A picosecond time-resolved photoluminescence microscope with detection at wavelengths greater than 1500 nm

J. M. Smith,[a] P. A. Hiskett, I. Gontijo,[b] L. Purves, and G. S. Buller
Department of Physics, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, United Kingdom

(Received 1 December 2000; accepted for publication 19 February 2001)

We describe a picosecond resolution time-resolved photoluminescence microscope with high detection sensitivity at wavelengths extending beyond 1500 nm. The instrument performs time-correlated single photon counting using an InGaAs/InP single photon avalanche diode as a detector, and provides temporal resolution of less than 300 ps (full width at half maximum) and spatial resolution down to 4 μm at a sample temperature between 4 and 300 K. Analysis of noise characteristics indicates the ability to measure the excess carrier lifetimes of semiconductor devices with excited carrier densities of less than $10^{14}\,\mathrm{cm}^{-3}$. © *2001 American Institute of Physics.*
[DOI: 10.1063/1.1366635]

I. INTRODUCTION

Time-resolved photoluminescence (TRPL) measurements allow the investigation of electronic relaxation mechanisms in optically active materials. In semiconductors, for instance, the luminescence intensity depends upon the excess carrier density, which reduces on a time scale ranging from picoseconds to microseconds, as a result of intraband carrier dynamics and various interband recombination processes. Understanding of the dominant processes in the relaxation is of great value in the characterization of material quality and design of new devices.

Time-correlated single photon counting (TCSPC)[1] using single photon avalanche diodes (SPADs) offers TRPL measurement of unrivaled sensitivity and with picosecond time resolution, but has until now been limited to detection wavelengths shorter than 1500 nm, this being the long wavelength limit of sensitivity for Ge homojunction devices.[2,3] The low-loss telecommunications window between wavelengths of 1500 and 1600 nm, in which most long haul telecommunications devices are designed to operate, has therefore not been accessible, except by less sensitive methods such as pump-probe[4] or up-conversion detection.[5]

As a result of extensive investigations into the use of commercial InGaAs/InP separate absorption, grading, and multiplication layer (SAGM) avalanche photodiodes as photon counters,[6] we have now identified devices which can be operated in photon counting mode to offer noise equivalent powers (NEPs) as low as $3\times10^{-17}\,\mathrm{W\,Hz}^{-1/2}$ at an operating temperature of 77 K, combined with timing resolution of less than 300 ps. We have incorporated one such detector into a TRPL microscope, and performed preliminary measurements on a semiconductor device structure that indicates sensitivity to excess carrier densities as low as $10^{14}\,\mathrm{cm}^{-3}$.[7]

This article describes in detail the microscope hardware, the measurement method, and key performance issues. It provides a more detailed, general analysis of the sensitivity of TCSPC to low light levels, and derives some results that are both helpful in optimizing this instrument, and which indicate that sensitivity to signal intensities even lower than those demonstrated previously[7] are measurable.

II. APPARATUS AND OPERATION

A schematic diagram of the TRPL apparatus is shown in Fig. 1. For ease of viewing, the electrical configuration primarily occupies the left-hand side of the diagram and the optical configuration the right.

The pulsed optical source is a passively *Q*-switched InGaAsP/InP semiconductor diode laser emitting at a wavelength of 1305 nm. The *Q*-switching mechanism is based on an efficient saturable absorber formed by heavy ion bombardment, which permits reliable production of optical pulses of duration ~15 ps and pulse energy of ~5 pJ.[8] Such lasers have been produced for operation at numerous other wavelengths, including 800 and 1450 nm. The excitation laser diode is driven by an Avtech AVO-9 electrical pulse generator, which also provides a synchronized trigger for a second pulse generator unit, a Hewlett Packard HP81110A, to supply both the ''start'' pulse to the time-to-amplitude converter (TAC) via a constant fraction discriminator (CFD) and a −5 V gate pulse to bias the cooled InGaAs/InP SPAD detector into the Geiger mode. The SPAD is otherwise held at a dc bias just below its breakdown voltage (approx 30 V at 77 K). A variable time delay in the Avtech pulse generator is adjusted to ensure that the optical signal arrives at the detector during the gated-on period.

