JOURNAL OF MODERN OPTICS, 2003, VOL. 50, NO. 14, 2251–2269

High-rate quantum key distribution at short wavelength: performance analysis and evaluation of silicon single photon avalanche diodes

MASSIMO GHIONI, ANDREA GIUDICE, SERGIO COVA and FRANCO ZAPPA

Politecnico di Milano, Dipartimento di Elettronica e Informazione
Piazza Leonardo da Vinci 32, 20133 Milano – Italy
e-mail: ghioni@elet.polimi.it; tel: +39-02-23996093; fax: +39-02-2367604

(*Received 12 November 2002; final version 3 February 2003*)

Abstract. Experimental results obtained with silicon single photon avalanche diodes (SPADs) in quantum key distribution (QKD) at short wavelengths reveal remarkable potential for application in local area networks (LAN) and for free-space transmission at high rate. Actual application prospects, however, depend on the performance level and on the suitability of practical systems using the available silicon SPAD devices. They can be essentially divided in two groups: planar p-n junction structures with a thin depletion layer (typically 1 μm); and reach-through structures with a thick depletion layer (from 20 μm to 150 μm). The physical mechanisms that control the device behaviour were investigated and the effect on the key parameters of the detector (quantum detection efficiency, dark counting rate, afterpulsing probability and photon-timing jitter) were thoroughly assessed. A quantitative analysis was made of the influence of such parameters on the quantum bit error rate (QBER). Actual parameters were measured and the attainable performance and system suitability of the two device types evaluated. Comparable performance is obtained, but from a system viewpoint thin SPADs appear inherently better suited to high-rate QKD applications, because of their faster response time, ruggedness, low voltage, low power dissipation and fabrication technology, which is simple, efficient, economical and compatible with monolithic integration of detector and associated circuits.

1. Introduction

Recently, there has been growing interest in short operating wavelengths (0.77 to 0.85 μm) for free-space and short-distance fibre LAN applications, due to the lower cost of electro-optic components and assemblies. For these applications, quantum cryptography offers the unique possibility of secure key distribution between remote locations. Practical systems require efficient, low noise, single photon avalanche diodes (SPADs) to achieve this goal [1–4].

Silicon SPADs have been investigated extensively and are nowadays well developed; considerable progress has been achieved in design and fabrication techniques and devices with good characteristics are commercially available [5–7]. The devices so far reported can be divided into two groups, according to the depletion layer of the pn junction, which can be thin, typically 1 μm [8–11], or thick, from 20 μm to 150 μm [5–7, 9, 12–14]. Their main features can be summarized as follows. For planar SPADs with a thin depletion layer: (i) breakdown voltage V_B

Journal of Modern Optics ISSN 0950–0340 print/ISSN 1362–3044 online © 2003 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0950034031000113677

from 10 V to 50 V; (ii) small active area with diameter ranging from 5 μm to 100 μm; (iii) photon detection efficiency varying from 45% at 500 nm to about 10% at 820 nm; (iv) very high resolution in photon-timing, remarkably better than 100 ps full-width at half maximum (FWHM); in particular, devices with small active area (about 10 μm diameter) attain better than 30 ps FWHM at room temperature and better than 20 ps when cooled to −65°C [8]. For reach-through SPADs with thick depletion layer: (i) breakdown voltage V_B from 200 V to 500 V; (ii) fairly wide active area, with diameter ranging from 100 μm to 500 μm; (iii) photon detection efficiency better than 50% over all the range from 540 to 850 nm wavelength; (iv) fairly good resolution in photon-timing: better than 350 ps FWHM for reach-through types [7] and around 150 ps FWHM for devices having a smoother field profile [14].

Planar SPADs with thin depletion layer (called in the following thin SPADs) were developed by our research group in various generations of devices. The SPAD devices employed are described in detail in reference [10]; they were fabricated thanks to the support of ST Microelectronics (Cornaredo, Italy) and are presently not available commercially.

Reach-through SPADs with thick depletion layer (called in the following thick SPADs) were developed by the research team led by R.J.McIntyre and P.Webb at the former RCA Electrooptics Canada, now PerkinElmer Optoelectronics and they are commercially available in the PerkinElmer Single-Photon Counting Modules SPCM [6].

The purpose of this paper is to compare the performance of the available thick and thin SPAD devices in quantum key distribution (QKD) systems at short wavelengths. In section 2 the figure of merit of a QKD system is briefly discussed, in order to highlight the critical detector parameters. Section 3 gives an accurate description of the physical mechanisms that limit the performance of SPAD devices. Key parameters such as detection efficiency, dark counting rate, after-pulsing probability and timing jitter are considered and thoroughly discussed. These parameters have been experimentally measured for thin and thick SPAD devices and the results are shown in section 4. A comparison between the two types of SPAD devices is performed in section 5 from a system point of view.

2. Practical limits of quantum cryptography

In principle, QKD systems should be able to perform key distribution at a high rate without introducing errors. A quantitative measure of the system performance is therefore obtained from the bit rate and the bit-error rate values.

The raw bit rate, i.e. the number of exchanged bits per second after sifting but before any error correction, is given by [15]:

$$R_{\mathrm{raw}} = q_{\mathrm{pr}} \mu \nu \eta_{\mathrm{t}} \eta_{\mathrm{d}}, \tag{1}$$

where μ is the average photon number per pulse at the transmitter end, η_{t} is the transfer efficiency between the sender (Alice) and the receiver (Bob), η_{d} is the photon detection efficiency of the SPAD and ν is the pulse repetition frequency, q_{pr} is a systematic factor that depends on the chosen encoding protocol. If, for instance, the BB84 four-state protocol [16] is used, $q_{\mathrm{pr}} = 1/2$ since half of the time the bases chosen randomly by Alice and Bob are not compatible.

In ideal QKD systems errors only arise from tampering of an eavesdropper. Actually, the intrinsic non-idealities of the optical components and the photodetectors are responsible for some discrepancies between the transmitted and the received bit sequence. The bit-error rate or, more precisely, the quantum bit error rate (QBER), is generally defined as the ratio of wrong bits to the total amount of received bits, which is equivalent to the probability of getting a false detection to the total probability of a detection per pulse [15]:

$$QBER = \frac{p_{\text{opt}} p_{\text{phot}} + p_{\text{dark}}}{p_{\text{phot}} + 2p_{\text{dark}}} \cong p_{\text{opt}} + \frac{p_{\text{dark}}}{p_{\text{phot}}} \equiv QBER_{\text{opt}} + QBER_{\text{det}}, \tag{2}$$

where p_{dark} and p_{phot} are, respectively, the probabilities of getting a dark count and a photon count and p_{opt} is the probability that a photon is detected by the wrong detector, due to the limited interference fringe visibility or due to poor polarization alignment. Equation (2) holds for a system implementing the BB84 protocol and it has to be slightly adapted when other protocols are used.

