Performance and design of InGaAs/InP photodiodes for single-photon counting at 1.55 μm

Philip A. Hiskett, Gerald S. Buller, Alison Y. Loudon, Jason M. Smith, Ivair Gontijo, Andrew C. Walker, Paul D. Townsend, and Michael J. Robertson

The performance of selected, commercially available InGaAs/InP avalanche photodiodes operating in a photon-counting mode at an incident wavelength of 1.55 μm is described. A discussion on the optimum operating conditions and their relationship to the electric field distribution within the device is presented.

OCIS codes: 230.0040, 230.5170, 230.5160, 270.5290.

1. Introduction

Time-correlated single-photon counting (TCSPC) is a widely used technique in many applications that depend on the detection of low-level and fast optical signals.[1–7] The TCSPC technique has been used for several years in applications such as quantum cryptography,[2,8–12] optical time-domain reflectometry (OTDR),[6,13,14] time-of-flight ranging,[7,15] and time-resolved photoluminescence (TRPL).[3,16,17]

Early TCSPC systems incorporated photomultiplier tubes (PMT's) to detect single photons. These PMT's required high operating voltages, were quite fragile, and had relatively poor detection efficiencies.[18,19] A ubiquitous example of a photocathode is the S1 fabricated from AgOC.[19] These photocathodes are photosensitive only at wavelengths below 1.1 μm. Recently produced PMT's, for example, the Hamamatsu R5509-72, are more rugged, and alternative photocathode materials enable operation up to wavelengths of ~1.6 μm. The detection efficiency of such PMT's, however, remains low,[20] whereas the operating voltage remains very high, ~1700 V for the Hamamatsu device.[20]

Silicon avalanche photodiodes (APD's) operated in the Geiger mode[7,18,21] have shown much higher detection efficiencies than PMT's at wavelengths below 1.1 μm. A Geiger-mode APD, also known as a single-photon avalanche diode (SPAD) detector, is a reverse-bias diode that operates at fields above breakdown, such that a single carrier can initiate a self-sustaining, and hence readily detectable, avalanche current. As a consequence of their enhanced efficiency, silicon SPAD's have been used at wavelengths below 1.1 μm in applications including TRPL microscopy[17,21] and time-of-flight ranging.[7]

The high-efficiency detection of single photons between 1.1 and 1.6 μm is not possible with silicon SPAD's because of the low absorption of this material at these wavelengths. Two applications of TCSPC are quantum cryptography and OTDR. The transmission range of both a quantum cryptography system and an OTDR system across a telecommunication network can be substantially increased when photons of wavelengths 1.3 and 1.55 μm are used because of the low absorption and scattering loss within silica-based optical fibers at these specific wavelengths. The detection of photoluminescence at long wavelengths would allow the investigation of carrier dynamics of a wider range of samples, notably in the InGaAs/InP material system. A time-of-flight ranging system based on a wavelength of 1.55 μm would allow eye-safe operation.

Extensive research carried out by several groups[1–3,8,22,23] has shown that germanium SPAD's are capable of detecting single photons with high detection efficiency and low noise-equivalent power (NEP) at 77 K at a wavelength of 1.3 μm. TRPL

When this research was performed, P. A. Hiskett (p.a.hiskett@hw.ac.uk), G. S. Buller, A. Y. Loudon, J. M. Smith, I. Gontijo, and A. C. Walker were with the Department of Physics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, United Kingdom. I. Gontijo is now with the Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, California 90095-1594. P. D. Townsend and M. J. Robertson are with the Corning Research Centre, Adastral Park, Martlesham Heath, Ipswich IP5 3RE, United Kingdom.

Received 27 January 2000; revised manuscript received 25 August 2000.

0003-6935/00/366818-12$15.00/0

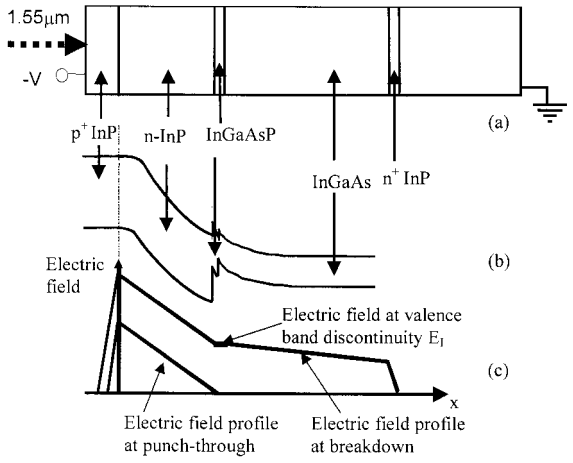

Fig. 1. (a) Schematic diagram of the InGaAs/InP SAGM APD. (b) The band profile of the device clearly showing the valence-band discontinuity at the heterointerface. (c) The electric field of the device at the breakdown point and at punch-through.

measurements on semiconductor material with germanium SPAD's have been used to study the carrier dynamics at a wavelength of 1.3 μm with low photogenerated carrier densities (e.g., $<10^{15}$ cm^{-3}).[3,16] These devices required cryogenic cooling to limit the effects of thermal noise in the device. The unfortunate consequence of this cooling was a shift in the band edge of the material causing a drop in the detection efficiency to <1% at a wavelength of 1.55 μm.[22]

The low detection efficiency of germanium-based SPAD's at a wavelength of 1.55 μm has led to research in detectors grown from alternative material systems. The separate absorption, grading, and multiplication (SAGM) InGaAs/InP APD's have been shown to be capable of detecting single photons at a wavelength of 1.55 μm.[4,13,24] The devices characterized in Refs. 4, 13 and 24 were designed for room-temperature, non-photon-counting operation; therefore the photon-counting performance of these devices was found to vary significantly.

In this paper we concentrate on the comprehensive characterization of a number of commercially available InGaAs/InP APD's which, although not specifically designed for single-photon detection, are operated as single-photon detectors. Included is a comprehensive study of the dark noise, detection efficiency, timing response, and NEP of these commercially available InGaAs/InP detectors at temperatures between 77 and 225 K.

2. Avalanche Photodiodes

The SAGM InGaAs/InP APD has been commercially available for several years now. A schematic of the microstructure, the band structure, and the electric field profile of a SAGM is shown in Fig. 1.

The device, which is intended for room-temperature, non-photon-counting telecommunications use, is based around a heterojunction formed between InP and InGaAs. Under reverse-bias operation, a high electric field, large enough to induce impact ionization, is present in the InP layer, which forms the multiplication region in the device. Long-wavelength photons (1.2 μm ≤ λ ≤ 1.7 μm) incident on the device pass through the InP layer and are absorbed in the narrower-gap InGaAs layer producing electron–hole pairs. The device is designed so that, under normal operating conditions, the depletion layer extends from the p^+n junction in the InP layer into the InGaAs layer. Photogenerated carriers created in the InGaAs layer will undergo drift because of the presence of the electric field. It has been shown[25] that, to obtain low multiplication noise, the carrier with the higher-impact ionization coefficient should initiate the avalanche process. For InP, holes have the higher-impact ionization coefficient,[26] and so devices are designed to be p^+n; hence photogenerated holes are swept into the multiplication region.

An important feature of the band profile of the device is the valence-band discontinuity between the InP (1.35-eV) and the InGaAs (0.73-eV) layers. This hinders the transportation of holes from the narrow-gap InGaAs material to the wider-gap InP material. To increase the transition rate of the holes, one or more layers of the quaternary material, InGaAsP, are placed between the InGaAs and InP layers. The quaternary layer has an intermediate bandgap between that of the InP and InGaAs and hence grades the valence-band discontinuity. As a result, the holes can cross the barrier more efficiently, and the bandwidth of the device is increased.