The optics for this microscope are mounted on a purpose-built baseplate, which can be positionally adjusted, using micrometer precision actuators, relative to the sample, which is fixed. This slotted baseplate technology was developed for previous experiments in optical interconnects[9] and supports the stable positioning of optical components mounted in cylindrical modules. Its ability for rapid optical reconfiguration makes this instrument ideal for the study of a

[a]Electronic mail: j.m.smith@hw.ac.uk
[b]Current address: Department of Electrical Engineering, UCLA, Los Angeles, California 90095-1594.

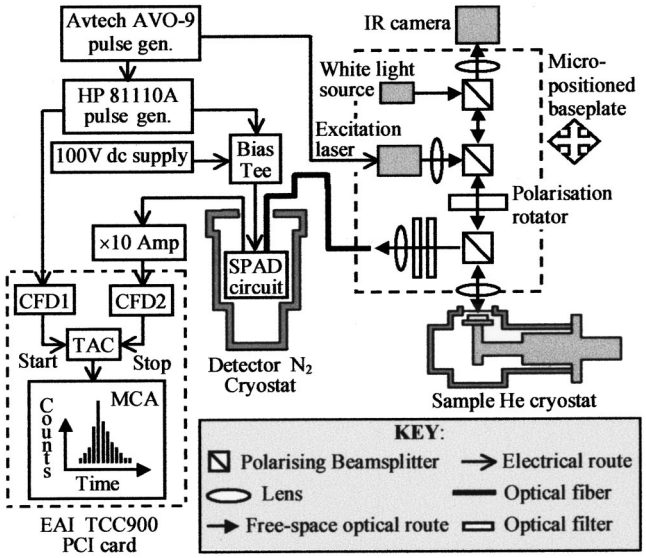

FIG. 1. Schematic diagram of the TRPL microscope. For ease of viewing, the electrical and optical configurations occupy the left- and right-hand sides of the diagram, respectively.

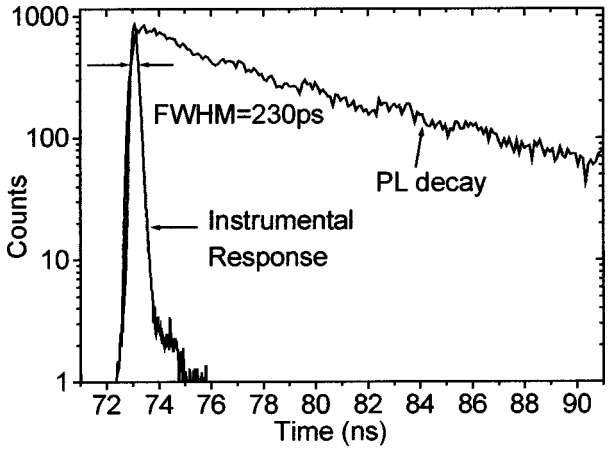

FIG. 2. Instrumental response ($\lambda_{PL}=1305$ nm) and sample PL decay from an InGaAsP semiconductor heterostructure ($\lambda_{PL}=1525$ nm) using the TRPL microscope.

wide range of samples, whereby frequent interchange between different excitation sources, passive optical components, and detectors may be required.

Polarizing beamsplitters coated for 1200–1600 nm wavelength operation and a 90° polarization rotator combine to direct the collimated excitation pulse towards the sample, and it is focused onto the sample surface using one of a selection of microscope objectives. The same lens is used to collect the (now unpolarized) luminescence, which is then filtered spectrally using interference filters [~50% peak transmission, 12 nm full width at half maximum (FWHM)] and attenuated as necessary, and focused into a single mode optical fiber for delivery to the detector. The choice of objective lens used at the sample depends upon the desired spatial resolution and collection efficiency, and upon whether the sample is to be placed in the continuous flow, vacuum loading cryostat (Oxford Instruments CF1104) in which case the minimum working distance is limited to about 5 mm. Using a 36× reflecting objective lens with a numerical aperture (NA) of 0.5 we can currently obtain 4 μm spatial resolution under these conditions.

