Single photon avalanche diode for single molecule detection

Li-Qiang Li and Lloyd M. Davis
University of Tennessee Space Institute, Tullahoma, Tennessee 37388

(Received 30 November 1992; accepted for publication 27 January 1993)

A commercially available single photon avalanche photodiode in a passively quenched circuit is used with time-correlated single photon counting modules to achieve subnanosecond time response together with high quantum efficiency and low dark noise. These characteristics are required for experiments in single molecule detection and spectroscopy in which time-gated detection schemes are used. By tightly focusing the input light onto the active area, a quantum efficiency of over 50% and a single photon timing jitter of 168 ps full width at half-maximum are achieved. In addition, the full width at one-hundredth maximum, which is of greater importance for time gating, is 790 ps, comparable to that from a microchannel plate photomultiplier. Measurements of the detector dead time, and the quantum efficiency, dark counts, time response, and pulse height distribution for different operating conditions are also reported.

I. INTRODUCTION

Ultrasensitive detection and spectroscopy experiments require detectors with single photon detection capability, high quantum efficiency, and low dark noise. Also, for time resolved spectroscopy and time-gated detection schemes a detector with subnanosecond single photon timing jitter is needed. Previous experiments[1] for the detection of single dye molecules in solution have used a microchannel plate photomultiplier to achieve most of these requirements except that the quantum efficiency was only 5%–8%. However, due to recent improvements in solid state semiconductor photodetectors, considerably higher quantum efficiency, together with adequate time response and dark count rate, can now be provided by a single photon avalanche diode (SPAD).

Because of their simplicity, smaller size, low bias voltage, and ease of operation, these low light level detectors have the potential for becoming the device of choice for applications in lidar, photon correlation spectroscopy, optical range finding, optical fiber test and fault location, fluorescence lifetime measurements, ultrasensitive spectroscopy, and other applications, where the light being measured can be focused to a small spot.

The SPAD is essentially a *p-n* junction reverse biased above the breakdown voltage such that the junction electric field is sufficiently high to sustain the flow of an avalanche current triggered by a photogenerated carrier.[2–5] Thus the SPAD is like a bistable triggering circuit with a very large gain, sufficient for the detection of single photons. By contrast, an ordinary avalanche photodiode (APD) is operated below the breakdown voltage as an analog amplifier with modest gain and large dynamic range for use in applications with medium-to-high light levels.[6] To reduce thermally generated dark counts and ensure fast single photon timing, SPADs are designed with considerably smaller active areas than APDs.

The avalanche current in a SPAD continues until it is quenched by lowering the current below the latch current. Quenching may be achieved by either a passive or an active circuit. In passive quenching,[7] once the avalanche is initiated, the device capacitance, including junction and stray capacitance, discharges very quickly through the SPAD, with current limited by the SPAD space-charge resistance. The power supply then drives the avalanche current, causing the voltage to increase rapidly across a large current limiting resistor, R_Q, (typically ~ 100 kΩ), in series with the SPAD. The bias voltage across the SPAD drops below the breakdown value causing the current to drop below the latch current and the avalanche is quenched. The bias voltage on the SPAD then returns to a value above breakdown with a time constant R_QC, where C is the device capacitance. The quenching circuit should have a sufficiently fast recharge time so that the diode becomes ready for detecting another photon, but R_Q must be large enough to quench the avalanche. In active quenching,[8] fast transistors, triggered by the breakdown pulse, are used to quickly lower and then reset the bias voltage, hence producing a shorter dead time.

For time resolved single photon counting, the timing jitter, which is the statistical distribution of the difference between the photon arrival time and the electrical output of the detector, is of key importance. According to different applications, SPADs have different structures and hence different characteristics in timing and quantum efficiency. There is a trade-off between fast timing and high quantum efficiency in the design of a SPAD.[9] The timing jitter ranges from ~ 320 ps full width at half-maximum (FWHM) for the EG&G/RCA C30902S,[10] to ~ 20 ps for specially designed devices fabricated with a shallow depletion layer and very small active area (~ 5 μm) on an epitaxial substrate.[11–13] The latter value is comparable to the time resolution of the fastest microchannel plate photomultiplier tube (MCP-PMT).[14] However, these ultrafast SPADs have low quantum efficiency because the photon may pass through the very thin depletion layer without being absorbed.