QBER is the sum of two terms: $QBER_{\text{opt}}$ and $QBER_{\text{det}}$. In practical systems, $QBER_{\text{opt}}$ can be easily reduced below 1% [15].

The probability of getting a photon count is given by:

$$p_{\text{phot}} = \mu \eta_{\text{t}} \eta_{\text{d}} \tag{3}$$

and

$$p_{\text{dark}} = n_{\text{dark}} \Delta t, \tag{4}$$

where n_{dark} is the SPAD dark counting rate (dark counts per second) and Δt is the detection time window. We therefore obtain:

$$QBER_{\text{det}} = \frac{n_{\text{dark}}}{\eta_{\text{d}}} \frac{\Delta t}{\mu \eta_{\text{t}}}. \tag{5}$$

$QBER_{\text{det}}$ is inversely proportional to the system's transmission efficiency. Therefore the transmission distance is limited by the detector performance. Equation (5) shows that the photodetector figure of merit in QKD applications is the ratio between the dark counts in the detection time window ($n_{\text{dark}}\Delta t$) and the photon detection efficiency, rather than the noise equivalent power (NEP) used in most other applications. Anyway, the behaviour of the two figures of merit is quite similar.

In principle, $QBER_{\text{det}}$ can be reduced by using a small time window $\triangle t$ around the expected photon arrival time, but, as will be shown in the next section, the SPAD timing jitter sets an ultimate limit to the width of this window.

Despite errors and potential eavesdropping, a certifiably secret key can be distilled by means of error correction [17] and privacy amplification [18] procedures. However, the number of bits that have to be discarded during these procedures increases with the *QBER*. In practice, the distillation of a secure key becomes very inefficient as *QBER* approaches $\sim 15\%$ [19, 20].

3. Key detector features

According to equation (5), the ideal SPAD for quantum cryptography applications would feature 100% photon detection efficiency, zero dark counting rate and low timing jitter. Though such requirements cannot be fulfilled, they can be fairly well approached by good SPAD devices coupled to suitably designed quenching circuits [21].

In the following we will briefly discuss the physical mechanisms that limit the performance of SPAD devices in QKD systems.

3.1 *Photon detection efficiency*

The photon detection efficiency is the product of the quantum efficiency and the avalanche triggering probability η_{at}, i.e. the probability that a primary electron–hole pair initiates a self-sustaining avalanche process. The quantum efficiency depends only on the photodetector structure and on the presence of a properly designed anti-reflection coating. Conversely, the avalanche triggering probability depends on both the SPAD structure and the excess bias voltage V_E, i.e. the difference between the bias voltage and the breakdown voltage. η_{at} increases with V_E and it can be calculated, provided the electric field profile and the ionization coefficients for electrons and holes are known. The triggering probability can be approximated by [5]:

$$\eta_{at} = 1 - e^{-V_E/V_C}, \tag{6}$$

where the characteristic voltage V_C depends on the depletion layer thickness and on k_{eff}, which is a weighted average of the ratio of the electron ionization coefficient to the hole ionization coefficient [5]. V_C typically varies from few volts for SPADs with thin depletion layer, to 16 V for SPADs with thick (20–35 μm) depletion region [5]. Accordingly, the quenching circuit should be capable of handling excess bias voltages up to 20–30 V, in order to achieve a η_{at} value approaching unity.

3.2 *Dark-counting rate: thermal generation and afterpulsing*

3.2.1 *Thermal generation*

Thermal generation effects produce current pulses even in absence of illumination, and the Poissonian fluctuation of these dark counts represents the internal noise source of the detector. The SPAD dark counting rate increases with the excess bias voltage.

The dark counting rate includes primary and secondary pulses [22, 23]. Primary dark pulses are due to carriers thermally generated within the SPAD junction, so that the counting rate increases with the temperature just as the dark current in ordinary photodiodes. The rate also increases with V_E due to two effects, namely (i) field assisted enhancement of the emission rate from generation centres and (ii) increase of the avalanche triggering probability.

3.2.2 *Afterpulsing*

Afterpulsing effects produce secondary dark pulses, and may strongly enhance the total dark counting rate. During the avalanche some carriers are captured by deep levels in the depletion layer and then released with a statistically fluctuating delay, whose mean value depends on the deep levels actually involved [22, 23].

In order to take the afterpulsing effects into account $QBER_{\mathrm{det}}$ must be modified as follows:

$$QBER_{\mathrm{det}} = \frac{p_{\mathrm{dark}} + p_{\mathrm{ap}}}{p_{\mathrm{phot}} + 2(p_{\mathrm{dark}} + p_{\mathrm{ap}})} \tag{7}$$

where p_{ap} is the probability of getting a secondary dark pulse.

The value of p_{ap} depends on whether the SPAD operates in gated mode or in free-running condition with an active quenching circuit (AQC) [21].

Afterpulsing in gated operation The SPAD device is gated-on for a short time period T_{g} around the expected arrival time of the photon nT, where n is a generic integer and T is the pulse repetition period. In this case, the probability of observing an afterpulse in the time period T_{g} centred around nT is given by the sum of the probabilities that an avalanche current pulse has been triggered in the time periods T_{g} around $(n-1)T$, $(n-2)T$ and so forth, multiplied by the probability that the carriers trapped in these events are released at the time nT causing a correlated avalanche pulse:

$$p_{\mathrm{ap}} = \left(p_{\mathrm{phot}} + p_{\mathrm{dark}} + p_{\mathrm{ap}}\right) \sum_{1}^{\infty} {}_{n} \int_{nT}^{nT+T_{\mathrm{g}}} P_{\mathrm{a}}(t)\mathrm{d}t, \tag{8}$$

where $P_{\mathrm{a}}(t)\mathrm{d}t$ is the probability of detecting an afterpulse in the time interval between t and $t+\mathrm{d}t$ after an avalanche current pulse. We have made two assumptions in the derivation of equation (8): (i) the probability of coincident events is negligible, and (ii) the trap population does not saturate. Both these assumptions are correct if p_{ap}, p_{phot} and p_{dark} are sufficiently small (a few per cent).

Rearranging equation (8) we obtain:

$$\begin{aligned} p_{\mathrm{ap}} &= \frac{\left(p_{\mathrm{phot}} + p_{\mathrm{dark}}\right) \sum_{1}^{\infty} {}_{n} \int_{nT}^{nT+T_{\mathrm{g}}} P_{\mathrm{a}}(t)\mathrm{d}t}{1 - \sum_{1}^{\infty} {}_{n} \int_{nT}^{nT+T_{\mathrm{g}}} P_{\mathrm{a}}(t)\mathrm{d}t} \\ &= \frac{\left(p_{\mathrm{phot}} + p_{\mathrm{dark}}\right) p_{\mathrm{ap_T}}\left(T, T_{\mathrm{g}}\right)}{1 - p_{\mathrm{ap_T}}(T, T_{\mathrm{g}})}, \end{aligned} \tag{9}$$

where

$$p_{\mathrm{ap_T}}\left(T, T_{\mathrm{g}}\right) = \sum_{1}^{\infty} {}_{n} \int_{nT}^{nT+T_{\mathrm{g}}} P_{\mathrm{a}}(t)\mathrm{d}t \tag{10}$$

is the total probability of detecting a secondary pulse after the triggering of an avalanche with the SPAD operated in gated mode.