The fiber into which the luminescence is focused couples to the single mode pigtail of the SPAD, which feeds into the top of an Oxford Instruments Optistat DN exchange gas cryostat containing the detector. This allows control of the detector temperature T_{SPAD}, a parameter that is critical in determining its noise characteristics and maximum operation frequency, down to 77 K.

Prior to the luminescence measurement, alignment of the instrument is simplified by shining light through the collection fiber in the reverse direction and viewing both the resultant image of the detection aperture and the excitation spot on the sample surface using an infrared camera. This step has the additional advantage of revealing directly the relative sizes of the excitation and detection areas, which can be important in the avoidance of carrier diffusion effects in the time-resolved measurements.

As well as being gated by the HP pulse generator, the SPAD is passively quenched using a 100 kΩ resistor also mounted inside the cryostat to minimize the flow of current through the detector upon avalanche breakdown. This serves to limit the filling of deep traps in the semiconductor material, from which subsequent carrier release is thought to be responsible for triggering unwanted ''afterpulses'' during the next detector gate.[10]

The voltage across a further 50 Ω resistor in series with the SPAD and quench resistor is amplified and used as the ''stop'' signal of the TAC via a second CFD. The amplitude of the voltage spikes that occur across the 50 Ω resistor due to the fast transients in the detector gating is reduced below that of the avalanche events by setting the gate rise time to 20 ns. It is the ability to adjust the gate transient time that earns the HP81110A its place in this apparatus. Having established the minimum rise time for successful operation, it should be possible to replace the HP81110A with a more basic pulse generator combined with a simple low-pass filter to achieve the same result.

The CFDs, TAC, and multichannel analyzer (MCA) used to generate the data histogram are conveniently integrated on an Edinburgh Instruments TCC900 data acquisition card, running on Windows in a PC.

III. TEMPORAL RESOLUTION

The instrumental response of the complete microscope, with a detector temperature of 77 K and excitation repetition rate of 1 kHz, is shown in Fig. 2. This histogram was acquired by placing a mirror in place of the sample and recording photons from the reflected (and suitably attenuated) excitation pulse, and so the wavelength of light being detected is that of the excitation laser, 1305 nm. Also shown in the figure is a portion of an example TRPL histogram measured from an InGaAsP multiple quantum well heterostructure at a detection wavelength of 1525 nm. The decay takes the form of a single exponential superimposed with both Poissonian and temporally correlated noise, the origins of which will be discussed in the next section.

The key feature of the instrumental response is the full width at half maximum (FWHM) of 230 ps. This is determined principally by the timing jitter of the SPAD, since the combined jitter of the other timing components amounts to at most a few tens of picoseconds. The physical origin of the SPAD jitter is a matter of debate,[11,12] but it is a weak function of the wavelength of light to be detected as a result of the absorption spectrum of the InGaAs layer, with 1550 nm light increasing the jitter to 270 ps. Iterative reconvolution of the measured luminescence data from the instrumental response can in practice facilitate the fitting of decay lifetimes shorter than the FWHM. While the maximum extent to which this can be achieved is subject to question,[1] a number of authors have demonstrated satisfactory fitting to data with decay lifetimes of around one fifth of the instrumental FWHM,[13,14] and on this basis we estimate the ultimate timing resolution of this microscope, under ideal conditions, to be 60 ps.

This resolution is, however, an upper limit when the influence of noise on the signal is negligible. This is rarely the case, and indeed is often undesirable since such a situation generally only occurs at the expense of long measurement times. For this reason, it is better to approach each measurement individually: by knowing what accuracy is required, and with some *a priori* knowledge of the photoluminescence decay characteristics, we can determine, and to some extent minimize, the length of time that each measurement will take to perform.

To this end, the following section contains a description of the origins and manifestations of noise in these measurements, and of how this understanding can be used to optimize instrumental performance. Note that much of what follows is applicable to any TRPL measurement using TCSPC, but that it is of particular importance here due to the limitations placed on the operating frequency of this detector.