In this paper, we demonstrate how to use the SPAD from a commercially available module (EG&G SPCM-200) to simultaneously provide high quantum efficiency, fast single photon timing, and reasonably low dark noise. Section II of the paper describes the implementation of

 0034-6748/93/061524-06$06.00

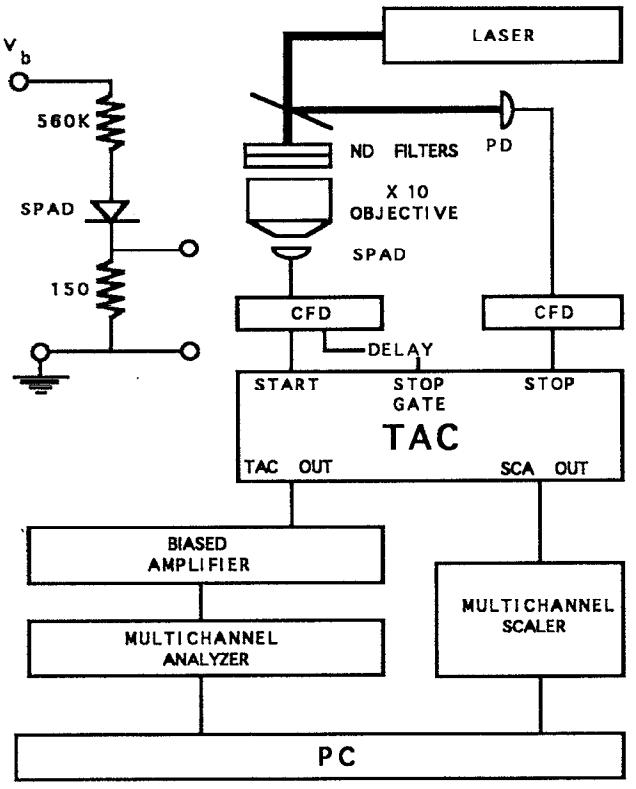

FIG. 1. Passive quenching circuit and experimental setup for time resolved single photon counting using the SPAD: PD—fast photodiode; N.D.—neutral density filters; SPAD—single photon avalanche diode; CFD—constant fraction discriminator; TAC—time to amplitude converter; SCA—single channel analyzer.

the SPAD, the time-correlated single photon counting (TCSPC) modules which are used to achieve fast timing, and the laser system and optics. In Sec. III we report the results of measurements of quantum efficiency and dark count rate vs bias voltage, the shape of the output pulses, the pulse height distribution vs bias, the dead time due to the passive quenching circuit, and the subnanosecond single photon timing jitter. In Sec. IV, we describe the use of the SPAD for time-gated detection for the experiments in single molecule detection and briefly discuss the advantages of SPADs over photomultipliers.

II. EXPERIMENTS

The tested SPAD is from the EG&G SPCM-200 photon counting module manufactured by EG&G Canada Ltd., but is custom modified by EG&G to give low stray capacitance across the SPAD and direct output pulses with lower load resistance. The SPAD is mounted on a two-stage thermoelectric cooler, in series with a 560 KΩ chip resistor for passive quenching and a 150 Ω chip resistor as a load resistor, all hermetically sealed behind a flat glass window in a TO-8 package. The reverse bias voltage, provided by a high voltage power supply (Pacific Instruments 110), is applied across all three components, from the quenching resistor through the SPAD to the grounded contact of the load resistor, as shown in the upper left of Fig. 1. In order to maintain a fast rise time in the output pulse, stray capacitance across the load resistor is minimized by soldering a RG-174, 50 Ω cable directly to the contacts on either side of the load resistor, and high-frequency connectors (Lemo series 00) are used. Two thermistors, mounted beside the SPAD and connected in parallel within the TO-8 package, are connected to a locally constructed feedback circuit and power supply for driving the thermoelectric cooler. Measurement of the current supplied to the thermoelectric cooler (Marlow Industries 2021), according to the data sheet, indicates that the SPAD temperature is -45 ± 4 °C, although the actual temperature may be somewhat higher because of the small heat load caused by the SPAD. At this temperature, the SPAD breakdown voltage is -330 V, and it increases with operating temperature by ~ 0.4 V °C^{-1}.