By substituting equation (9) into equation (7) and rearranging, we obtain:

$$QBER_{\mathrm{det}} = \frac{p_{\mathrm{dark}} + p_{\mathrm{phot}} p_{\mathrm{ap_T}}\left(T, T_{\mathrm{g}}\right)}{p_{\mathrm{phot}}\left(1 + p_{\mathrm{ap_T}}(T, T_{\mathrm{g}})\right) + 2p_{\mathrm{dark}}}. \tag{11}$$

The value of $p_{\mathrm{ap_T}}(T, T_{\mathrm{g}})$depends on the lifetime of the trap levels and on the gate-off interval $T(T \gg T_{\mathrm{g}})$. Considering, for example, k different trap levels with

lifetime τ_j, we have:

$$P_a(t) = \sum_1^k {}_j A_j e^{-(t/\tau_j)}, \tag{12}$$

where

$$\sum_1^k {}_j A_j \tau_j = \int_0^{+\infty} P_a(t) dt$$

is the total probability of observing an afterpulse after the triggering of an avalanche.

By substituting equation (12) into equation (10) we finally get:

$$p_{ap_T}(T, T_g) = \sum_1^k {}_j \frac{A_j \tau_j \left(1 - e^{-(T_g/\tau_j)}\right) e^{-(T/\tau_j)}}{\left(1 - e^{-(T/\tau_j)}\right)}. \tag{13}$$

If $T \gg \tau_j$ all carriers are detrapped in the time interval between two adjacent gate pulses and, as expected, $p_{ap_T}(T, T_g)$ becomes negligible. In this case, no afterpulses are detectable, and $QBER_{det}$ is limited by the dark counting rate. On the other hand, if $p_{ap_T}(T, T_g)$ is finite and p_{dark} is negligible with respect to p_{phot}, we get

$$QBER_{det} \cong \frac{p_{dark}}{p_{phot}} + p_{ap_T}(T, T_g) = QBER_{det,dark} + QBER_{det,ap}, \tag{14}$$

where $QBER_{det,dark}$ and $QBER_{det,ap}$ are the contributions to $QBER_{det}$ arising from the dark counts and the afterpulses.

In some experimental set-ups, a shorter detection window with amplitude Δt, centred within T_g is used to further reduce the dark count and the afterpulse probabilities. In this case equation (14) can be simply modified as follows:

$$QBER_{det} \cong \frac{p_{dark}}{p_{phot}} + p_{ap_T}(T, \Delta t) = QBER_{det,dark} + QBER_{det,ap}, \tag{15}$$

where

$$p_{ap_T}(T, \Delta t) = \sum_1^\infty {}_n \int_{nT}^{nT+\Delta t} P_a(t) dt. \tag{16}$$

By looking at equation (13) we note that $p_{ap_T}(T, T_g)$ can be reduced by reducing the factors A_j. These factors deserve some further comments. By performing a simple carrier balance consideration, it turns out that

$$A_j \tau_j = n P_{cj} \eta_{at}, \tag{17}$$

where n is the average number of electron–hole pairs that flow through the junction during the avalanche, P_{cj} is the probability that a minority carrier (electron for n-p SPAD junction) gets trapped in the jth level and η_{at} is the triggering probability. The factors A_j may be reduced either by employing carefully designed fabrication processes, which minimise the density of trap levels (and thus P_{cj}), or by reducing the charge flowing through the junction during the avalanche pulse. For a SPAD operating in gated mode,

$$n = \frac{V_E}{R_d q} T_g, \tag{18}$$

where V_E is the excess bias voltage, R_d is the series resistance of the avalanching junction and q the electron charge. Since the value of V_E is usually dictated by the requirement of high photon detection efficiency, in the first instance the trapped charge per pulse has to be minimized by shortening the gate-on time T_g.

Afterpulsing in free running operation In this operation mode, an AQC is used to quench the current as soon as the avalanche pulse is detected. Quenching is performed by actively lowering the SPAD bias voltage below the breakdown level [21] for a well-defined hold-off time. After that, the bias voltage is swiftly restored in order to detect another photon. The minimum time interval between the detection of two photons is called dead time T_d: this time is typically a few tens of nanoseconds in the state of the art AQCs [21].

In free running, the probability of detecting an afterpulse p_{ap} may be calculated by using the same considerations made in the preceding subsection. Suppose we evaluate p_{ap} in a detection time window Δt centred around the generic time instant nT where a photon is expected to arrive. As happens in the gated mode, photons may strike the detector only at discrete time instants $(n\text{-}i)T$ preceding nT ($T > T_d$). However, in free-running mode both dark counts and afterpulses may happen at any time t preceding nT (except for a short interval T_d, where the SPAD detector is insensitive). Therefore equation (8) must be modified as follows:

$$p_{ap} = (p_{dark} + p_{ap}) \int_{T_d}^{+\infty} P_a(t)\mathrm{d}t + p_{phot} \sum_{1}^{\infty} n \int_{nT}^{nT+\Delta t} P_a(t)\mathrm{d}t, \tag{19}$$

which reduces to

$$\begin{aligned} p_{ap} &= \frac{p_{phot} \sum_{1}^{\infty} n \int_{nT}^{nT+\Delta t} P_a(t)\mathrm{d}t + p_{dark} \int_{T_d}^{+\infty} P_a(t)\mathrm{d}t}{1 - \int_{T_d}^{+\infty} P_a(t)\mathrm{d}t} \\ &= \frac{p_{phot} p_{ap_T}(T,\Delta t) + p_{dark} p_{ap_T}(T_d)}{1 - p_{ap_T}(T_d)}, \end{aligned} \tag{20}$$

where

$$p_{ap_T}(T_d) = \int_{T_d}^{+\infty} P_a(t)\mathrm{d}t. \tag{21}$$

$p_{ap_T}(T_d)$is the total probability of detecting a secondary pulse after the triggering of an avalanche with the SPAD operated in free running mode.

By substituting equation (20) into equation (7) we obtain:

$$QBER_{det} = \frac{p_{dark} + p_{phot} p_{ap_T}(T,\Delta t)}{p_{phot}(1 - p_{ap_T}(T_d) + 2p_{ap_T}(T,\Delta t)) + 2p_{dark}}, \tag{22}$$

which reduces to

$$QBER_{det} \approx \frac{p_{dark}}{p_{phot}} + p_{ap_T}(T,\Delta t) = QBER_{det,dark} + QBER_{det,ap} \tag{23}$$

if both $p_{ap_T}(T_d)$ and $p_{ap_T}(T,\Delta t)$are much smaller than unity and p_{dark} is negligible with respect to p_{phot}. These conditions are usually achieved in well-designed systems.