For the device to possess a high quantum efficiency and high bandwidth, it is necessary for the electric field at the heterointerface between the quaternary and the InGaAs to be nonzero, i.e., the depletion layer must punch through into the absorption layer. The extent by which the depletion layer extends into the InGaAs layer determines the electric field at the heterointerface. The magnitude of the field at the barrier strongly influences the quantum efficiency and the bandwidth of the device because of the effect the field has on the probability and rate of transport of carriers across the barrier. The field at the heterointerface also greatly affects the dark current; the tunneling current within the InGaAs layer will be too large if the field at this point is too high.

The tunneling current density in the InGaAs layer is given by[27]

$$\frac{I}{A} = \gamma \exp(-\theta m_0^{1/2} \varepsilon_g^{3/2} / q\hbar E_I), \tag{1}$$

where $\theta = \alpha(m^*/m_0)^{1/2}$. α depends on the details of the tunneling barrier, and it is usually assumed that $\alpha \cong 1$ and

$$\gamma = (2m^*/\varepsilon_g)^{1/2}(q^3 E_I V / 4\pi^2 \hbar^2),$$

where m^* is the effective tunneling mass of an electron, m_0 is the mass of the free electron, ε_g is the bandgap of InGaAs, E_I is the electric field at the

heterointerface, V is the voltage drop across the depletion region within the InGaAs layer, $\hbar$ is Planck's constant divided by 2π, and q is the elementary charge.

It is widely accepted that the current density at the InGaAsP/InGaAs barrier should be less than 1×10^{-4} A cm^{-2}.[28,29] Above current densities of 10^{-4} A cm^{-2}, tunneling will contribute a large dark current that would inhibit the performance of the device. The upper limit of the electric field at the heterointerface to satisfy the low tunneling current criteria will depend on the specific device design. The doping density and thickness of the InGaAs layer and the value of the field at the heterointerface will determine the electric field distribution within the InGaAs layer and therefore the exact value of the voltage drop across the layer. Generally, for an absorption layer of doping $\sim 10^{15}$ cm^{-3} and of a thickness of 3 μm, the maximum field at the heterointerface is between 150 and 200 kV cm^{-1}. However, there is also a lower limit of the electric field, below which the holes will traverse the barrier slowly, and so the bandwidth of the device will be greatly impaired. This lower limit cannot be determined directly because the characteristics depend on the configuration of the valence-band deformation at the heterointerface.[28,30,31]

For any given design of a SAGM it is therefore necessary to determine accurately the electric field profile of the depletion layer as it extends through the device. To calculate this electric field profile it is first necessary to determine the critical electric field, at the p^+n junction, at the point of breakdown. At a given temperature, we can determine the critical value of the electric field by considering the multiplication of an electron in the multiplication region of the APD. This can be defined by[32]

$$M_e = \frac{1}{1 - \int_0^W \alpha \exp\left[-\int_0^x (\alpha - \beta)\, \mathrm{d}x'\right] \mathrm{d}x}, \qquad (2)$$

where w is the width of the depletion layer, α is the (electric-field-dependent) impact ionization coefficient for electrons, and β is the (electric-field-dependent) impact ionization coefficient for holes.

From Eq. (2) it is clear that at the breakdown point, where $M \to \infty$, the integral in the denominator must equal 1, i.e.,

$$\int_0^W \alpha \exp\left[-\int_0^x (\alpha - \beta)\, \mathrm{d}x'\right] \mathrm{d}x = 1. \qquad (3)$$

Solving Eq. (3) for the breakdown condition will determine the value of the critical electric field for the given structure.

Once we obtain the value of the critical electric field, it is straightforward to determine the electric field profile throughout the device using Poisson's equation. At a low applied bias, the electric field may well drop to zero within the multiplication layer, i.e., the depletion region does not extend into the absorption layer where all the photons are being absorbed. Under such biasing conditions, the quantum efficiency of the device is effectively zero; no photogenerated holes are swept into the multiplication layer. As the applied voltage is increased, a point will be reached where the electric field becomes nonzero at the heterointerface barrier. The voltage value at which this occurs is known as the punch-through voltage. Above the punch-through voltage, carriers created within the InGaAs layer by photon absorption can undergo drift because of the nonzero electric field and can be swept into the multiplication layer.

It is expected that to utilize these detectors successfully to count single photons, the detectors will have to be cryogenically cooled to limit thermally excited dark events (see Subsection 5.D). It is important therefore to consider the operation of these SAGM devices at low temperatures.

As the temperature of the APD is decreased, the critical electric field and breakdown voltage also decrease because of the reduced carrier scattering by optical phonons. For a cooled device, the breakdown condition for the InP layer can be reached at a voltage below that of punch through, which is relatively unaffected by temperature (unless a freeze out of the junction dopants occurs). That is, breakdown occurs in the InP but there is no depletion layer present in the absorption layer, and hence few photogenerated holes will drift into the multiplication layer. Therefore, the position of the breakdown voltage with respect to the punch-through voltage at the desired operating temperature is of crucial importance.

The reduction in the critical electric field with temperature also reduces the electric field at the valence-band discontinuity; therefore holes may start to be impeded by the barrier. The two detectors discussed in this paper were designed for room-temperature operation; hence it is assumed that the electric field at the heterointerface was optimized so as to limit dark current and increase quantum efficiency and bandwidth at or near room temperature. At cryogenic temperatures, the electric field at the heterointerface may be zero, resulting in a device with low detection efficiency.

3. Description of a Time-Correlated Single-Photon-Counting System

A schematic of the TCSPC system, which we used to measure a fast or weak laser pulse, is shown in Fig. 2. The time-to-amplitude converter (TAC) measures the time difference between a start and a stop event. In the configuration shown in Fig. 2, the start event is coincident with the laser pulse, and the stop event is coincident with the detection of a single photon by the SPAD. The TAC is fundamentally a capacitor, which is linearly charged by a constant current source after the receipt of a start signal. Upon receiving the stop pulse, the capacitor stops charging, and the voltage across the plates of the capacitor is then proportional to the time difference between the start and the stop event (if no stop pulse

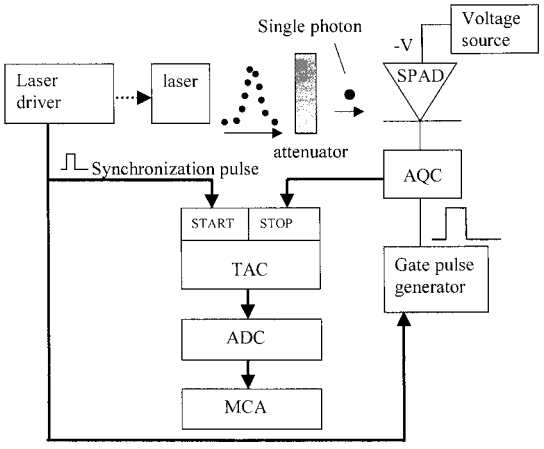

Fig. 2. Schematic of a TCSPC system. ADC, analog-to-digital converter; TAC, time-to-amplitude converter; MCA, multichannel analyzer; AQC, active quench circuit.

arrives, the TAC resets in readiness for the next start signal, i.e., the next laser pulse). The measured TAC output voltage is digitized by an analog-to-digital converter and relayed to a multichannel analyzer (MCA). The MCA consists of 4096 channels, covering the entire range of TAC outputs, each channel representing a finite voltage (time) range.