IV. SENSITIVITY AND NOISE ANALYSIS

The TCSPC histogram is the aggregate of all the temporally correlated signal counts and the uncorrelated ''background'' counts registered during the acquisition time t_{acq}. Temporally uncorrelated background counts originate mostly from two sources; stray photons incident on the detector and ''dark'' counts, which are avalanche events caused not by photogenerated carriers, but rather by carriers that have been generated thermally or by trap release in the detector.

In the hypothetical case of a noiseless histogram, the background counts are equal in number in each channel, and so can be simply subtracted from the acquired data to leave only the signal. Such a simple treatment is in practice hindered by two factors: temporally correlated noise, manifested as inhomogeneity in the detection probability across the gate and Poissonian noise. The latter is a result of the stochastic nature of TCSPC and is equal in magnitude to the square root of the number of counts in a given channel. The former can result from a variety of effects; a nonuniform gate voltage, electromagnetic pickup in the detection circuit, and nonlinearity in the data acquisition hardware being the most likely causes. The detection probability function, or gating function, can be measured directly by illuminating the gated SPAD with temporally uncorrelated photons and acquiring a large enough data set that Poissonian noise, as a fraction of the total number of counts, may be neglected. An example of such a data set, for a gate of width ~200 ns, is shown in Fig. 3.

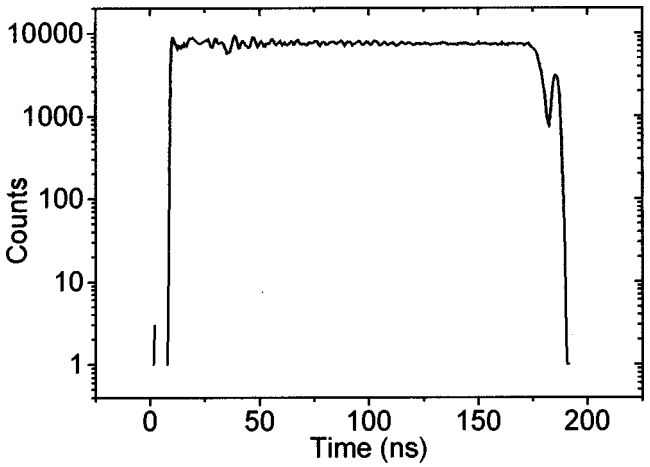

FIG. 3. A typical gating function data set with a count level of $\sim 10^4$ and thus a relative Poissonian noise amplitude of 1%. A gate duration of 200 ns has been used with a gate transient time of 20 ns. The earlier part of the gate is seen to be less uniform, due to post-transient ringing in the detector circuit.

Each subsequent data set can be normalized by dividing through by the gating function. To follow this process religiously is somewhat arduous, however, since for best results the gating function must be measured directly prior to each set of measurements under a given set of conditions (detector temperature, gate width, MCA channel width, etc.), and can take some time to acquire. It is clearly preferable to take care in generating a flat, uniform gating function when constructing the apparatus, so that all but the most demanding of measurements can be performed without the need for this adjustment.

Even when the above precautions are taken, the gate normalization procedure does not in practice account for all of the temporally correlated noise, due to slowly varying elements such as drift of the excitation laser and in the data acquisition chain. What remains is a reduced temporally correlated noise component, with an amplitude that can be represented as a constant fraction α of the count level in a given channel. In our preliminary investigations using the apparatus described, α is approximately 0.1. It is hoped that this value can be reduced significantly in our apparatus with the adoption of an optical trigger to start the TAC, which will eliminate from the time measurement circuit the drift between the optical pulse and the electrical trigger within the laser pulse generator. Obtaining a small value of α is important, since, as we shall see in the following analysis, it alone determines the ultimate sensitivity of our measurements.