By comparing equations (14) and (23), we therefore conclude that $QBER_{det}$ for free running operation and gated operation are equivalent, assuming that $\Delta t = T_g$ and $T > T_d$.

When the pulse repetition period T becomes shorter than the dead time T_d, the probability $p_{ap_T}(T,\Delta t)$ must be modified as follows:

$$p_{ap_T}(T,\Delta t,T_d) = \sum_{h}^{\infty} {}_n \int_{nT}^{nT+\Delta t} P_a(t)\mathrm{d}t, \tag{24}$$

where $hT = T_d$.

If k different trap levels with time constants τ_j are considered, equation (24) can be written as:

$$p_{ap_T}(T,\Delta t,T_d) = \sum_{1}^{k} {}_j A_j \tau_j \mathrm{e}^{-(T_d/\tau_j)} \frac{\left(1-\mathrm{e}^{-(\Delta t/\tau_j)}\right)}{\left(1-\mathrm{e}^{-(T/\tau_j)}\right)}. \tag{25}$$

By comparing equations (25) and (13) we note the free running mode has advantage over gated mode when $T < T_d$.

In free running mode, care must be taken when evaluating the probability of detecting a photon in the time window Δt, since this probability is conditioned by the probability of not having a dark pulse or an afterpulse in the time interval T_d preceding Δt. Therefore:

$$p_{phot} = \mu\eta_t\eta_d(1-p_d(T_d))(1-p_{ap}(T_d)) \approx \mu\eta_t\eta_d(1-p_d(T_d)-p_{ap}(T_d)), \tag{26}$$

where $p_d(T_d)$ and $p_{ap}(T_d)$ are the probabilities of having a dark pulse and an afterpulse in the time interval T_d. These probabilities can be simply evaluated by using equations (4) and (20). As a matter of fact, they are negligible in a well-designed system. Moreover, at high pulse repetition rates p_{phot} is reduced due to the active quenching circuit dead time, and it must be corrected as follows [24]:

$$p_{phot} = \mu\eta_t\eta_d(1-\mu\eta_t\eta_d\nu T_d). \tag{27}$$

As in the case of gated mode operation, the afterpulsing effects in free running mode may be minimized by a suitable fabrication technology that ensures direct control of the density of trapping centres and/or their lifetime. If the trapped charge cannot be reduced to sufficiently low levels in this way, the only alternative is to reduce the charge flowing through the junction during the avalanche. It has been shown that in most cases this task is best accomplished by using a mixed active–passive quenching circuit [21], leading to:

$$n = \frac{C_T V_E}{q}, \tag{28}$$

where C_T is total capacitance (junction + stray) across the SPAD terminals.

A strong reduction of C_T is thus mandatory in free-running operation mode. Monolithic integration of the AQC is advantageous from this standpoint [25]. In a hybrid sensor where the SPAD is mounted chip-on-chip on the integrated AQC, stray capacitances below 1 pF can easily be achieved.

3.3 *Thermal effects*

Self–heating of the SPAD may seriously degrade the detector performance. The avalanche current dissipates considerable energy in the device, particularly in thick devices; since they operate with high bias voltage, from 300 to 400 V, they may dissipate more than 5 W during the avalanche pulse. The thermal resistance from the diode junction to the heat sink strongly depends on the type of mounting (packaged device, chip on carrier, etc.) and it may range from less than 0.1°C/mW to 1°C/mW. At high counting rate, the mean power dissipation causes a significant temperature increase, particularly in SPADs with high breakdown voltage V_B. The breakdown voltage strongly depends on the junction temperature; its thermal coefficient depends on the SPAD device structure and is fairly high, typically around 0.2–0.3% per °C [5, 22]. At a constant supply voltage V_A, the increase of V_B with temperature causes a decrease of the excess bias voltage V_E. The percentage variation of V_E with temperature is greater than that of V_B by a factor V_B/V_E [21], being very strong at low V_E level, about 30% per degree kelvin, and fairly high also at high V_E level, about 3% per degree kelvin. The reduction of V_E gives rise to degradation of the photon detection efficiency and of the time resolution. The degradation of detector performance is remarkable when the photodiode chip is not mounted on an efficient heat sink and the mean counting rate pulses varies [5]. It is therefore important to accurately stabilize the junction temperature in working conditions.

3.4 *Time resolution*

The time resolution curve of a SPAD is given by the statistical distribution of the delay from the photon arrival time to the detection time. It can be measured in a time correlated photon counting (TCPC) set-up by using a source of picosecond optical pulses [26]. Time resolution is usually quoted as the full width at half maximum (FWHM) of the time resolution curve. In quantum cryptography, time resolution is an important parameter for two reasons. First, it is of the utmost importance to avoid a photon being incorrectly registered in an adjacent time slot. If the time resolution curve is truly Gaussian, about 98% of photon counts occur within a time interval of 2FWHM centred around the peak; it follows that the time resolution FWHM should be sufficiently shorter than half of the pulse period. The second and more stringent requirement is that the time resolution FWHM should be smaller than half of the detection time width Δt, in order to avoid photon losses. The QBER reduction obtainable by limiting Δt (see equation (3)) is thus limited by the detector time resolution.

It has been ascertained by experiments [7–11, 14, 21] and explained by theoretical studies [27] that the resolution in single photon timing improves at higher electric field, hence at higher V_E.

4. Experimental tests

Experimental tests were carried out by operating the SPADs in optimized conditions, i.e. such as to achieve the best compromise between quantum efficiency, dark counting rate, afterpulsing effects, and time resolution. A compact AQC was specifically designed to this purpose. This AQC makes it possible to obtain quenching pulses up to 25 V amplitude and fast active reset of the detector. An overall dead time of about 36 ns was attained in operation with 20 V excess bias

voltage above breakdown level of the photodiode, that is, a saturated counting rate exceeding 25 Mcounts/s. The avalanche pulse charge was minimized by means of a mixed passive-active quenching approach [21], thus reducing self-heating and afterpulsing effects in the photodiode.

The thick SPAD tested (mod. Slik^{TM}) were available by courtesy of PerkinElmer-Canada. These devices are not sold separately as components for building instrumentation, but they are commercially available in compact modules (SPCM) including the SPAD and the electronic circuitry. Originally designed as analogue amplifying avalanche photodiodes with very low multiplication noise, these devices have a smooth field profile in the p-n junction, intended to minimize the value of k_{eff} for a given thickness; hence the name Slik^{TM}, which stands for ''super-low k'' [5, 6]. Due to this field profile and to the improved fabrication technology, which includes very effective gettering steps, the performance in Geiger mode is greatly improved compared to previous types, such as the C30902S and C30921S. The Slik^{TM} device actually employed in the tests had a breakdown voltage of 389 V at room temperature, active area diameter of 250 μm and was mounted on a double-stage thermoelectric cooler (Marlow TE2121), sealed in a TO-8 package. The operating temperature was monitored by using a thermistor and maintained within 0.1°C of the preset value by a proportional-integral regulator (Marlow SE5020). This ensures stable performance, despite changes in the ambient temperature. Measurements were performed at 20 V excess bias voltage.