The experimental setup is basically a time-correlated single photon counting system as shown in Fig. 1. The output pulses from the SPAD are directed to a 2 GHz noninverting 20 dB amplifier (B&H AC2010H) and then to a quad constant fraction discriminator (CFD) (Tennelec TC455-CFD-MCP). This provides three negative NIM output signals of variable width, one of which is used to provide the *start* signal to a time-to-amplitude converter (TAC) (EG&G Ortec 567). A beamsplitter directs part of the laser excitation pulses onto a fast photodiode (HP 5082-4203) in a locally constructed circuit to give 0.7 ns FWHM pulses with an amplitude of ~ 1 V. These are conditioned by a second channel of the quad CFD, which then provides a continuous train of *stop* pulses for the TAC, each separated by only 13.2 ns. The TAC is operated over its fastest range of 50 ns and, in order to use the central region of this range to obtain best linearity, a second output signal from the CFD of the SPAD, with delay and width suitably adjusted, is used to gate the *stop* signal of the TAC.

For time-gated photon counting, the output of the TAC is directed through an internal single-channel analyzer, the output of which passes to a PC-based multichannel scaler. For measurement of the overall time spectrum, the TAC output pulses pass through a biased amplifier (Tennelec TC252) to a PC-based multichannel analyzer (EG&G Ortec Adcam 921). The system time calibration is measured to be 1.49 ps/channel with the use of an Ortec 462 Time Calibrator.

The excitation source is provided by either a mode-locked frequency doubled Nd:YAG laser (Coherent Antares 76-s) which provides 70 ps FWHM pulses at 532 nm at a repetition rate of 76 MHz, or a synchronously pumped dye laser (Coherent 702-1) operated at 584 nm to give pulses < 5 ps FWHM. The laser beam is focused onto the SPAD using a $\times 10$ microscope objective (Newport M-10X) to a spot of radius ~ 8 μm. Several calibrated neutral density filters (Oriel 50204) are placed at small tilt angles in a holder immediately before this objective, in order to attenuate the laser intensity to known values. The SPAD is mounted on a three-dimensional translation stage (Newport 462 series) to allow fine adjustment of its position relative to the focused laser beam.

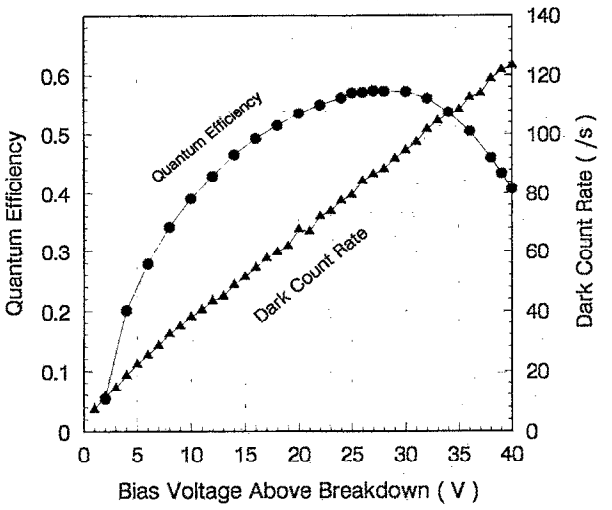

FIG. 2. Quantum efficiency and dark count rate vs bias voltage above breakdown.

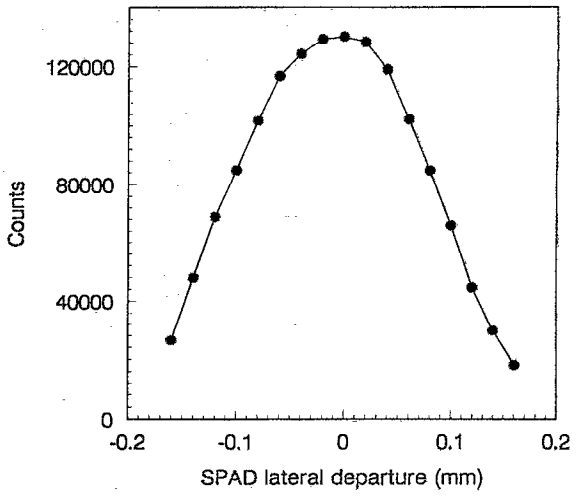

FIG. 3. Quantum efficiency vs SPAD lateral position.