Measurement of the time difference between start and stop events (i.e., the arrival time of the single photon) is repeated typically several thousand times; each time the MCA stores the measured voltage value in the relevant channel. After a large number of repetitions, the probability distribution of the TAC output voltage is obtained. We can convert these voltage values to times by calibrating the timing electronics using an accurate pulse generator. The time per channel of the MCA in this system was 7.9 ps/channel.

An important feature of the TCSPC system in Fig. 2 is the active quenching circuit (AQC). The AQC has two functions. The first is to control the applied bias to the SPAD. The voltage source, shown in Fig. 2, applies a constant reverse-bias voltage to the p^+ side of the SPAD to bias the device above breakdown. The voltage applied to the n side of the device is controlled by the AQC and can be either 0 V or -6 V. If -6 V is applied, the potential difference across the SPAD is reduced by 6 V. Provided that the SPAD is biased less than 6 V above breakdown, the application of -6 V by the AQC will cause the bias to drop below breakdown. When the bias is below breakdown, a self-sustaining avalanche is not possible; the SPAD is essentially switched off.

The AQC output can be modulated by use of a gate pulse generator. The duration and frequency of the SPAD above breakdown is determined by the gate generator. For the majority of the measurements presented in this paper, the SPAD detectors were gated on (i.e., above breakdown) for a time of 162 ns at a repetition rate of approximately 1 kHz. At this low repetition rate, the contribution to the overall dark count rate caused by the influence of previous avalanche effects (see Subsection 5.D) is negligible. During gated operation, the arrival time of the single photons is time correlated to the SPAD being above breakdown.

The second function of the AQC is to quench a self-sustaining avalanche. When the SPAD is held above breakdown, a self-sustaining avalanche can be caused by either a photon or dark event. During a self-sustaining avalanche, a potentially damaging, large current flows through the device. To avoid damage to the device and the filling of trap centers within the material, the avalanche is quenched by the AQC superimposing -6 V onto the n side of the SPAD, hence lowering the applied bias to below breakdown. Simultaneously with the application of the quench pulse, a -1-V pulse is outputted from the AQC to stop the TAC. After quenching an avalanche, the AQC holds the SPAD below breakdown until the next gate pulse is received. Another popular quenching technique is that of passive quenching; a description of this technique has been widely published (see, for example, Ref. 33).

It is also possible to use a photon-counting system without a quench circuit. In such a system a voltage source is used to bias the SPAD at a voltage just below breakdown, and a negative pulse is superimposed to increase the bias above breakdown. The duration of the pulse is kept deliberately very short (e.g., <10 ns) so that the SPAD is only above breakdown for a very short time. A self-sustaining avalanche can only flow for the remainder of the short gate pulse, at which point the avalanche will cease when the SPAD is restored below breakdown. Other groups[10,11] use this technique successfully in quantum cryptography because photons arrive within a small time window, so the gate pulse can be restricted to a short duration. The time for which an avalanche is permitted to flow determines the maximum gate frequency at which a SPAD can be operated (explained in Subsection 5.D). In applications such as TRPL and OTDR, the photons have a much wider range of arrival times and so the gate pulse will often be required to be in excess of 100 ns. With such a long gate pulse, it is imperative to use some form of quenching (whether active or passive) to restrict the amount of time the large avalanche current flows through the device.

4. Experimental Methods

In this paper we concentrate on the characterization of a range of commercially available InGaAs/InP SAGM Fujitsu APD's. From a sample of 12 APD's, two were chosen for more detailed investigation. The first device, a Fujitsu FPD13U81SR, 80 μm in diameter, was selected because it is punched through (i.e., the breakdown voltage exceeds the punch-through voltage) by ~23 V at 77 K. This device is referred to here as device A. The other device, a Fujitsu FPD13W31RT, has a diameter of 30 μm and was selected because it is not punched through at 77 K. This device is referred to here as device B.

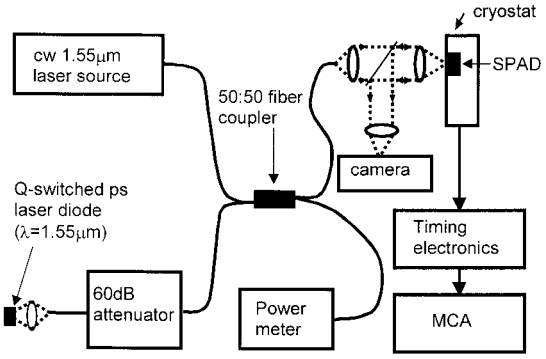

Fig. 3. Schematic of the system used to characterize the free-space Fujitsu devices. A 60-dB attenuator was used to reduce the 10^5 photons/pulse coupled into the fiber to the appropriate level of 0.1 photon/pulse.

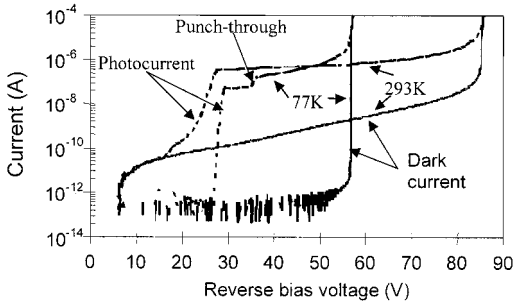

Fig. 4. Dark current and photocurrent of the 80-μm-diameter Fujitsu APD, device A, at 77 and 293 K. The punch-through position can clearly be seen at ~34 V. The *I–V* tracer used for these experiments could not resolve below 1 pA, therefore the current looks erratic below this value on the 77 K dark curve.

These devices were characterized at temperatures between 77 and 225 K by the system shown in Fig. 3.

The laser used is a passively *Q*-switched laser diode with the saturable absorber formed by heavy ion bombardment.[34] The laser produced short pulses of 35-ps full width at half-maximum (FWHM) (measured independently by a streak camera). A 60-dB attenuator was used to reduce the number of photons per pulse to 0.1. This low photon level ensures that the probability of two or more photons reaching the SPAD is negligible, which is important for detection efficiency measurements (see Subsection 5.B). The passively *Q*-switched laser diode was coupled into one of the two inputs of the 50:50 coupler. To align the photons from the *Q*-switched laser with the APD in the cryostat, a white-light source and imaging optics were used to image the APD on an infrared vidicon camera. A 1.55-μm wavelength continuous-wave laser source was connected to the other input arm of the fiber coupler. This laser was used to illuminate the APD during measurements of the photocurrent current–voltage (*I–V*) characteristic (see Subsection 5.A).

5. Experimental Results

A. Current–Voltage Characteristics

The most useful starting point in SPAD characterization is to measure the *I–V* characteristic of the device over the range of desired temperatures. To obtain the photocurrent *I–V* characteristic, the experimental setup shown in Fig. 3 was used; however, the cw laser source was used rather than the *Q*-switched laser. A Hewlett-Packard semiconductor parameter analyzer (Model HP4145b) was used to measure the *I–V* characteristic of the SPAD. The photocurrent *I–V* characteristic will give the breakdown and punch-through voltage, thereby determining if the device is capable of detecting photons at the desired temperature. The punch-through voltage can be easily identified on the photocurrent *I–V* characteristic. At the punch-through voltage there is a sudden jump in the photocurrent caused by photogenerated holes in the InGaAs being swept into the multiplication layer (see Figs. 4 and 5).