The thin SPAD employed in the tests had an active area diameter of 10 μm and a breakdown voltage of 17 V [10]. The device was operated at 5 V excess bias voltage at room temperature. Thanks to their low operating voltage, the power dissipation of thin SPADs is negligible even at high counting rates.

4.1 *Photon detection efficiency*

Figure 1 shows the measured photon detection efficiency of the SPADs as a function of wavelength.

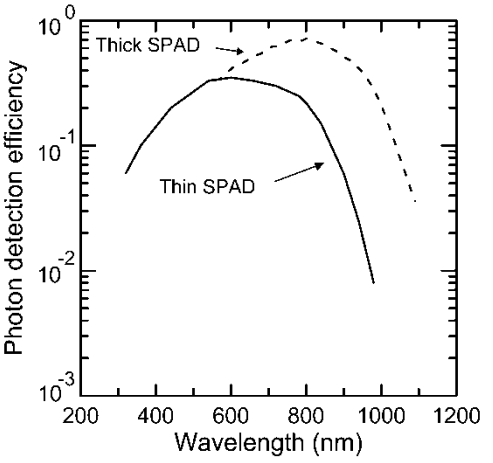

Figure 1. Photon detection efficiency of thick and thin SPAD as a function of wavelength.

Over most of the spectral range covered with silicon detectors, the Slik^{TM} efficiency was significantly better than that of thin SPAD. A peak efficiency of about 70% was measured for the Slik^{TM} at 820 nm, a value consistent with the fact that an antireflection (AR) coating centred at this wavelength is deposited on the silicon device, while no AR coating is present on the glass window of the TO-8 case. It is worth mentioning that by using different AR coatings [5] a detection efficiency in excess of 80% at 500 nm can be attained. The peak efficiency for the thin SPAD was about 30% at 633 nm, without any antireflection coating on the device.

4.2 *Dark counts*

Figure 2 shows the measured dark counting rate of the SPADs as a function of the temperature. For the Slik^{TM}, the total dark counting rate was about 17 kcounts/s at room temperature and was strongly reduced (down to a hundreds of counts/s) by lowering the detector temperature to $-20°C$. As expected, dark counts approximately halved for every 10°C decrease in temperature. Conversely, dark counting rate for the thin SPAD was less sensitive to temperature: it varied from about five hundred counts at room temperature to two hundred counts at $-20°C$. This effect is due to the high electric field in the thin depletion layer, which causes a reduction of the potential barrier seen by the electrons trapped in the generation centres, thus making the electron emission less sensitive to temperature.

A comparison of the experimental data clearly shows that from the standpoint of dark-counting rate the Slik^{TM} performance is better than that of thin SPAD. In fact, the comparison must be made taking into account that the depleted silicon volume of thin SPADs is more than three orders of magnitude smaller than that of Sliks^{TM}.

This point deserves some comment. It is well known that in silicon processing technology efficient gettering processes have been developed, which minimize both the concentration of generation centres, responsible for the primary dark current pulses, and of deep levels, acting as traps of avalanche carriers. Processes of this

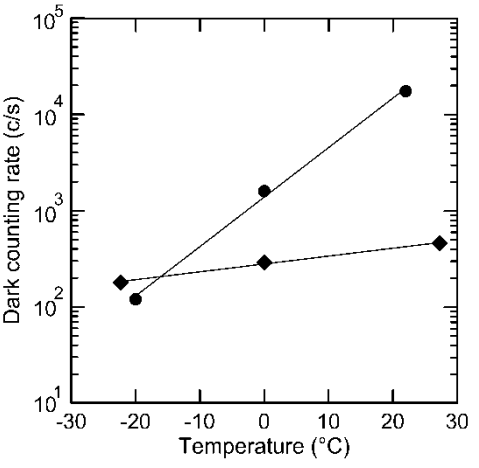

Figure 2. Dark counting rate of thick (●) and thin SPAD (◆) as a function of temperature.

kind are employed in the fabrication of thick SPADs [28, 29] with remarkable success, as confirmed by the experimental data. Conversely, in the fabrication of these thin SPADs specific efforts to optimise the gettering steps had not been made. It can be concluded that there is considerable room to improve the performance of thin SPADs by including more efficient gettering steps in the fabrication process. Thin devices with an active area of a few tens of micrometres diameter and dark counting rates of a few tens of cycles per second can be reasonably targeted. Moreover, the electric field profile across the junction may be better tailored to limit the maximum field value and consequently increase the rate of reduction of the dark counting rate with temperature. This would enhance the effectiveness of device cooling to reduce the dark-counting rate.

4.3 *Afterpulsing*

The afterpulsing effects have been fully characterized by using the experimental procedure reported in reference [23]. The AQC used in the measurements has a dead time of about 36 ns. The total capacitance C_T between SPAD terminals is about 8 pF. Measurements were performed at room temperature.

Figure 3 shows the probability density in time for the occurrence of an afterpulse after an initial avalanche pulse. The data can be fitted with a sum of exponentials and the minimum chi-square value and the best residual misfit distribution were always achieved for both thick and thin SPAD devices.

Table 1 reports the measured lifetimes and amplitudes of the exponential components. Slik^{TM} devices have a total afterpulsing probability (i.e. the integral of the probability density function) of about 0.4%. As a matter of fact, the afterpulsing probability was sufficiently low to be completely negligible in QKD applications. The thin SPAD tested exhibited substantially higher afterpulsing effects: the mean value of the total afterpulsing probability measured on various samples was about 4%.

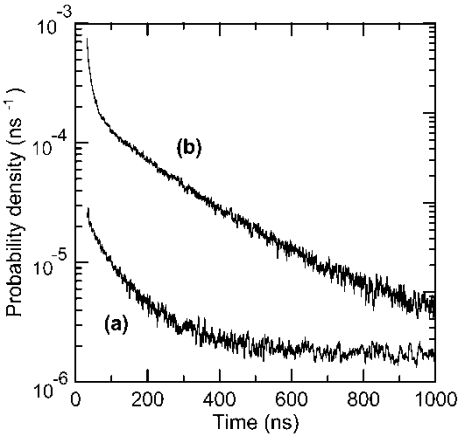

Figure 3. Measured afterpulsing probability density for a Slik^{TM} (curve a) and an average sample of available thin SPAD (curve b) at room temperature. Curves are plotted after correcting data for the first stop effect [26] and subtracting the constant background due to thermally generated carriers.

Table 1. Measured lifetimes and amplitudes of the exponential components resulting from the fitting of the afterpulsing probability densities shown in figure 3.