III. TEST RESULTS

A. Quantum efficiency and dark count rate

The intensity of the 532 nm laser beam is measured using a precision power meter (Laser Precision RK-5720) to be 7.10 mW, and is further attenuated using neutral density (ND) filters to give a total density of ND 11.4. The throughput of the focusing objective is separately measured using the power meter to be 0.92. Thus the photon flux focused onto the SPAD is calculated to be 69 600 s^{-1}, within 10% error due mostly to the specified tolerance of the neutral density filters.

Figure 2 shows the quantum efficiency and dark count rate vs bias voltage above breakdown. The quantum efficiency, measured at a wavelength of 532 nm, peaks at a bias of ~27 V above breakdown at a value of 57%, and then decreases at higher bias voltages. The reason for this decrease is yet to be understood. From the previously reported[15] wavelength dependence of the relative quantum efficiency, the peak quantum efficiency for single photon detection at a wavelength of ~630 nm is expected to be >60%.

The dark count rate increases steadily with bias at a rate of 3 s^{-1} V^{-1}, taking a value of 80 s^{-1} for a bias of 25 V above breakdown. This value is somewhat high for some applications, but lower values of dark count may be obtained from the selected SPADs of the EG&G SPCM-200 modules. It should also be possible to make a small improvement in the dark count rate by cooling the SPAD further, although eventually charge trapping will cause afterpulsing and an increase in the dark count rate.[16]

The effective quantum efficiency depends on the focused laser beam spot size on the SPAD. Figure 3 shows that the count rate falls off rapidly if the SPAD is translated laterally by more than 75 μm, in agreement with the specified 150 μm active diameter. Also, if the SPAD is translated far enough along the beam focusing direction, the count rate again falls off but not as rapidly, since most of the expanded beam spot size remains within the active diameter. As to be expected, high quantum efficiency can be achieved only by focusing the laser beam within the active area.

B. Pulse shape

The SPAD output pulse shape is measured using a 25 ps sampling head (Tektronix S-4) and a boxcar integrator (EG&G 4402/4422), with averaging over a large number of pulses. The SPAD output pulses are directed to a 50 Ω power splitter and one output is amplified to provide the boxcar trigger while the other is delayed by about 50 ns to provide the signal to the sampling head. The result, shown in Fig. 4 has been scaled by a factor of 2 to account for the 50 Ω power splitter. The pulse amplitude depends on the bias voltage and this curve, obtained at −20 V above breakdown, shows an amplitude of 180 mV, with a 10%–90% rise time of 1.3 ns and FWHM of 2.9 ns.

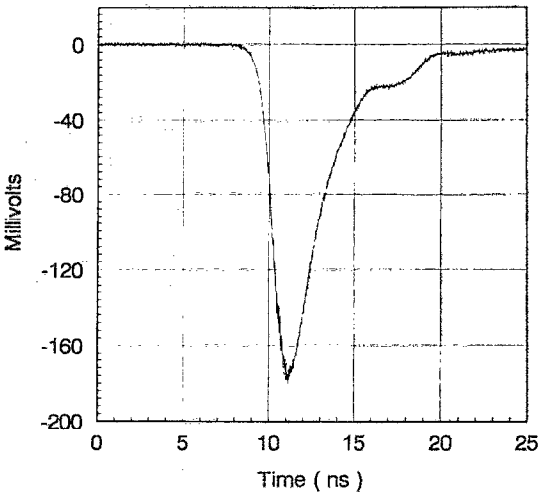

FIG. 4. SPAD output pulse at a bias of 20 V above breakdown.

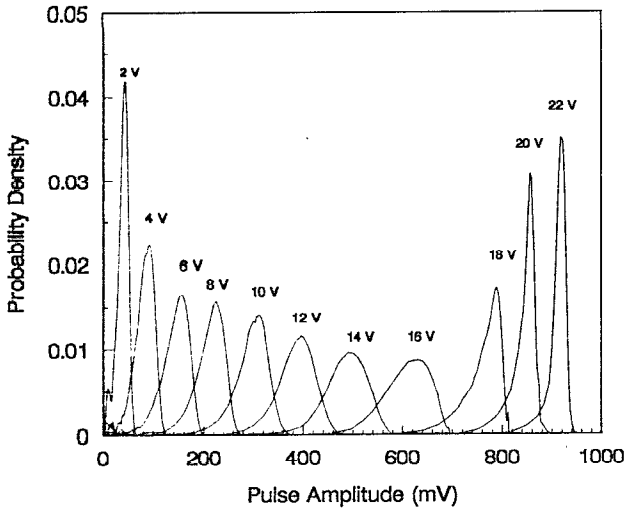

FIG. 5. Pulse height distribution of the amplified SPAD pulses as seen by the CFD, for a range of bias voltages above breakdown.