We start our noise analysis by defining the signal-to-noise ratio (SNR) in the ith channel of the MCA as the ratio of the number of signal counts to the sum of the Poissonian noise and the temporally correlated noise in that channel:

$$\mathrm{SNR}^i = \frac{N^i_{\mathrm{sig}}}{\sqrt{N^i_{\mathrm{sig}} + r_{\mathrm{bkgnd}}\delta t n_{\mathrm{rep}}} + \alpha(N^i_{\mathrm{sig}} + r_{\mathrm{bkgnd}}\delta t n_{\mathrm{rep}})}, \quad (1)$$

where r_{bkgnd} is the normalized background count rate, δt is the time width per MCA channel (indeed $r_{\mathrm{bkgnd}}/\delta t$ is, by definition of r_{bkgnd}, the probability of a background count occuring in a single time window of duration δt), and n_{rep} is the number of pulse repetitions performed in the measurement.[15]

The number of signal counts in the ith channel of the MCA can be expressed as a fraction, β^i, of the peak value. For a single exponential decay with a lifetime τ_{decay} that is slow compared with the instrumental response time, the resulting expression is

$$N^i_{\mathrm{sig}} = \beta^i n_{\mathrm{rep}} P_{\mathrm{phot}} \frac{\delta t}{\tau_{\mathrm{decay}}}, \quad (2)$$

where P_{phot} is the probability of measuring a signal photon in any given pulse. For a decay that is faster than the instrumental response FWHM, and therefore requires reconvolution analysis, a good approximation is achieved by replacing δt with the FWHM in Eq. (2).

Equation (1) illustrates the fact that as data is acquired, Poissonian noise reduces relative to the temporally correlated components, and in the limit of large n_{rep} the temporally correlated noise is dominant. Since no more than a factor of 2 in the SNR is to be gained by continuing to accumulate data beyond the point at which the two noise components are equal in magnitude, we choose this condition to provide a practical indication of the achievable sensitivity, and the number of pulsed excitations required to attain it. Using Eqs. (1) and (2), the SNR is thus given by

$$\mathrm{SNR}^i_{\mathrm{inst}} = \frac{\beta^i P_{\mathrm{phot}}}{2\alpha(\beta^i P_{\mathrm{phot}} + r_{\mathrm{bkgnd}}\tau_{\mathrm{decay}})} \quad (3)$$

and the corresponding number of pulsed excitations is

$$n_{\mathrm{rep}} = \frac{\tau_{\mathrm{decay}}}{\alpha^2 \delta t(\beta^i P_{\mathrm{phot}} + r_{\mathrm{bkgnd}}\tau_{\mathrm{decay}})}. \quad (4)$$

Note that the SNR according to Eq. (3) cannot exceed $1/2\alpha$. The inherent difficulty in fitting decay lifetimes to high accuracy as a result of this limit must be compensated for by fitting over a broad range of β. More quantitatively, the attainable accuracy of the fit to data spanning a given range of β is approximately $\beta_{\min}/(\mathrm{SNR}_{\min}\times\beta_{\max})$, where $\beta_{\min}$ and $\beta_{\max}$ define the range of β and $\mathrm{SNR}_{\min}$ is the SNR at $\beta_{\min}$. For example, a SNR of 1 at $\beta=0.1$ would enable a decay time constant to be fitted to the peak of a signal with an accuracy of 10%. It should be noted that, in the case of a decay that is fast compared with the instrumental response and therefore requires deconvolution analysis, significantly higher SNR, or fitting over a correspondingly greater range of β would be required to achieve a similar accuracy.

With the above result in mind, Eq. (3) permits an estimate of the minimum signal photon detection probability for which such accuracy can be obtained. For an example situation of a monomolecular decay with $\tau_{\mathrm{decay}}=10\,\mathrm{ns}$, and using $r_{\mathrm{bkgnd}}=100\,\mathrm{s}^{-1}$, the resulting photon detection probability is

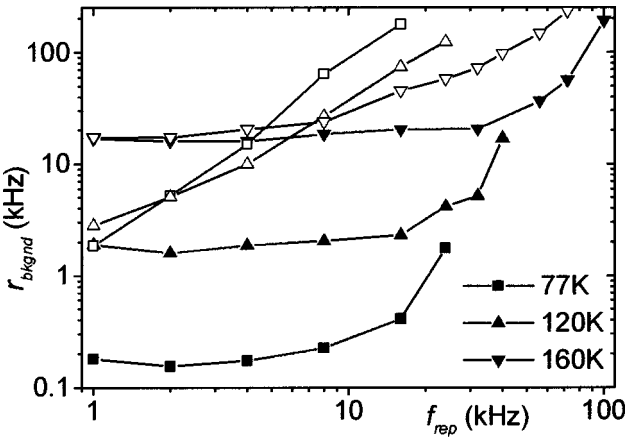

FIG. 4. Plot of r_{bkgnd} against and f_{rep} for $T_{\mathrm{SPAD}}=77$, 120, and 160 K. Closed shapes are with the detector unlit and open shapes are with the detector illuminated to give a total TAC ''stop'' probability of 5%.