	Thick SPAD	Thin SPAD
τ_1 (ns)	245	192
A_1 (ns^{-1})	1.077×10^{-5}	2.24×10^{-4}
τ_2 (ns)	25	10
A_2 (ns^{-1})	1.57×10^{-4}	3.07×10^{-2}

It must be stressed, however, that this does not represent an inherent limit of thin SPADs. In fact, figure 3 reports the data obtained for an average-level sample, whereas for selected device samples total afterpulsing probabilities down to 0.15% have been observed. Such a remarkably non-uniform density of trap centres across the wafer surface emphasizes the necessity of introducing efficient gettering steps in the fabrication process. A very remarkable improvement in the performance can be obtained by enhancing the quality of the semiconductor material in the depletion layer, as confirmed by the results already obtained in the fabrication of thick SPADs.

It is interesting to note that the analysis in section 3 leads to the conclusion that an afterpulsing probability of a few per cent does not necessarily worsen the performance of a QKD system, provided that the SPAD detector is suitably operated. In fact, equations (14), (15) and (23) show that afterpulsing effects add to the detector error rate a contribution $QBER_{\mathrm{det,ap}}$ dependent on the SPAD operation mode. We have calculated this contribution for a thin SPAD with the features reported in Table 1.

$QBER_{\mathrm{det,ap}}$ is shown in figure 4 as a function of the pulse repetition period. Different operation modes have been considered: (i) simple gated mode ($T_g = 2.5$ ns), (ii) gated mode enclosing a shorter detection window ($T_g = 2.5$ ns, $\Delta t = 0.5$ ns) and (iii) free running mode operation ($T_d = 50$ ns, $\Delta t = 0.5$ ns).

By assuming a maximum tolerable $QBER_{\mathrm{det,ap}}$ of 1%, it turns out that simple gated mode (curve a) is not suitable for pulse repetition rates higher than 40 MHz. Gated mode with shorter detection window shows better performance, but it is still limited to pulse repetition rates lower than 100 MHz. This limit may be overcome by introducing more sophisticated gating strategies. It has been shown that a sufficiently long hold-off time imposed after an avalanche strongly reduces the afterpulsing probability in the subsequent gate-on pulse [30]. In practice, this hold-off mechanism can be simply implemented by suppressing an appropriate number of gate pulses after each avalanche detection [31]. Finally, free running turns out to be the best performing operation mode: a maximum pulse repetition rate of 500 MHz is achievable, while keeping $QBER_{\mathrm{det,ap}}$ below 1%.

In conclusion, thin SPADs affected by moderately low afterpulsing effects can be employed in QKD applications without an appreciable increase of the QBER, provided that free-running mode is adopted. Thin SPADs fabricated with improved processes are expected to show a strongly reduced total afterpulsing probability, hopefully well below 1%. With such devices the maximum pulse repetition rate could be pushed beyond 1 GHz, being ultimately limited by the minimum achievable dead time.

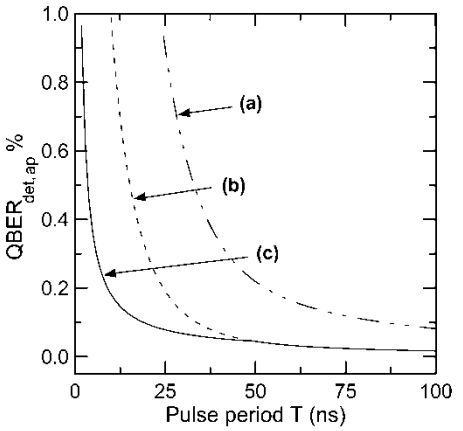

Figure 4. Calculated $QBER_{\mathrm{det,ap}}$ as a function of the pulse repetition period, for a thin SPAD operated in different modes: (a) simple gated mode ($T_g = 2.5$ ns); (b) gated mode with a shorter detection window Δt enclosed ($T_g = 2.5$ ns, $\Delta t = 0.5$ ns); (c) free running mode ($T_d = 50$ ns, $\Delta t = 0.5$ ns).

4.4 *Thermal effects*

When a mixed active/passive quenching circuit is employed, the energy dissipated by the SPAD in each avalanche pulse is approximately given by $V_E V_B C_T$. For the Slik^{TM} operating at 20 V excess bias, the energy dissipated within the device is about 45 nJ per avalanche pulse. At a pulse-counting rate of 1 MHz the mean power dissipation is about 45 mW. The thin SPAD operating at 5 V excess bias, dissipates an energy of about 0.6 nJ per avalanche pulse, which causes a negligible mean power dissipation of 0.6 mW at a pulse-counting rate of 1 MHz.

Thermal effects were investigated by illuminating the SPAD with a LED source emitting at 615 nm wavelength (BNC light pulse generator—model 6010). In order to check for the deviation from linearity, we measured the photon counting rate as a function of the incident optical power. The optical source was operated in stabilized CW mode, and its linearity was assessed by means of a calibrated power meter. A copper block surrounding the TO-8 case was employed as a heat sink, in order to keep the hot side of the Peltier element as close to room temperature as possible. Figure 5 shows the linearity of the Slik^{TM}. The reported counting rate is corrected for the dead-time effect [24].

Thermal effects turn out to set an intrinsic limit to thick SPADs. In fact, the device shows full linearity up to 3 MHz, while a progressively increasing deviation from linearity occurs at higher counting rate. For instance, at 7 MHz the deviation is about 20%.

This behaviour is consistent with the photodetector mounting features. The Slik^{TM} is flip-chip mounted on a ceramic support; to control the temperature, the ceramic support is glued on the cold face of the thermoelectric cooler and a thermistor is mounted in good thermal contact with the ceramic surface. Even with perfectly stabilized temperature of the ceramic support, the temperature rise of the avalanching junction can be considerable, because of the significant thermal resistance between the photodiode junction and the temperature sensor. We

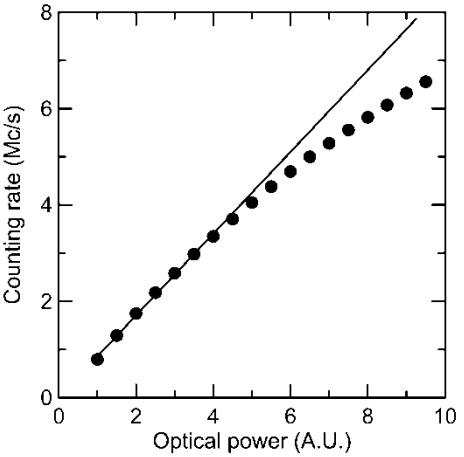

Figure 5. Counting rate of the Slik^{TM} (●) as a function of the incident optical power. The continuous line represents the ideal (linear) response.

experimentally checked that this resistance is 0.1°C/mW, by measuring first the breakdown voltage V_B versus the chip temperature (at low counting rate) and then V_B versus the power dissipation in the chip (with steady avalanche current). At 5 MHz counting rate, the 225 mW power dissipation produces an increase of 20°C in the junction temperature. Since the temperature coefficient is ~0.5 V/°C, V_B increases by ~10 V, the excess bias voltage decreases by the same amount and the detection efficiency is correspondingly reduced. The thermal behaviour could be significantly improved by reducing the thermal resistance between the silicon chip and the thermoelectric cooler and, possibly, by integrating the temperature sensor within the silicon chip. A reduction of the parasitic capacitance seen at the anode of the photodetector would also be desirable to reduce the power dissipation.