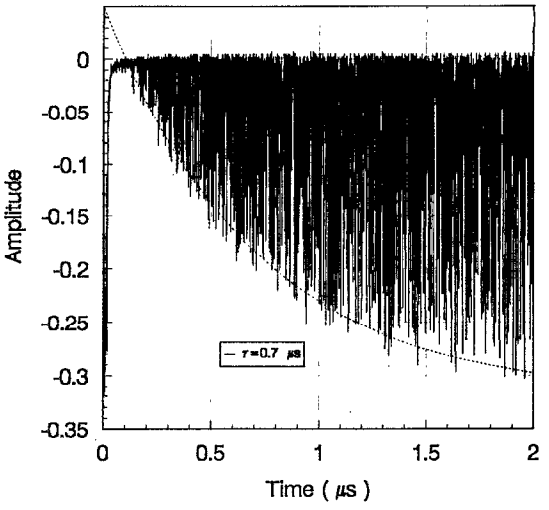

FIG. 6. SPAD dead-time measurement.

The apparent shape and amplitude of the output pulses depend on the bandwidth of the instrument used to measure the pulses. When the same pulses are viewed directly on a 250 MHz oscilloscope, they appear to have amplitudes from 90 to 110 mV, with a mean amplitude of only 105 mV. The bandwidth of the circuitry within the CFD is ~200 MHz, so that the pulses as seen by the CFD would be even smaller.

C. Pulse height distribution

Variation in the amplitude of the SPAD output pulses would give rise to significant (>200 ps) time walk at a leading edge discriminator. This is considerably reduced with the use of a CFD, which nevertheless still gives timing errors depending upon the mean and the range of the input pulse amplitudes. In our experiments, best timing is obtained when the SPAD pulses are amplified before being directed to the CFD, as discussed below in Sec. E. Since the apparent amplitudes of the pulses depend on the bandwidth of the instrument used to measure the pulses, we measure the pulse height distribution of the amplified pulses using the CFD directly. This is achieved by using a constant flux light and recording the falloff in the count rate as the threshold level of the discriminator within the CFD is increased. The derivative of the resultant error function shaped curve is then normalized to give the pulse height distribution.

Figure 5 shows the results for an increasing range of SPAD bias voltages. The pulse amplitude increases as the bias increases and at the same time the relative width of the distribution becomes narrower, particularly for bias voltages above 16 V above breakdown. This decreasing width in the distribution is mostly due to the SPAD, since the fast amplifier is not expected to saturate for input pulses less than 0.4 V (i.e., output <4 V) and the CFD for pulses less than 2 V. The pulse height distribution obtained from the SPAD is considerably narrower than that from a photomultiplier.

D. Dead time

The passive quenching circuit of the SPAD causes the bias to quickly drop below breakdown after a photon is detected so that the SPAD will not detect incoming photons until the bias recharges above breakdown, with a time constant determined by the quenching resistor and the device capacitance. The detector dead time is therefore an important parameter to know, particularly in experiments which require a high instantaneous count rate, such as the photon burst scheme employed for single molecule detection.[1]

The dead time is measured using a constant light flux which gives a mean count rate of $\sim 2\times 10^5\ s^{-1}$ by recording and overlaying a large number (96) of successive sweeps of the SPAD output, using a boxcar integrator with 2 ns sample gate and 2 μs time range. Figure 6 shows the result for a bias voltage of 25 V above breakdown. The SPAD is completely unresponsive for ~200 ns and has reduced response as the bias voltage recovers with a time constant of ~700 ns. Since the quenching resistor is known to be 560 kΩ, the device capacitance is estimated to be ~1.3 pF.

E. Single photon time resolution

For measurements using time resolved single photon detection, the statistical distribution of the time intervals between the mode-locked laser excitation pulse and the timing signal from the discriminator triggered by the SPAD output due to the detection of an incoming photon determines the overall time resolution. This is measured by employing electronics configured in a conventional TCSPC scheme[17] to collect a histogram of the time intervals.

