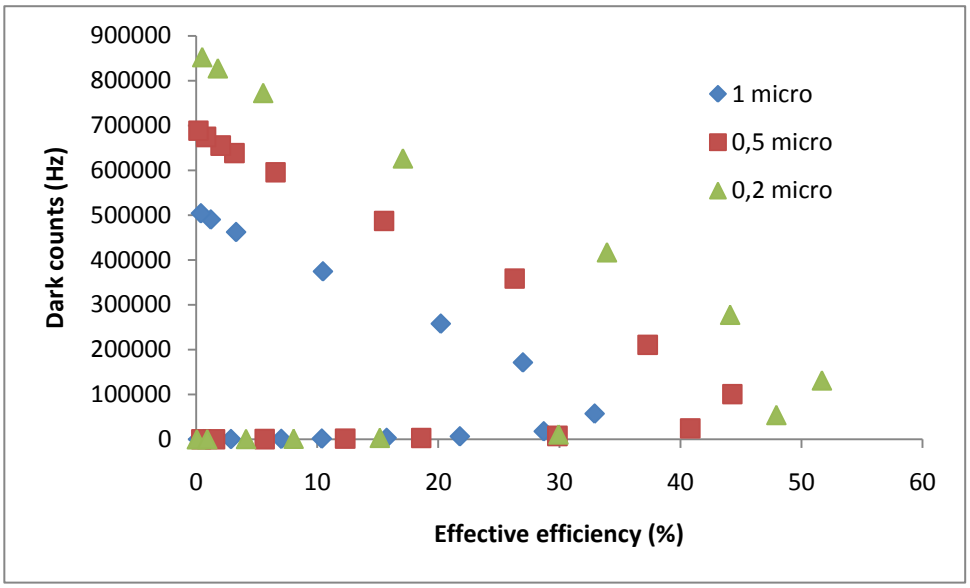


Figure A6.2. Evolution of quantum, (a) and (b), and effective, (c) and (d), efficiencies of the NFAD tested in free-running mode with 10^5 photons per second for four different values of the hold-off time.

Finally, in figure A6.3 the evolution of the dark counts frequency versus the effective efficiency is shown, for the three values of hold-off time for which the NFAD was tested now and with a zoom of the first part of the graph.

(a)



(b)

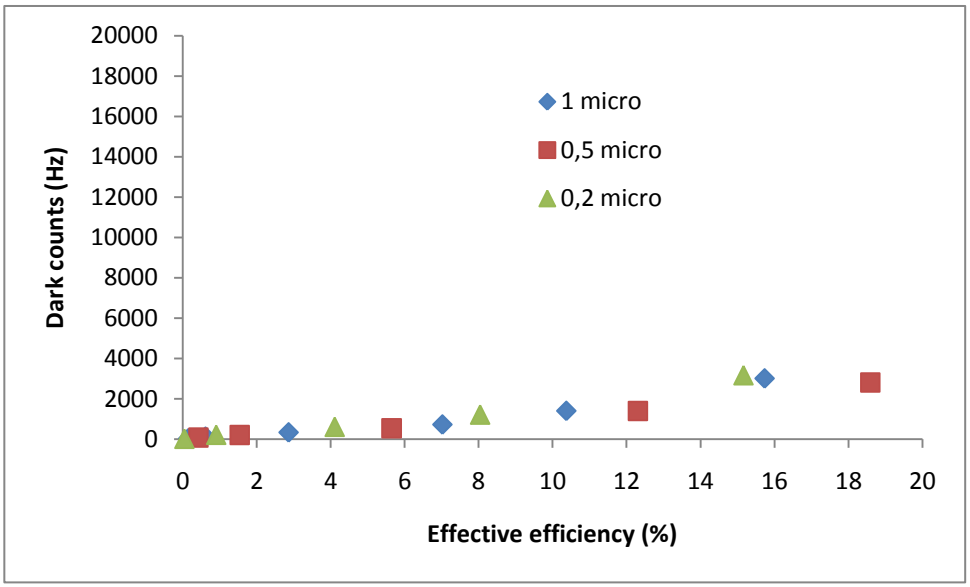


Figure A6.3. Free-running mode; dark counts frequency versus effective efficiency (a), with a zoom of the first part of the graph (b), when 10^6 photons per second are sent to the NFAD.