Device A (the 80-μm-diameter device) was chosen because the device punches through by a considerable amount, ~23 V, at 77 K. Figure 4 shows the *I–V* characteristic obtained for this device both at room temperature and at 77 K. Figure 4 also shows the *I–V* characteristic for both dark operation and with incident 1.55-μm wavelength light at the two temperatures. The punch-through can clearly be seen in the 77 K photocurrent curve to be at ~34 V, and the breakdown voltage of this device at 77 K is 57 V. This device is therefore capable of detecting photons at a wavelength of 1.55 μm at all temperatures between 77 K and room temperature because, even at temperatures as low as 77 K, the depletion layer extends into the absorption layer. Holes created in this layer by photon absorption can be swept into the multiplication region by the electric field.

Device B was selected as a comparison because it is not punched through at 77 K. This can clearly be seen from the *I–V* trace in Fig. 5. The punch through is at ~33 V whereas the breakdown voltage at 77 K is 30.1 V. To identify the temperature where

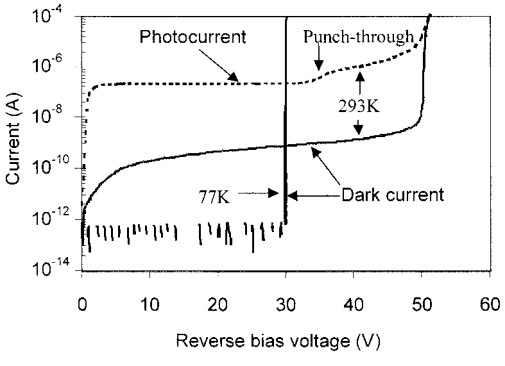

Fig. 5. *I–V* dark characteristic of the 30-μm-diameter Fujitsu APD, device B, at 77 and 293 K and also the photocurrent curve taken at 293 K using 1.55-μm wavelength light. The punch-through position can clearly be seen at ~33 V. The *I–V* tracer used for these experiments could not resolve below 1 pA, therefore the current looks erratic below this value on the 77 K dark curve.

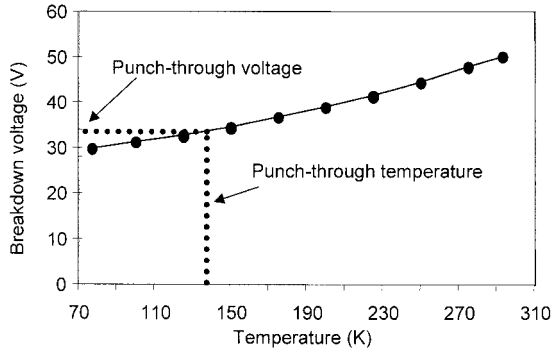

Fig. 6. Graph of breakdown voltage versus temperature for device B. The temperature at which punch-through occurs, 138 K, is indicated.

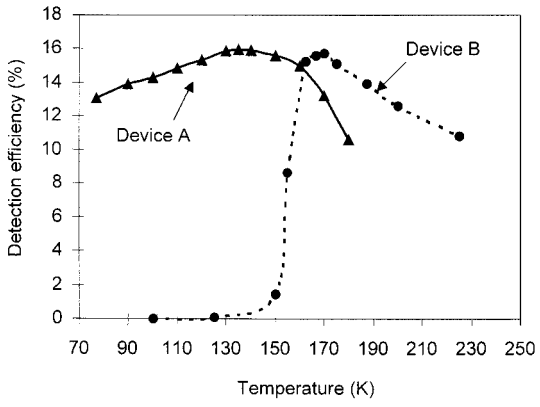

Fig. 7. Detection efficiency of both devices A and B as a function of temperature. At each temperature the devices were operated at an excess bias of 3 V. The detection efficiency of device B is effectively zero at temperatures less than 150 K because the device is not punched through. Curves are drawn between the measured points to aid the reader.

the breakdown voltage exceeds the punch-through voltage, the breakdown voltage was recorded at various temperatures between 77 K and room temperature. This relationship is shown in Fig. 6. Device B does not punch through until a temperature of 138 K is reached.

B. Detection Efficiency

The detection efficiency of a single-photon detector can be defined as the probability that a single photon, incident on the device, will trigger a self-sustaining avalanche. The detection efficiency of a SPAD is defined as the product of the quantum efficiency and the trigger probability.[4] The quantum efficiency of the detectors discussed in this paper is the probability of a photon being absorbed in the InGaAs layer and the subsequent photogenerated hole being swept into the InP multiplication region. The trigger probability is defined as the probability that a hole, once inside the multiplication region, will initiate a self-sustaining avalanche. The trigger probability has been shown[35] to be dependent on the relative excess bias, i.e., the voltage above breakdown relative to the actual breakdown voltage.

To perform single-photon-counting measurements with a pulsed laser, it was necessary for us to attenuate the optical pulse to reduce the number of photons coupled into the fiber (typically 10^5) to the single-photon level. To perform detection efficiency measurements, the attenuation must be high enough to ensure that a maximum of one photon per laser pulse reaches the SPAD. This is a requirement of the single-photon-counting system because, after receiving a start signal from the laser driver, the TAC begins to charge. After receiving the first stop signal (whether it be from a photon event or dark event), the TAC shuts off and remains off until the next start pulse. Therefore in any single laser pulse the TAC can record only a maximum of one photon (or dark) event. To achieve the required low photon level, a 60-dB attenuator is used to reduce the average number of photons, from each laser pulse, reaching the SPAD to just 0.1. At this reduced level, the probability of two or more photons from any single laser pulse reaching the SPAD is negligible.

The measurements of detection efficiency as a function of temperature for the two devices, operated at 3-V excess bias (i.e., 3 V above breakdown), are shown in Fig. 7. There is a definite peak in the detection efficiency at 135 K for device A and at 170 K for device B. The increase in detection efficiency with temperature arises because of the increase in quantum efficiency, i.e., the probability increases that a photon will be absorbed and the subsequent photogenerated hole will drift into the multiplication region. Also the rise in temperature increases the amount by which the device punches through, thereby increasing the electric field at the heterointerface between the InP and the InGaAs layers.

As mentioned above, the product of the quantum efficiency and the trigger probability determines the detection efficiency, and the trigger probability is dependent on the relative excess bias. Both devices A and B were kept at a constant 3-V excess bias when the temperatures were varied, so because the breakdown voltage increases with temperature, 3 V constitutes a lower relative excess bias at higher temperatures. The trigger probability therefore decreases with increasing temperature when a constant excess bias is applied. The results for both devices at 3 V above breakdown indicate that the detection efficiency is dominated by the low quantum efficiency at low temperatures and by the low trigger probability at higher temperatures. At low temperature, the trigger probability is high whereas the quantum efficiency is low. As the temperature rises the quantum efficiency begins to increase whereas the trigger probability begins to fall.

The efficiency of device B can be seen to be approximately zero below 150 K. This is a consequence of the device not being punched through below 138 K. Above 150 K there is a rise in the detection efficiency

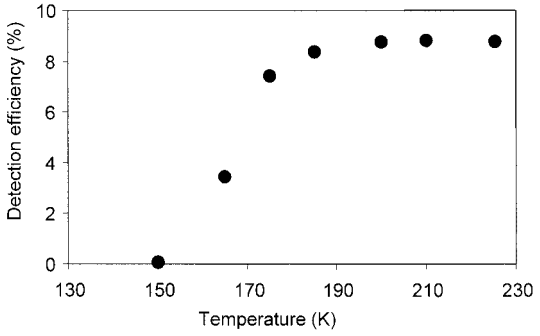

Fig. 8. Detection efficiency versus temperature for device B at a constant 5% excess bias. With a constant relative excess bias, the trigger probability of the device is the same for each temperature. The detection efficiency initially increases because of the large increase in quantum efficiency caused by an increase in the probability of photogenerated holes that surpass the heterointerface barrier as the device becomes more punched through. The detection efficiency levels off above 190 K. At this point it is assumed that the quantum efficiency reaches a maximum value; and, because the trigger probability remains constant, the detection efficiency also remains constant.

until 170 K, at which the efficiency begins to fall again because of the reduction in trigger probability.