Figure 7 shows the single photon time resolution of the SPAD measured for a bias of 20 V above breakdown using <5 ps FWHM dye laser pulses at 584 nm. The SPAD output pulses are amplified using the fast amplifier to a voltage of ~1.8 V and the walk and threshold of each CFD for the SPAD and the photodiode channels are cor-

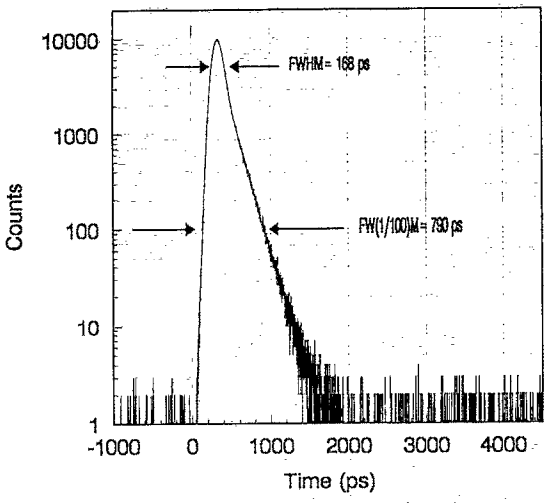

FIG. 7. SPAD single photon time resolution for operation at a bias of 20 V above breakdown, measured using <5 ps FWHM laser pulses at $\lambda=584$ nm and a fast amplifier.

rectly adjusted. A full width at one-hundredth maximum [FW(1/100)M] of 790 ps is achieved, due to the tail which has a decay constant of ~ 150 ps, while the FWHM of the prompt is 168 ps.

While tight focusing of the laser beam onto the SPAD is essential for obtaining high quantum efficiency, it is equally important for achieving fast time resolution with a short tail, as indicated by a small value of the FW(1/100)M. This can be seen in Fig. 8, which shows the dependence of the prompt function on the spot size of the light focused onto the SPAD. The spot size is separately measured by scanning a razor edge across the beam and fitting an error function to the signal collected by a large area photodiode. At the focal plane, the $1/e^2$ intensity diameter is 16 μm, which is considerably smaller than the 150 μm active area of the SPAD, and it expands to 140 μm as the SPAD is translated 0.4 mm from the focal plane. A similar improvement in the timing jitter has previously been observed[10] when the light is tightly focused onto the active area of the EG&G/RCA C30902S SPAD, and has been attributed to the generation of a lower number of carriers which diffuse slowly into the field region to give delayed avalanche pulses and an increase in the duration of the tail.[18]

We also find that the use of the 2 GHz 20 dB amplifier between the SPAD and CFD improves the overall timing jitter, probably because the CFD gives less time walk for larger amplitude input pulses. Without the amplifier, after reoptimization of the CFD threshold and walk, the time resolution gives a FWHM of 256 ps and a FW(1/100)M of 1012 ps, still adequate for many applications. Since the rise time of the SPAD output pulse is only 1.3 ns, improved timing is also expected with a slower amplifier.

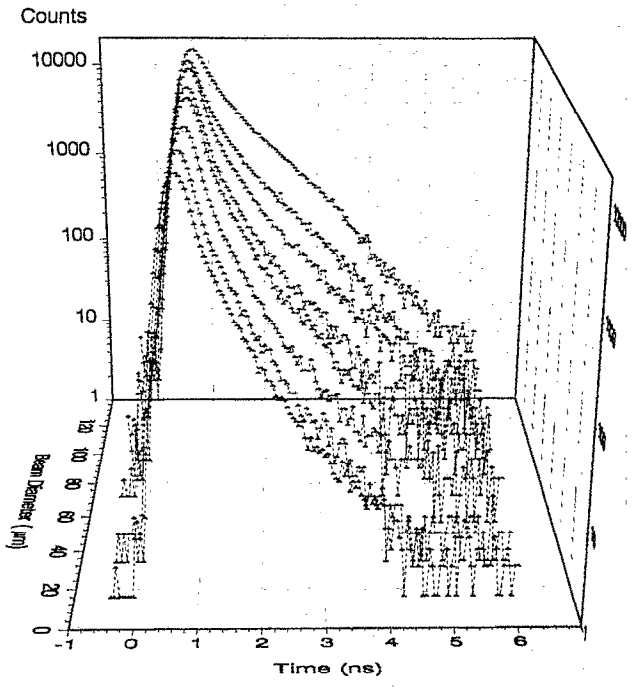

FIG. 8. Single photon time resolutions for a range of focused spot sizes on the SPAD (beginning with the curve at the front, the beam spot diameters are 16, 34, 56, 85, 96, 117, 126, and 140 μm).