Bibliography

[1] N. Gisin, G. Ribordy, W. Tittel, H. Zbinden, « Quantum cryptography » ,Reviews on modern physics, Vol.74, 2002

[2] B.T. Levine, C.G. Bethea, and J.C. Campbell, « Room-temperature 1.3 µm optical time reflectometer using a photon counting InGaAs/InP avalanche detector", Appl. Phys. Let. 46, 333-335 (1985)

[3] B.K Garside, and R.E. Park, "A photon counting optical time domain reflectometer for distributed sensing applications", Proc. SPIE, Vol. 1169, p. 89-97 (1990)

[4] J.G. Timothy, "Photon-counting detector systems. Current status and future prospects", Proceedings of the 10th Symposium on Photoelectronic Image Devices, 85-96, London, 2-6 September 1991

[5] D.V. O'Conner and D. Phillips: "Time-correlated single photon counting", Academic, New York, (1988)

[6] W.J. Kindt, "Geiger Mode Avalanche Photodiode Arrays for spatially resolved single photon counting », Delft University Press, Delft, (1999)

[7] D. Contini, A. Pifferi, L. Spinelli, A.Torricelli, R. Cubeddu, F. Martelli, , G.Zaccanti, A. Dalla Mora, A.Tosi, F. Zappa, S. Cova, 'Functional diffuse reflectance spectroscopy at small source-detector distances based on fast-gated single-Photon avalanche diodes, in *Advanced Photon Counting Tech. IV,* edited by M.A. Itzler, J.C. Campbell, Proceedings of SPIE Vol. 7681 02

[8] P. Eraerds, "Advances in single-photon counting and fiber metrology", Doctoral Thesis, GAP-Optique de Genève

[9] B.F. Levine and C. G. Bethea, Appl. Phys. Lett. 46, 333 (1985)

[10] http://www.everyphotoncounts.com/physics-ingaas.php

[11] Cova, S., Tosi, A., Gulinatti, A., Zappa, F., and Ghioni, M., "Avalanche Diodes and Circuits for Infrared Photon Counting and Timing: Restrospect and Prospect,", IEEE LEOS Newsletter, 25-28 (October 2006).

[12] Y. Zhaoa, S. He, "Multiplication characteristics of InP/InGaAs avalanche photodiodes with a thicker charge layer" Optics Communications, Vol. 265, Iss. 2, pp. 476-480 (2006)

[13] Tosi, A. Dalla Mora, S. Tisa, F. Acerbi, F. Zappa, S. Cova, 'InGaAs/InP SPADs for near-infrared applications : device operating conditions and dedicated electronics', Proc. SPIE, Vol. 7681, 2010

[14] G. Ribordy, N Gisin, O. Guinnard, D. Stucki, M. Wegmuller and H. Zbinden, " Photon counting at telecom wavelengths with commercial InGaAs avalanche photodiodes: current performance", Journal of Modern Optics, vol. 51, no. 9-10, 1381-1398, 2004

[15] F. Zappa, A. Giulinatti, P. Maccagnani, S. Tisa, S. Cova, 'SPADA, Single-Photon Avalanche Diode Arrays ',IEEE photonics technology letters, vol. 17, no. 3, march 2005

[16] http://www.sensorsinc.com/GaAs.html from the website of the Goodrich ISR Systems – Princeton

[17] A. Tosi, A. dalla Mora, F. Zappa, S. Cova, M.A. Itzler, X. Jang, « InGaAs/InP Single-photon Avalanche Diodes show low dark counts and require moderate cooling », Proc. of

[18] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, “Avalanche photodiodes and quenching circuits for single-photon detection”, APPLIED OPTICS, Vol. 35, No. 12, 1996

[19] Gol’tsman, G. N., Okunev, O., Chulkova, G., Lipatov, A., Dzardanov, Smirnov, K., Semenov, A., Voronov, B., Williams, C. and Sobolewski, R., “Fabrication and properties of an ultrafast NbN hot-electron single-photon detector”, IEEE Trans. On Applied Superconductivity 11, 574-577, (2001)

[20] M. A. Itzler, Xudong Jiang, B. Nyman, K. Slomkowski, “InP-based Negative Feedback Avalanche Diodes”, Proc. Of SPIE, Vol. 7222, K1-K12, (2009)

[21] M.M. Hayat, D. A. Ramirez, G. J. Rees, M. A. Itzler “Modeling negative feedback in single photon avalanche diodes”, SPIE Defense, Security + Sensing Symposium 2010

[22] A. Gallivanoni, I. Rech, and M. Ghioni, “Progress in quenching circuits for Single Photon Avalanche Diodes”, IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 57, NO. 6, DECEMBER 2010

[23] R.H. Haitz, “Mechanism contributing to the noise pulse rate of avalanche diodes”, J. Appl. Phys. 36, 3123-3131 (1965).