To illustrate the effect of keeping a constant relative excess bias, we repeated detection efficiency measurements for device B, over a range of temperatures between 150 and 225 K, while we maintained the SPAD at 5% relative excess bias. In this situation, the trigger probability will be constant. The results, shown in Fig. 8, indicate a rise in detection efficiency with temperature up to 190 K where the detection efficiency begins to level off. We believe that the efficiency rises because of the increase in quantum efficiency which is attributable to the increase in the electric field at the valence-band discontinuity as discussed above. The efficiency is believed to level off because the device becomes sufficiently punched through so that the holes have a high probability of successfully crossing the heterointerface. On reaching this condition, if the temperature is increased it will not increase the probability of hole transition any further, and hence the quantum efficiency will become constant. With a constant trigger probability, the detection efficiency of the device will also become constant.

C. Instrumental Response Time

The instrumental response time of a photon-counting system is a measurement of the maximum sensitivity of the TCSPC system and is usually expressed as a FWHM. It is the convolution of the shape of the laser pulse and the response of the photon-counting apparatus.[36] To determine the instrumental response, an attenuated pulse from the Q-switched 1.55-μm wavelength laser was focused into the active area of the SPAD. The measured probability distribution on the MCA is the instrumental response time of the entire system, including the detector, the timing electronics, and the laser pulse.

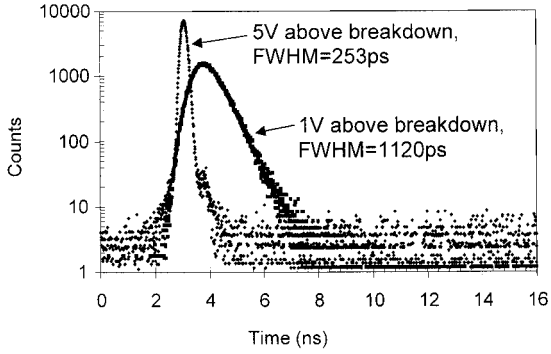

Fig. 9. MCA probability histograms showing the timing response of device A at two different bias settings above breakdown. There is a noticeable decrease in instrumental response time at higher excess bias.

An actual MCA probability distribution recorded with this 35-ps FWHM laser pulse is shown in Fig. 9. Two instrumental response times are shown for device A biased at 5 V above breakdown and at 1 V above breakdown at 77 K. The FWHM values were measured to be 253 and 1120 ps for 5 V above breakdown and 1 V above breakdown, respectively, both of which are considerably longer than the 35-ps FWHM of the laser pulse. There is a contribution to jitter from the timing electronics in the system; however, previous studies[21] indicate that this contribution is low (approximately tens of picoseconds). Most of the jitter must therefore originate from the SPAD itself.

The timing jitter found in SPAD detectors can originate from at least four sources: (1) the timing difference arising from photon absorption at different depths and the subsequent variance in drift time; (2) the slow emission of holes across the heterointerface[35]; (3) the stochastic nature of the multiplication process itself[25]; and (4) the lateral movement of the multiplication across the detector active area, which is strongly influenced by the location of the seed point of the avalanche.[37] In the latter three cases an increase in the excess bias reduces the jitter; however, we show that increasing the excess bias also has the effect of increasing the dark count probability.

The instrumental response times of the two devices were measured at temperatures between 77 and 225 K at 3-V excess bias. These results are shown in Fig. 10. The response time of device A can be seen to increase slightly with temperature, whereas the response time for device B decreases with temperature. Device A is punched through by ~23 V at 77 K; if the temperature is increased, it will further increase the amount the breakdown voltage exceeds the punch-through voltage. For a device punched through by this amount, it is believed that the holes encounter little resistance when they cross the heterointerface barrier, even at 77 K, and that there is little variation of transit times of carriers that cross the barrier.

As the temperature is increased, the electric field at the heterointerface also increases. This increased electric field usually decreases the jitter in the transit time of the carriers that cross the barrier;

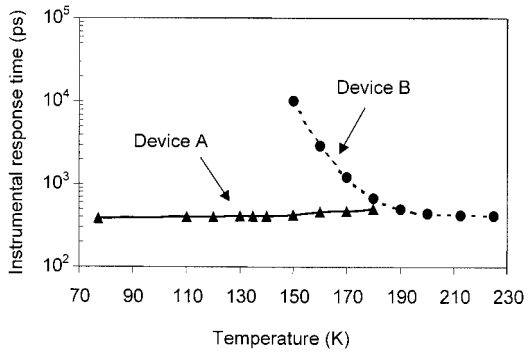

Fig. 10. Instrumental response time (FWHM) versus temperature for both device A and device B at an excess bias of 3 V.

however, it is believed that, because the field at the heterointerface is so high even at low temperatures, the increase in temperature has little effect on the jitter.

The instrumental response time of device B shows a dramatic improvement as the temperature is increased above 150 K. The instrumental response time continues to shorten up to 190 K at which point it levels off. At this point, the field at the heterointerface of the device is now believed to be sufficient to minimize the variation in time that holes take to surpass the heterointerface barrier. Any increase in the temperature will not further decrease the jitter. The main factor that contributes to the instrumental response time of device B up to 190 K is the voltage by which the device is punched through. Below 150 K the device is not punched through; at 150 K the detection efficiency becomes nonzero, and hence the value of the instrumental response time can be measured. At 150 K the instrumental response time is very high (10 ns); this is expected because the electric field at the heterointerface of the device will be very small ($\ll$150 kV cm^{-1}), hence holes will encounter a large obstacle at this barrier as they are swept toward the InP layer. With each increment in temperature the device is punched through by a greater voltage, thus increasing the field at the heterointerface and assisting the transition of holes across the heterointerface barrier. The device is still only punched through by $\sim$7 V at breakdown at 225 K (compared with a punch through of $\sim$40 V for device A at the same temperature).

D. Dark Counts

A dark count is a recorded count that originates from a noise event rather than a photon event. The TCSPC system cannot distinguish between self-sustaining avalanches caused by photon absorption or those caused by dark events. An understanding of the origins of dark counts and the limit to which they restrict the detection of single photons is therefore vital. Dark counts in a single-photon detector are caused by three separate phenomena: (1) thermal noise, (2) tunneling, and (3) trap release.

Thermal noise is caused by thermal excitation of electrons into the conduction band from the valence band and can therefore be significantly reduced when the detectors are cooled. However, it was shown in Subsection 5.C that it can be beneficial to work at temperatures above 77 K to increase the detection efficiency of the device, even, as shown below, at the expense of a higher dark count rate.

Tunneling, where an electron tunnels into the conduction band from the valence band, can occur at high fields. This tunneling can be suppressed when the device is designed to have low doping in both InP and InGaAs layers and when the amount of excess bias is limited. The origin of tunneling in InGaAs/InP SAGM APD's was discussed above.

Trap centers within a device are caused by defects in the material. During a self-sustaining avalanche, the large charge flow through the device causes these trap centers to fill. After the avalanche is quenched and the SPAD is restored below breakdown, the trap centers begin to release the trapped carriers. If the gate repetition rate, at which the SPAD is being operated, is sufficiently high, the SPAD may be restored above breakdown while the traps from the previous avalanche are still emptying. These released carriers can potentially activate another self-sustaining avalanche, thus causing the system to record a dark count. The frequency at which the SPAD is gated must be sufficiently low to allow all the traps to empty. Traps that empty while the SPAD is below breakdown cannot initiate a self-sustaining avalanche.