IV. USE FOR SINGLE MOLECULE DETECTION

The high quantum efficiency and subnanosecond time resolution for single photon counting provided by the SPAD enables increased signal in experiments for the detection of single molecules in solution. In these experiments, single molecules of rhodamine 6G (R6G) in aqueous solution flow into a tightly focused laser beam and are cycled $\sim 2.5\times 10^4$ times before photodegradation, resulting in a small burst of detected photons. The use of a mode-locked laser and time-gated photon detection are essential for discriminating this weak single molecule fluorescence from the large number of promptly scattered Raman pho-

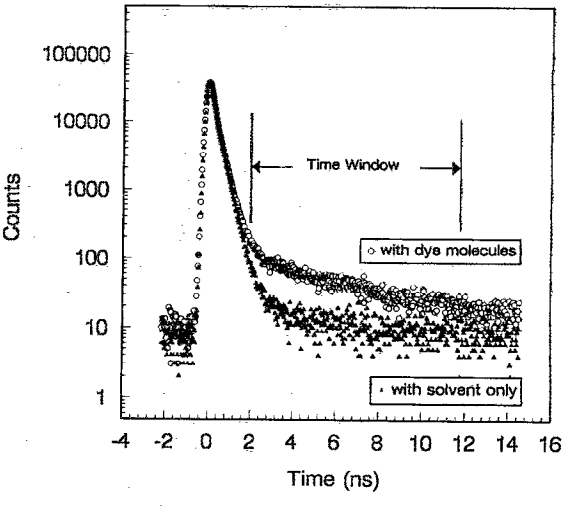

FIG. 9. Prompt function from pure solvent and fluorescence decay from ultradilute solution collected in the single molecule detection experiment. The graph illustrates the principle of time-gated detection with the time-gate window set to reduce promptly scattered photons and dark noise.

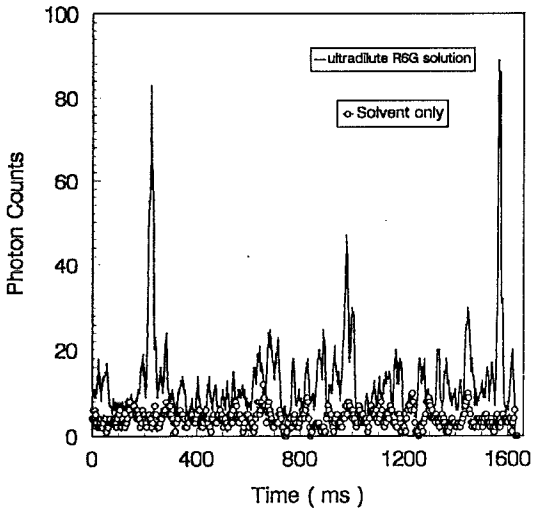

FIG. 10. Photon bursts from single molecule detection experiment, similar to Ref. 1 but using the SPAD detector with the following experimental parameters: Rhodamine 6G concentration 10^{-13} M, laser beam waist 8.8 μm, power 10 mW, solution flow rate 0.78 μm/ms. Photon bursts have not been processed using the weighted quadratic filter of Ref. 1, but are simply binned over a 12 ms interval. Molecules which pass through the center of the laser beam give bursts of $\sim$80 photons while smaller bursts are due to molecules which pass through the edges of the beam.

tons at the same wavelength originating from the $>10^{13}$ solvent molecules in the <1 pl sample volume.

In the previous experiments,[1] a MCP-PMT was employed because it provides fast single photon timing, but with a quantum efficiency of $\sim$5%, only $\sim$10 photons per molecule are detected. A GaAs photocathode side-on photomultiplier would give increased quantum efficiency of $\sim$16% but with poorer time response. While conventional PMTs have larger sensitive areas than SPADs, giving easier alignment, in single molecule detection experiments the light originates from a small sample volume, and so can be effectively focused to a small area.