[24] S. Cova, A. Longoni, and A. Andreoni, “Towards picoseconds resolution with single-photon avalanche diodes”, Rev. Sci. Instrum. 52, 408-412, (1981).

[25] S. Cova, M. Ghioni, A. Lotito, I. Rech and F. Zappa, “Evolution and prospects for single-photon avalanche diodes and quenching circuits”, Journal of Modern optics, vol. 51, no. 9-10, 1267-1288 (2004)

[26] A. Goetzberger, B. McDonald, R. H. Haitz, and R. M. Scarlett, “Avalanche effects in Silicon p-n Junctions. II. Structurally Perfect Junctions”, Journal of Applied Physics 34, 1591 (1963)

[27] R. H. Haitz, “Model for the Electrical behavior of a Microplasma”, Journal of Applied Physics 35, 1370 (1964)

[28] R.H. Haitz, « Mechanisms Contributing to the Noise Pulse Rate of Avalanche Diodes », ”, Journal of Applied Physics 36, 3123 (1965)

[29] R.J. McIntyre, “The distribution of gains in uniformly multiplying avalanche photodiodes: Theory,” IEEE Trans. Electron Devices **ED-19**, 703 (1972).

[30] R. P. Webb, R. J. McIntyre, and J. Conradi, “Properties of avalanche photodiodes,” *RCA Rev.*, vol. 35, pp.234-278, 1974

[31] F. Zappa, S. Cova, and M. Ghioni, “Monolithic circuit of active quenching and active reset for avalanche photodiodes”, US Patent: 6.541.752 B” (2003)

[32] Jie Zhang, “ Semicondutor optical single-photon detectors “, University of Rochester

[33] http://learn.hamamatsu.com/articles/photomultipliers.html

[34] http://www.idquantique.com/scientific-instrumentation/id210-ingaas-apd-single-photon-detector.html

[35] P. Antognetti, S. Cova, and A. Longoni, “A study of the operations and performances of an avalanche diode as a single photon detector”, in 2nd ISPRA Nuclear Electronic Symposium, May 20-23, 1975.

[36] S. Cova, A. Longoni, and G. Ripamonti, “Active- quenching and gating circuits for single photon avalanche diodes (SPADs)”, IEEE Trans. Nucl. Sci., vol. 29, pp. 599-601, Feb. 1982

[37] S. Cova, “Active Quenching Circuit for Avalanche Photodiodes”, U.S. Patent 4,963,727, Oct. 20, 1990, (Italian patent 22367 A/88)

[38] S. Komiyama, O. Astafiev, V. Antonov, T. Kutsuwa, and H. Hirai, “A single-photon detector in the far-infrared range,” *Nature*, vol. 403, pp. 405–407, Jan. 2000.

[39] R. T. Thew, D. Stucki, J.-D. Gautier,H. Zbinden, and A. Rochas, “Freerunning InGaAs/InP avalanche photodiode with active quenching for single photon counting at telecom wavelengths,” *Appl. Phys. Lett.*, vol. 91, no. 20, pp. 201114-1–201114-3, Nov. 2007.

[40] Personal correspondence with Mark Itzler of Princeton Lightwave

[41] W.G. Oldham, R.R. Samuelson and P. Antognetti, “Triggering phenomena in avalanche diodes”, IEEE Transactions on electron devices, vol. 19, no. 9, 1056-1060, 1972

[42] Itzler, Mark A. , Jiang, Xudong , Entwistle, Mark , Slomkowski, Krystyna , Tosi, Alberto , Acerbi, Fabio, Zappa, Franco and Cova, Sergio(2011) 'Advances in InGaAsP-based avalanche diode single photon detectors', Journal of Modern Optics, 58: 3, 174 — 200, First published on: 31 January 2011 (iFirst)

[43] S.Cova, A.Lacaita and G.Ripamonti, "Trapping phenomena in avalanche photodiodes on nanosecond scale", IEEE Electron.Dev.Lett. **12**, 685-687, (1991)

[44] G.Ripamonti, F.Zappa and S.Cova, "Effects of trap levels in single-photon optical time-domain reflectometry: evaluation and correction", IEEE J.Lightwave Technol. **10**, 1398-1402 (1992)

[45] Itzler, Mark A. , Jiang, Xudong , Entwistle, Mark , Slomkowski, Krystyna , Tosi, Alberto , Acerbi, Fabio, Zappa, Franco and Cova, Sergio(2011) 'Advances in InGaAsP-based avalanche diode single photon detectors', Journal of Modern Optics, 58: 3, 174 — 200, First published on: 31 January 2011