Similar measurements were performed on thin SPADs. Since the mean power dissipation is negligible in this case, no cooling elements or dissipators were used. As expected, no deviation from linearity was observed for counting rates up to 10 MHz.

4.5 *Time resolution*

Figure 6 shows the time resolution curve of thick and thin SPADs performed with a switched-gain laser diode emitting 10 ps laser pulses at 820 nm wavelength with 10 kHz repetition rate. Measurements were performed at room temperature.

The thin SPAD response shows a sharp peak (FWHM = 45 ps) and an almost exponential tail, with a lifetime of 270 ps. The latter is due to the photons absorbed in the neutral region beneath the active region, as explained in reference [10]. This tail can be completely eliminated by resorting to more advanced device structures [11]. The Slik^{TM} response has a FWHM of 350 ps and a diffusion tail with 160 ps decay time, which is consistent with the photon absorption occurring in a 0.8 μm thick p + neutral layer placed at the bottom of the Slik^{TM} structure.

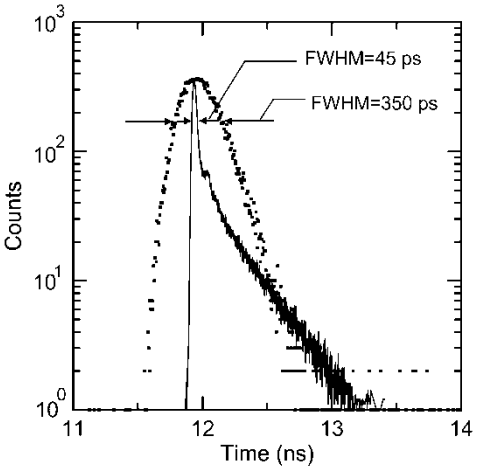

Figure 6. Time resolution curves of the thick (■) and thin SPAD (—) detectors.

5. System performance comparison

From a system point of view, the comparison between thick and thin SPAD must be performed on the basis of potential key distribution rate and QBER. These parameters have been calculated with reference to a first-window fibre optic transmission system operated at a pulse rate of 400 MHz.

The raw key creation rate can be calculated using equation (1), where η_t is given by:

$$\eta_t = 10^{-((L_f l + L_B)/10)}, \tag{29}$$

where L_f is the specific loss in the fibre in dB/km, l is the length of the link in kilometres and L_B are the internal losses at the receiver end in dB.

QBER has been calculated according to equations (23) and (24), by assuming free running operation with $\mu = 0.1$, $L_f = 2.5$ dB/km, $L_B = 3$ dB, $p_{opt} = 1\%$, $\Delta t = 0.5$ ns, $T_d = 50$ ns. The relevant SPAD parameters are summarized in Table 2.

A $-10°$C operating temperature has been considered in the comparison, since this value can easily be obtained by using a Peltier thermo-electric cooler. Finally, the useful key creation rate after error correction and privacy amplification has been evaluated as a function of *QBER*, according to reference [15].

Figure 7 shows the useful bit rate B and the QBER as a function of transmission distance, for thick and thin SPAD devices. The additional horizontal axis on top of the graph makes it possible to easily evaluate the performance of other systems (e.g. free-space systems) as a function of the optical link loss.

Both photodetectors yields satisfactory results. The performance of thin SPAD is actually slightly inferior, due to the lower quantum detection efficiency, which is not completely compensated by the lower dark counting rate. If we assume 100 bits/s as a lower limit for a reasonable key distribution, we get a maximum transmission distance of 15 and 16.5 km for thin and thick SPADs respectively. The maximum bit rate is a few Mbits/s, which is comfortably below the maximum

Table 2. Relevant SPAD parameters used for performance comparison in a QKD system.

	Thick SPAD	Thin SPAD
V_E (V)	20	5
Dark counting rate (c/s)	420	230
Photon detection efficiency @ 820 nm (%)	55	15
Operating temperature (°C)	−10	−10

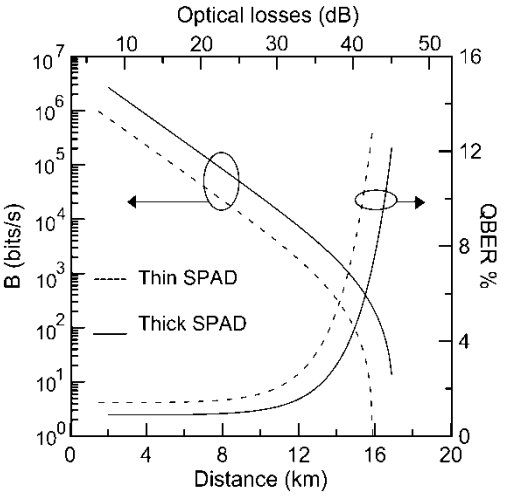

Figure 7. Useful bit rate B and $QBER$ calculated as a function of transmission distance, for thick (—) and thin SPAD (- - -).

(saturated) value of 2×10^7 counts set by the AQC dead time. Such bit rates can be attained without problem by thin SPADs and also by thick SPADs equipped with cooling systems that ensure efficient removal of the dissipated power.

6. Conclusions

In this paper, we report a full characterization and evaluation of two types of silicon SPADs that can be used in QKD applications at short wavelength.

Thick SPADs show high quantum efficiency, low dark counting rate at moderately low temperatures and negligible afterpulsing. They may operate up to an average counting rate of 3 MHz before the occurrence of thermal effects. They are very fine devices, but they have characteristics that make them inherently poorly suited for building compact and economical apparatus. Since they were designed for attaining the highest possible quantum detection efficiency also at long wavelengths, up to 1 μm, they have a thick depletion layer of 30 μm and are built with very high purity silicon wafers and dedicated fabrication processes. These features inherently involve several drawbacks, which can be summarized as follows: (a) the fabrication technology is not compatible with the integrated circuits technologies; (b) the fabrication is costly and gives a low or moderate

yield of good devices; (c) the device inherently operates with high bias voltage, from 300 to 400 V; (d) the power dissipation in the avalanche pulse is high, more than 5 W — already at moderate photon counting rate the self-heating of the device is remarkable; (e) the time resolution of the device (precision in the determination of photon arrival time) is a few hundred picoseconds, whereas better values would be desirable for high bit-rate quantum cryptography.