In our present experiments, light is collected from the sample volume using a $\times$40 microscope objective (numerical aperture 0.65) and is spatially filtered through a 800 μm pinhole, which is then imaged to a disk with a diameter $\sim$80 μm at the center of the SPAD active area. Figure 9 shows the resultant prompt obtained using clean water, and the fluorescence decay curve collected from a 10^{-13} M solution of R6G. For applications such as single molecule detection, the tail of the prompt determines how closely the time window may be set, and hence how efficient the time-gated detection scheme will be. In this respect, the FW(1/100)M rather than the FWHM provides a better measure of the prompt for time-gated detection. The tested SPAD provides a timing jitter with FW(1/100)M comparable to that of the previously used MCP-PMT (Hamamatsu R1564U), but the quantum efficiency is about an order of magnitude higher. Consequently, in our present experiments, a single molecule of R6G which passes through the center of the laser beam gives rise to a larger burst of $\sim$80 photons, as shown in Fig. 10. This increase in the rate of detected photons during the passage of an individual molecule is an important factor in the success of future experiments for the detection of molecules at faster flow rates and also for experiments in single molecule spectroscopy.[19] Also, the enhanced response of the SPAD in the infrared region has recently enabled the detection of single infrared dye molecules in ethanol.[20]

In experiments which demand high sensitivity and subnanosecond single photon timing, such as single molecule detection, the tested SPAD can provide considerable improvement over photomultipliers, although the dark count rate is a little higher. In addition, SPADs provide the advantages of solid state over vacuum-tube devices: small size, low bias voltage, and cost effectiveness.

ACKNOWLEDGMENTS

This work has been supported by a grant from Whitaker Foundation. We are grateful to S. I. Soltesz (S. I. S.) and C. J. Trottier from EG&G Canada for the modification and loan of the SPAD and we also thank S. I. S., R. J. McIntyre, and E. B. Shera for helpful discussions.

[1] E. B. Shera, N. K. Seitzinger, L. M. Davis, R. A. Keller, and S. A. Soper, Chem. Phys. Lett. **174**, 553 (1990).
[2] S. Cova, A. Longoni, A. Andreoni, and R. Cubeddu, IEEE J. Quantum Electron. **QE-19**, 630 (1983).
[3] T. E. Ingerson, R. J. Kearney, and R. L. Coulter, Appl. Opt. **22**, 2013 (1983).
[4] S. Cova, A. Longoni, and A. Andreoni, Rev. Sci. Instrum. **52**, 408 (1981).
[5] T. Louis, G. H. Schatz, P. Klein-Bölting, A. R. Holzwarth, G. Ripamonti, and S. Cova, Rev. Sci. Instrum. **59**, 1148 (1988).
[6] M. Noble, Lasers Optronics **10**, 32 (1991).
[7] R. G. W. Brown, K. D. Ridley, and J. G. Rarity, Appl. Opt. **25**, 4122 (1986).
[8] R. G. W. Brown, R. Jones, J. G. Rarity, and K. D. Ridley, Appl. Opt. **26**, 2383 (1987).
[9] A. Lacaita and M. Mastrapasqua, Electron. Lett. **26**, 2053 (1990).
[10] M. Ghioni and G. Ripamoni, Rev. Sci. Instrum. **62**, 163 (1991).
[11] A. Lacaita, M. Ghioni and S. Cova, Electro. Lett. **25**, 841 (1989).
[12] G. S. Buller, J. S. Massa, and A. C. Walker, Rev. Sci. Instrum. **63**, 2994 (1992).
[13] S. Cova, A. Lacaita, M. Ghioni, G. Ripamoni, and T. A. Louis, Rev. Sci. Instrum. **60**, 1104 (1989).
[14] H. Kume, K. Koyama, K. Nakatsugawa, S. Suzuki, and D. Fatlowitz, Appl. Opt. **27**, 1170 (1988).
[15] EG&G SPCM-200-PQ Data Sheet, 1991.
[16] A. Lightstone, R. J. McIntyre, and P. P. Webb (RCA internal communication).
[17] D. V. O'Connor and D. Phillips, *Time-Correlated Single Photon Counting* (Academic, London, 1984).
[18] G. Ripamoni and S. Cova, Solid State Electron. **28**, 925 (1985).
[19] S. A. Soper, L. M. Davis and E. B. Shera, J. Opt. Soc. Am. B **9**, 1761 (1992).
[20] S. A. Soper and Q. Mattingly, Proc. SPIE (to be published).