Silicon devices are relatively free from defects, and so trapping contributes little to the dark counting rate. However, InGaAs-based detectors cannot currently be produced with the same material quality as silicon detectors. InGaAs/InP detectors are also formed around a heterostructure and are expected to contain more defects. As a consequence of this, the InGaAs/InP devices must be operated in a gated mode. The effect of traps on the dark count probability is highly dependent on excess bias. It has been shown[1,4] that the greater the charge flow through the device during a self-sustaining avalanche, the greater the number of carriers that are trapped. The charge flow through the device is strongly dependent on the excess bias voltage, therefore high values of excess bias can severely limit the maximum gate frequency of the system. The length of time between the onset of the self-sustaining avalanche and the pulse from the quench circuit reaching the device is also a key factor in the limitation of traps. During this time between the onset of avalanche and quenching, a large current is able to flow through the device, filling trap states. In the system used in this paper it is estimated that the charge flows for approximately 15 ns before the SPAD is quenched. Part of the reason for the long quenching time is the transit time of the quenching pulse to the SPAD through a long length of cable ($\sim$1.5 m in this case) which was necessary to connect the active quench circuit (described above) to the SPAD in the liquid-nitrogen cryostat.

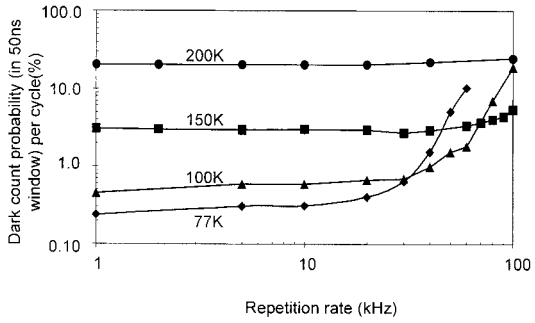

Fig. 11. Dark count probability (within a 50-ns window) per cycle versus a repetition rate of the gate pulse for an 80-μm-diameter Fujitsu device biased at 1% relative excess bias at different temperatures.

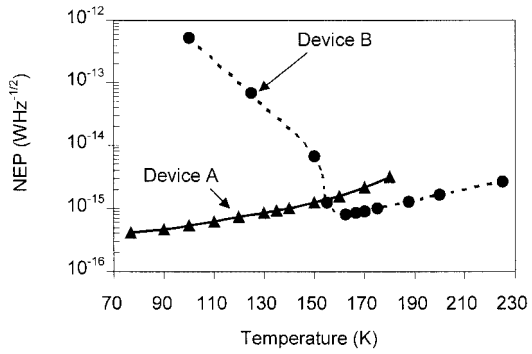

Fig. 13. NEP versus temperature for both device A and device B. At each temperature the device was operated at an excess bias of 3 V.

The dark count probability is defined as the probability that the SPAD will breakdown, in the absence of any photons. Figure 11 shows a graph of dark count probability (in a 50-ns window) versus a repetition rate for device A at temperatures between 77 and 200 K. At each temperature, the SPAD was biased at a relative excess bias of 1%. The same system that is used to measure the detection efficiency, shown in Fig. 3, is used to measure the dark count probability; however, the laser pulse is now blocked to ensure that the SPAD is in total darkness. All the dark count probabilities quoted in this paper are per 50 ns of the gate pulse.

The dark count probability (in a 50-ns window) of devices A and B as a function of temperature is shown in Fig. 12 with both devices operated at a constant excess bias of 3 V. The repetition rate of the system was restricted to approximately 1 kHz to ensure that the trap centers are emptied before the SPAD is restored above breakdown. Figure 12 shows that device A has a higher dark count probability than device B for a given temperature. There are many factors that could explain this; however, two of the most probable reasons are that device A is punched through by a considerably further distance than device B at each temperature and that the dark counts will, generally, be higher for the larger-area device. From the data obtained for both detection efficiency and dark count rate, the NEP of the two devices was then calculated. These results are shown in the next subsection.

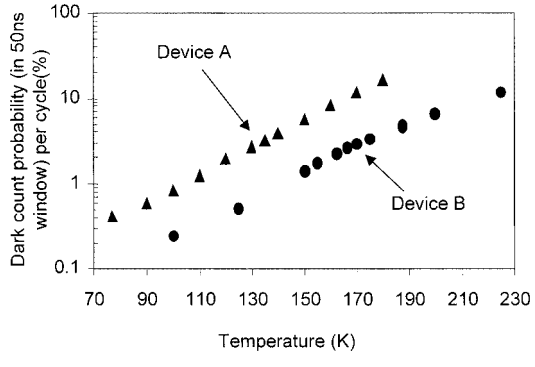

Fig. 12. Dark count probability (within a 50-ns window) per cycle of both devices A and B as a function of temperature. At each temperature the devices were operated at an excess bias of 3 V.

E. Noise-Equivalent Power

The NEP is the most useful of the figures of merit for photon-counting detectors because it incorporates both detection efficiency and dark count probability. The NEP is defined as the signal power required to attain a unity signal-to-noise ratio within a 1-s integration time.[4] The equation used to calculate the NEP is[4]

$$\mathrm{NEP} = \frac{h\nu}{\mathrm{DE}} \sqrt{2R}, \tag{4}$$

where DE is the detection efficiency of the device, R is the dark count rate integrated over 1 s, h is Planck's constant, and ν is the frequency of the photons.

The dark count rate R defined above is the total number of dark counts recorded per second for a detector that is operated continuously (i.e., not gated). If, however, the detectors are gated (as is the case in this paper), the dark count rate has to be scaled appropriately. For example, consider a detector that is gated on for 50 ns at a frequency of 1 kHz and has a dark count rate of 10 Hz; as this detector is only above breakdown for 50 μs (50 ns × 1000) in any 1 s, we must then scale the dark rate to continuous operation. We did this by assuming that the expected dark count over 1 s is 10 Hz × (1 s/50 μs), i.e., 200 000 Hz. From the detection efficiency and dark count results obtained for devices A and B, the NEP for both devices at 3-V excess bias was calculated over the range of temperatures from 77 to 225 K. The graphs of NEP versus temperature for the two devices are shown in Fig. 13.

Calculations reveal that the NEP for device A is a minimum at 77 K whereas the minimum NEP for device B is at 162 K. The effect of excess bias on the detection efficiency, instrumental response,

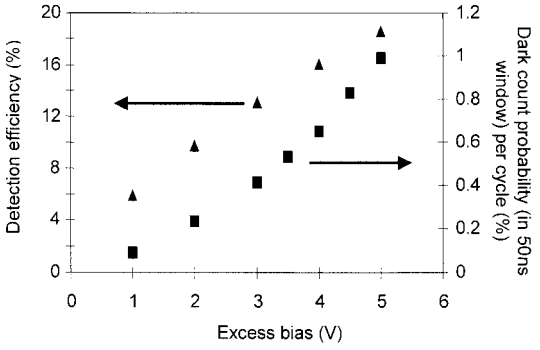

Fig. 14. Detection efficiency and dark count probability versus excess bias for device A at 77 K.