Thin SPADs have depletion layer thickness of about 1 μm. The quantum efficiency of these detectors is comparable to that of thick SPADs in the visible, up to about 600 nm wavelength, whereas it becomes progressively lower going further in the near infrared range. These thin SPADs are fabricated with material and fabrication technology compatible with the silicon integrated circuit technologies. Therefore, the fabrication cost is low and the yield of good devices is remarkably high. The devices work with moderate bias voltage, from 20 to 50 V. Therefore, they have low power dissipation during the avalanche pulse, less than 0.2 W, and the self-heating effects is moderate even at fairly high photon counting rate. The time resolution is remarkably better than that of thick SPADs: 50 ps resolution FWHM is currently available and values as good as 20 ps FWHM have been attained. The sensitive area of these detectors is fairly small, with diameter up to about 20 μm. Although in other applications this feature may be a drawback, this is not the case for quantum cryptography. In this case wide area detectors are not necessary, since a single-mode fibre with a core diameter of 10 μm or less can be used to transmit the optical signal. The fibre end can be butt-coupled to the detector thus requiring an active area with a diameter comparable to that of the core. The detector design can be improved by keeping the active area diameter small and by moderately increasing the depleted region thickness in order to enhance the ratio between quantum efficiency and dark counting rate.

From a system point of view, thin SPADs appear inherently better suited to high-rate QKD applications because of their faster response time, ruggedness, low voltage, low power dissipation and fabrication technology, which is simple, efficient, economical and compatible with monolithic integration of detector and associated circuits.

From the standpoint of the performance actually attained in QKD, a comparable level is verified for the available devices, with a slight advantage for thick SPADs. The performance of thick SPADs is already optimised, with little prospect left for further improvement. On the other side, there is notable prospect of improving thin SPADs by working on the gettering steps in the fabrication process for minimizing the concentration of both generation and trap centres.

The conclusion of the present analysis and evaluation is that thin silicon SPADs appear to be definitely better suited for QKD systems. Their features constitute relevant practical advantages in the design and execution of QKD systems; the presently available devices already offer performance level comparable to that of thick SPADs; there are definite reasons to expect remarkably improved performance from the next generations of devices.

References

[1] Buttler, W. T., Hughes, R. J., Lamoreaux, S. K., Morgan, G. L., Nordholt, J. E., and Peterson, C. G., 2000, *Phys. Rev. Lett.*, **84**, 5652.

[2] Hughes, R., Buttler, W., Kwiat, P., Lamoreaux, S., Morgan, G., Nordhold, J., and Peterson, G., 2000, *J. Mod. Opt.*, **47**, 549.

[3] GORMAN, P. M., TAPSTER, P. R., and RARITY, J. G., 2001, *J. Mod. Opt.*, **48**, 1887.
[4] TOWNSEND, P., 1998, *IEEE Photonics Technol. Lett.*, **10**, 1048.
[5] DAUTET, H., DESCHAMPES, P., DION, B., MACGREGOR, A. D., MACSWEEN, D., MCINTYRE, R. J., TROTTIER, C., and WEBB, P., 1993, *Appl. Opt.*, **32**, 3894.
[6] SPCM-AQ Single-photon Counting Module Data Sheet, PerkinElmer Optoelectronics, http://opto.perkinelmer.com.
[7] GHIONI, M., and RIPAMONTI, G., 1991, *Rev. Sci. Instrum.*, **62**, 163.
[8] COVA, S., LACAITA, A., GHIONI, M., RIPAMONTI, G., and LOUIS, T. A., 1989, *Rev. Sci. Instrum.*, **60**, 1104.
[9] LACAITA, A., GHIONI, M., ZAPPA, F., RIPAMONTI, G., and COVA, S., 1993, *Nucl. Instr. and Meth.*, **A326**, 290.
[10] LACAITA, A., GHIONI, M., and COVA, S., 1989, *Electron. Lett.*, **25**, 841.
[11] LACAITA, A., COVA, S., GHIONI, M., and ZAPPA, F., 1993, *IEEE Electron Device Lett.*, **14**, 360.
[12] BROWN, R. G., RIDLEY, K. D., and RARITY, J. G., 1986, *Appl. Opt.*, **25**, 4122.
[13] BROWN, R. G., JONES, R., RARITY, J. G., and RIDLEY, K. D., 1987, *Appl. Opt.*, **26**, 2383.
[14] LI, LI-QUIANG and DAVIS, L. M., 1993, *Rev. Sci. Instrum.*, **64**, 1524.
[15] ZBINDEN, H., BECHMANN-PASQUINUCCI, H., GISIN, N., RIBORDY, G., 1998, *Appl. Phys. B*, **67**, 743.
[16] BENNETT, C. H., and BRASSARD, G., 1984, Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India (New York: IEEE), pp. 175–179.
[17] BENNETT, C. H., BESSETTE, F., BRASSARD, G., SALVAIL, L., and SMOLIN, J., 1992, *J. Criptol.*, **5**, 3.
[18] BENNETT, C. H., BRASSARD, G., and ROBERT, J. M., 1988, *SIAM J. Comput.*, **17**, 210.
[19] LÜTKENHAUS, N., 1996, *Phys. Rev. A*, **54**, 97.
[20] LÜTKENHAUS, N., 1999, *Phys. Rev. A*, **59**, 3301.
[21] COVA, S., GHIONI, M., LACAITA, A., SAMORI, C., ZAPPA, F., 1996, *Appl. Opt.*, **35**, 1956.
[22] HAITZ, R. H., 1965, *J. Appl. Phys.*, **36**, 3123.
[23] COVA, S., LACAITA, A., and RIPAMONTI, G., 1991, *IEEE Electron. Dev. Lett.*, **12**, 685.
[24] NICHOLSON, W., 1974, *Nuclear Electronics* (New York: J. Wiley), pp. 374–376 and pp. 259–260.
[25] ZAPPA, F., GIUDICE, A., GHIONI, M., COVA, S., 2002, Proceedings of the 28th European Solid State Circuit Conference, ESSCIRC'02, Firenze, Italy, pp. 355–358.
[26] O'CONNOR, V., and PHILLIPS, D., 1984, *Time-correlated Single Photon Counting* (London: Academic Press).
[27] SPINELLI, A., and LACAITA, A., 1997, *IEEE Trans. Electron Devices*, **44**, 1931.
[28] MCINTYRE, R. J., 1990, Silicon avalanche photodiode with low multiplication noise, US Patent 4,972,242.
[29] MCINTYRE R. J., WEBB, P. P., 1996, Low-noise, reach-through, avalanche photodiodes, US Patent 5,583,352.
[30] LACAITA, A., FRANCESE, P. A., ZAPPA, F., and COVA, S., 1994, *Appl. Opt.*, **33**, 6902.
[31] VYLEGJANINE, K., 2000, Master Thesis (in Russian), Dept. of Physical Electronics, Norwegian University of Science and Technology (NTNU), http://www.vad1.com/qcr/kirill/.