[46] Aull, B.F.; Loomis, A.H.; Young, D.J., Stern, A.; Felton, B.J.; Daniels, P.J.; Landers, D.J.; Retherford,L.; Rathman, D.D.; Heinrichs, R.M.; Marino, R.M.; Fouche, D.G.; Albota, M.A.; Hatch, R.E.; Rowe, G.S.; Kocher, D.G.; Mooney, J.G.; O'Brien, M.E.; Player, B.E.; Willard, B.C.; Liau, Z.-L.; Zayhowski, J.J. Proc. SPIE 2004, 5353, 105–116.

[47] Special issue on 'Free-space communication techniques for optical networks', IEEE LEOS Newsl. 2005, 19, 6–39.

[48] Knill, E.; Laflamme, R.; Milburn, G.J. "A scheme for efficient quantum computation with linear optics » , Nature 2001, 409, 46–52.

[49] Rarity, J.G.; Tapster, P.R., Experimental violation of Bell's inequality based on phase and momentum", Phys. Rev. Lett. 1990, 64, 2495–2498.

[50] R.T. Thew, D. Stucki, J.-D. Gautier, and H. Zbinden, A. Rochas, “Free-running InGaAs/InP avalanche photodiode with active quenching for single photon counting at telecom wavelengths”, APPLIED PHYSICS LETTERS 91, 201114 (2007)

[51] A. Bouzid, J.-B. Park, S.M. Kin, S. Kwak, S. Moon, “Characterisation of a single-photon detectorat 1.55 μm operated with an active hold-off technique for quantum key distribution”, Curr. App. Phys. 11, (2011), 903-908

[52] D. Stucki, C. Barreiro, S. Fasel, J.-D. Gautier, O. Gay, N. Gisin, R. Thew, Y. Thoma, P. Trinckler, F. Vannel, and H. Zbinden, “Continuous high speed coherent one-way quantum key distribution”, Opt. Express 17, 13326 (2009).

[53] Internal report prepared by Nino Walenta.

[54] J. Zhang, R. Thew, J.-D. Gautier, N. Gisin, and H. Zbinden, “Comprehensive Characterisation of InGaAs-InP Avalanche Photodiodes at 1550 nm With an Active Quenching ASIC”, IEEE Journal of Quantum Electronics, vol.45, no. 7, (2009)

[55] R. Alléaume, J. Bouda, C. Branciard, T. Debuisschert, M. Dianati, N. Gisin,M. Godfrey, P.Grangier, T. Länger,A. Leverrier, N. Lütkenhaus, P. Painchault, M. Peev, A. Poppe, T. Pornin, J. Rarity, R. Renner, G. Ribordy, M. Riguidel, L. Salvail, A. Shields, H. Weinfurter, A. Zeilinger, “SECOQC White Paper on Quantum Key Distribution and Criptography”, 2009

[56] www.idquantique.com, www.magiqtech.com

[57] http://www.secoqc.net

[58] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Du˘sek, N. Lutkenhaus, M. Peev, "The Security of Practical Quantum Key Distribution", REVIEWS OF MODERN PHYSICS, VOLUME 81, JULY–SEPTEMBER 2009

[59] Bennett, C. H. and Brassard, G., "Quantum cryptography: Public-key distribution and coin tossing", Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India, December 1984, pp. 175-179

[60] A. K. Ekert, « Quantum cryptography based on Bell's theorem », Phys. Rev. Lett. 67, 661-663, (1991)

[61] W.K.Wootters and W. H. Zurek, « A single quantum cannot be cloned », *Nature* **299,** 802-803 (1982)

[62] www.everyphotoncounts.com

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

[64] Zhang J., Eraerds P., Walenta N., Barreiro C., Thew R., Zbinden H., « 2.23 GHz gating InGaAs/InP single-photon avalanche diode for quantum key distribution », Proc. SPIE, Vol. 7681, 76810Z, (2010)

[65] Namekata N., Sasamori S., Inoue S., « 800 MHz single-photon detection at 1550 nm using an InGaAs/InPavalanche photodiode operated with a sine-wave gating », Opt. Express, Vol. 14, No. 21, (2006)

[66] Yuan Z. L., Kardynal B.E., Sharpe A. W., Shields A. J., « High speed single-photon detection in the near infrared », Appl. Phys Lett. 91, 041114, (2007)