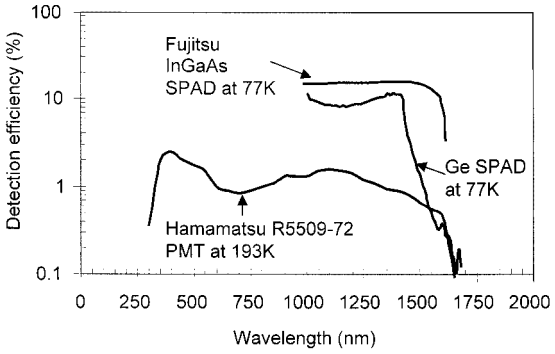

Fig. 17. Detection efficiency versus wavelength for various single-photon detectors.

dark performance, and NEP of device A was then investigated at 77 K, the optimum temperature for the operation of this device. The detection efficiency and dark count probability per cycle of device A can be seen in Fig. 14, the calculated NEP is shown in Fig. 15, and the instrumental response is shown in Fig. 16.

The NEP is relatively constant at all values of excess bias; however, the instrumental response time of the device reduces considerably at higher values of excess bias. For use in photon-counting applications, it would be beneficial to operate this detector at 5-V excess bias at 77 K to exploit both the fast time response and the low NEP.

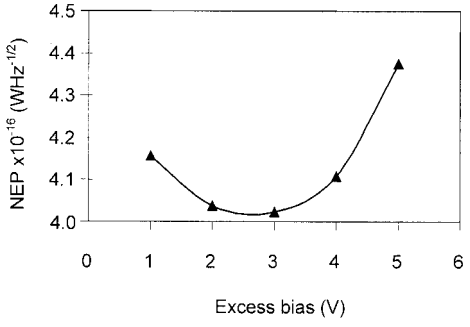

Fig. 15. NEP versus excess bias for device A at 77 K.

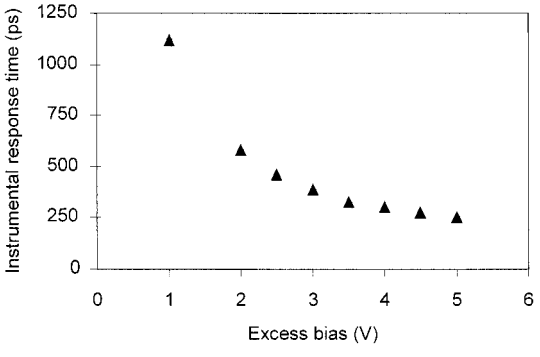

Fig. 16. Instrumental response versus excess bias for device A at 77 K.

6. Performance of Several Commercially Available Single-Photon Avalanche Diodes

The detection efficiency of several SPAD's grown with different material systems is shown in Fig. 17. The detection efficiency of the InGaAs/InP device (device A) was measured in the Geiger mode at a wavelength of 1.55 μm. To determine the spectral response, published data of the quantum efficiency[38] of this Fujitsu device were used. The detection efficiency, as stated above, is the product of this quantum efficiency and the trigger probability. There is no reason that the trigger probability, which depends on the relative excess bias, should be affected by a change in wavelength[35] in this spectral range, and so the Geiger-mode detection efficiency will be proportional to the quantum efficiency. We determined the spectral response by scaling the published quantum efficiency to the Geiger-mode detection efficiency, measured at a wavelength of 1.55 μm.

A similar method was used by Fancey[22] to determine the detection efficiency of the germanium device shown in Fig. 17. The quantum efficiency, below breakdown, was determined, and the Geiger-mode detection efficiency was measured at a wavelength of 1.3 μm. We again determined the spectral response by scaling the quantum efficiency results to the measured Geiger-mode value.

The InGaAs/InP-based detector has clearly the highest detection efficiency at a wavelength of 1.55 μm and also has a higher efficiency than the germanium device at a wavelength of 1.3 μm. The relationship between the NEP and the wavelength for these detectors is shown in Fig. 18. The InGaAs/InP SPAD has the lowest NEP at a wavelength of 1.55 μm. However, the germanium SPAD has the lowest NEP at a wavelength of 1.3 μm. It should be noted that the germanium SPAD has a diameter of 30 μm compared with the 80-μm-diameter of the InGaAs/InP device.

The Hamamatsu R5509-72 PMT, operated at 193 K, has a relatively poor detection efficiency and NEP compared with the InGaAs/InP and germanium devices. The NEP of the Hamamatsu PMT was calculated with data from the Hamamatsu data

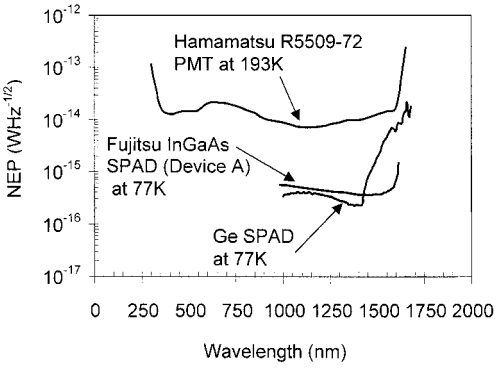

Fig. 18. NEP versus wavelength for various single-photon detectors.

sheet[20] where the dark count rate was quoted to be 200 kHz and the detection efficiency as a function of wavelength was published. The detection efficiency at 1.55 μm is only ~0.5%. However, this detector can be operated at far higher repetition rates than the InGaAs/InP and germanium devices, does not require gated biasing, and can be operated over a wider range of wavelengths.[20] As an additional comparison, an EG&G SLiK silicon SPAD has a measured NEP of 2×10^{-17} W Hz$^{-1/2}$ at a wavelength of 840 nm.[36] This is at least an order of magnitude lower than the lowest NEP of InGaAs/InP, germanium, and Hamamatsu devices; however, because of the large bandgap of silicon, such devices have detection efficiencies of $\ll$1% at wavelengths above 1.1 μm.

7. Conclusion

We have described the performance of selected, commercially available InGaAs/InP SAGM avalanche photodiodes operating above avalanche breakdown in a photon-counting mode. These devices were designed for linear multiplication at room temperature. We have related the photon-counting performance in terms of an electric field profile. We have shown the effect of punch-through and dark current on photon-counting performance and have given some indicators for selection of APD's for photon-counting operation. We have compared InGaAs/InP SPAD's to alternative photon-counting diodes and PMT's and found that the InGaAs/InP devices have the highest detection efficiency and lowest NEP at a wavelength of 1.55 μm. However, in the longer term it is clear that to make significant progress toward higher-efficiency, lower-noise, photon-counting APD's at 1.55 μm, it will be essential to design and to fabricate devices specifically for that application. These devices would exhibit clear punch through at the operating temperature, utilize small-area devices for minimal leakage, and have the maximum possible trigger probability.

The authors acknowledge many useful discussions with Steve Plimmer, John David, and Graham Rees of the Department of Electronic and Electrical Engineering, University of Sheffield, UK, and also John Rarity of the Defence Evaluation and Research Agency, Malvern, UK. The authors also acknowledge the support of the UK Engineering and Physical Sciences Research Council (GR/L8185).

References

1. A. Lacaita, P. A. Francese, F. Zappa, and S. Cova, "Single-photon detection beyond 1 μm: performance of commercially available germanium photodiodes," Appl. Opt. **33,** 6902–6918 (1994).
2. P. D. Townsend, "Quantum cryptography on multi-user optical fibre networks," Nature (London) **385,** 47–49 (1997).
3. G. S. Buller, S. J. Fancey, J. S. Massa, A. C. Walker, S. Cova, and A. Lacaita, "Time-resolved photoluminescence measurements of InGaAs/InP multiple-quantum-well structures at 1.3-μm wavelengths by use of germanium single-photon avalanche photodiodes," Appl. Opt. **35,** 916–921 (1996).
4. A. Lacaita, F. Zappa, S. Cova, and P. Lovati, "Single-photon detection beyond 1 μm: performance of commercially available InGaAs/InP detectors," Appl. Opt. **35,** 2986–2996 (1996).
5. D. V. O'Connor and D. Phillips, *Time-Correlated Single Photon Counting* (Academic, London, 1984).
6. G. Ripamonti and S. Cova, "Optical time domain reflectometry with centimetre resolution at 10-15W sensitivity," Electron. Lett. **22,** 818–819 (1986).
7. J. S. Massa, A. M. Wallace, G. S. Buller, S. J. Fancey, and A. C. Walker, "Laser depth measurement based on time-correlated single-photon counting," Opt. Lett. **22,** 543–545 (1997).
8. P. D. Townsend, "Simultaneous quantum cryptographic key distribution and conventional data transmission over installed fibre using wavelength-division multiplexing," Electron. Lett. **33,** 188–189 (1997).
9. R. J. Hughes, D. M. Alde, P. Dyer, G. G. Luther, G. L. Morgan, and M. Schauer, "Quantum cryptography," Contemp. Phys. **36**(3), 149–163 (1995).
10. H. Zbinden, H. Bechmann-Pasquinucci, N. Gisin, and G. Ribordy, "Quantum cryptography," Appl. Phys. B **67,** 743–748 (1998).
11. M. Bourennane, F. Gibson, A. Karlsson, A. Hening, P. Jonsson, T. Tsegaye, D. Ljunggren, and E. Sundberg, "Experiments on long wavelength (1550 nm) 'plug and play' quantum cryptography systems," Opt. Exp. **4,** 383–387 (1999).
12. G. Ribordy, J. Gautier, N. Gisin, O. Guinnard, and H. Zbinden, "Automated 'plug and play' quantum key distribution," Electron. Lett. **34,** 2116–2117 (1998).
13. B. F. Levine, C. G. Bethea, and J. C. Campbell, "1.52 μm room temperature photon counting optical time domain reflectometer," Electron. Lett. **21,** 194–196 (1985).
14. N. S. Rayit and L. J. Arnold, "A monomode optical time domain reflectometer using a photon counting technique," GEC J. Res. **4,** 223–227 (1986).
15. W. C. Priehorsky, R. C. Smith, and C. Ho, "Laser ranging and mapping with a photon-counting detector," Appl. Opt. **35,** 441–452 (1996).
16. S. J. Fancey, G. S. Buller, J. S. Massa, A. C. Walker, C. J. McLean, A. McKee, A. C. Bryce, J. H. Marsh, and R. M. De La Rue, "Time-resolved photoluminescence microscopy of GaInAs/GaInAsP quantum wells intermixed using a pulsed laser technique," J. Appl. Phys. **79,** 9390–9392 (1996).
17. J. S. Massa, G. S. Buller, A. C. Walker, G. Horsburgh, J. T. Mullins, K. A. Prior, and B. C. Cavenett, "Carrier recombination studies of ZnCdSe/ZnSe single quantum wells grown by molecular beam epitaxy," Appl. Phys. Lett. **66,** 1346–1348 (1992).
18. N. G. Woodard, E. G. Hufstedler, and G. P. Lafyatis, "Photon

counting using a large area avalanche photodiode cooled to 100K," Appl. Phys. Lett. **64,** 1177–1179 (1994).
19. J. Gower, *Optical Communication Systems*, 2nd ed. (Prentice-Hall, Englewood Cliffs, N.J., 1993).
20. Hamamatsu R5509–72 specification sheet (Hamamatsu, Corporation, Bridgewater, N.J., 1999).
21. G. S. Buller, J. S. Massa, and A. C. Walker, "All solid state microscope-based system for picosecond time-resolved photoluminescence measurements on II-VI semiconductors," Rev. Sci. Instrum. **63,** 2994–2998 (1992).
22. S. Fancey, "Single-photon avalanche diodes for time-resolved photoluminescence measurements in the near infra-red," Ph.D. dissertation (Heriot-Watt University, Edinburgh, UK, 1996).
23. P. C. M. Owens, J. G. Rarity, P. R. Tapster, D. Knight, and P. D. Townsend, "Photon counting with passively quenched germanium avalanche photodiodes," Appl. Opt. **33,** 6895–6901 (1994).
24. G. Ribordy, J. D. Gautier, H. Zbinden, and N. Gisin, "Performance of InGaAs/InP avalanche photodiodes as gated-mode photon counters," Appl. Opt. **37,** 2272–2277 (1998).
25. R. J. McIntyre, "Multiplication noise in uniform avalanche diodes," IEEE Trans. Electron Devices **13,** 164–168 (1966).
26. I. Umebu, A. N. M. M. Choudhury, and P. N. Robson, "Ionisation coefficients measured in abrupt InP junctions," Appl. Phys. Lett. **36,** 302–303 (1980).
27. S. R. Forrest, R. G. Smith, and O. K. Kim, "Performance of InGaAs/InP avalanche photodiodes," IEEE J. Quantum Electron. **18,** 2040–2048 (1982).
28. K. Taguchi, T. Torikai, Y. Sugimoto, K. Makita, and H. Ishihara, "Planar structure InP/InGaAsP/InGaAs avalanche photodiodes with preferential lateral extended guard ring for 1.0-1.6 micron wavelength optical communication use," J. Lightwave Technol. **6,** 1643–1655 (1988).
29. Y. Liu, S. R. Forrest, J. Hladky, M. J. Lange, G. H. Olsen, and D. E. Ackley, "A planar InP/InGaAs avalanche photodiode with floating point guard ring and double diffused junction," J. Lightwave Technol. **10,** 182–193 (1992).
30. S. R. Forrest, O. K. Kim, and R. G. Smith, "Optical response time of In(0.53)Ga(0.47)As/InP avalanche photodiodes," Appl. Phys. Lett. **41,** 95–98 (1982).
31. H. Ando, Y. Yamauchi, and N. Susa, "High-speed planar InP/InGaAs avalanche photodiode fabricated by vapour phase epitaxy," Electron. Lett. **19,** 543–545 (1983).
32. G. E. Stillman and C. M. Wolfe, "Avalanche photodiodes," in *Semiconductors and Semimetals, Vol. 12: Infrared Detectors II*, R. K. Willardson and A. C. Beer, eds. (Academic, New York, 1977).
33. R. G. W. Brown, K. D. Ridley, and J. G. Rarity, "Characterization of silicon avalanche photodiodes for photon correlation measurements. 1: Passive quenching," Appl. Opt. **25,** 4122–4126 (1986).
34. E. L. Portnoi, N. M. Stel'makh, and A. V. Chelnokov, "Characteristics of heterostructure lasers with a saturable absorber fabricated by deep ion implantation," Sov. Tech. Phys. Lett. **15,** 157–158 (1989).
35. F. Zappa, A. Lacaita, S. Cova, and P. Webb, "Nanosecond single-photon timing with InGaAs/InP photodiodes," Opt. Lett. **19,** 846–848 (1994).
36. F. Zappa, A. L. Lacaita, S. Cova, and P. Lovati, "Solid state single-photon detectors," Opt. Eng. **35,** 938–945 (1996).
37. A. Lacaita and M. Mastrapasqua, "Strong dependence of time resolution on detector diameter in single photon avalanche diodes," Electron. Lett. **26,** 2053–2054 (1990).
38. *Lightwave Semiconductors Data Book* (Fujitsu, Ltd., Tokyo, Japan, 1992).