$P_{\mathrm{phot}}=2.5\times10^{-6}$. This value compares favorably with the value of $P_{\mathrm{phot}}=10^{-5}$ deduced from the simpler model in Ref. 7. In an undoped semiconductor sample of planar geometry, exhibiting monomolecular photoluminescence decay, we estimate that such a detection probability would result from an excited carrier density of about $2.5\times10^{13}\,\mathrm{cm}^{-3}$.[16]

Equation (4) can be used to estimate the length of time taken to perform a given measurement, t_{acq}, by dividing through by the pulse repetition frequency f_{rep}. However, the background count rate r_{bkgnd} is strongly dependent on f_{rep} due to the afterpulsing effect described earlier. Figure 4 shows this dependence for $77\,\mathrm{K}\leqslant T_{\mathrm{SPAD}}\leqslant 160\,\mathrm{K}$ under two different illumination conditions. The first of these, shown in closed shapes, was measured with the detector in the dark to represent the weak signal limit; the second, shown in open shapes, was measured with a total detection probability per gate pulse of 0.05, which we have chosen to represent the allowable strong signal limit, pursuant to the avoidance of pulse pileup.[1]

In the weak signal limit at a detector temperature of 77 K, Fig. 4 shows that afterpulsing starts to become important at a pulse repetition frequency of around 5 kHz. At frequencies or temperatures higher than this, the ultimate sensitivity of the instrument will be reduced below that discussed above, correspondent with the increased r_{bkgnd} due to afterpulsing or thermally generated counts, respectively. Note that at a fixed f_{rep}, afterpulsing is reduced as T_{SPAD} is increased, a characteristic attributable to the thermionic nature of trap emission. The competition between this effect and the increased rate of thermally generated counts results in an optimum temperature that minimizes r_{bkgnd} for a given f_{rep}.

Estimating $\delta t=1\,\mathrm{ns}$ as the longest suitable MCA bin time width for measuring a 10 ns decay to $\beta=0.1$ with 10% accuracy, the shortest time in which the above measurement can be performed is estimated to be $1.6\times10^5\,\mathrm{s}$, or 44 h. While this may be regarded as an inordinately long time for a measurement, it is encouraging to note that by increasing P_{phot} by a factor of 4 to 10^{-5} (corresponding to a semiconductor excess carrier density of $10^{14}\,\mathrm{cm}^{-3}$), Eqs. (1) and (2)

predict that the same 10% accuracy in lifetime fitting can be achieved with t_{acq} reduced to 1.7 h.

For signal intensities larger than that corresponding to $P_{phot}=10^{-5}$, significant gains in measurement speed can be achieved by increasing the detector temperature. As a rule of thumb, t_{acq} is minimized by setting T_{SPAD} to maximize the f_{rep} that gives a background count level equal to the signal count level:

$$r_{bkgnd}=\frac{\beta^i P_{phot}}{\tau_{decay}}. \quad (5)$$

For example, if $P_{phot}=10^{-3}$ (and all other parameters are unchanged), the shortest acquisition time corresponds approximately to $r_{bkgnd}=10$ kHz, whereby the fastest repetition frequency according to Fig. 4 is approximately 40 kHz at $T_{SPAD}=120$ K (giving $t_{acq}\approx 8$ s). Note that afterpulsing characteristics may vary greatly from detector to detector, and so Fig. 4 should be generated afresh for a given SPAD.

The SPAD detection efficiency η_{SPAD} is about 8% at $T_{SPAD}=77$ K and $\lambda_{PL}=1550$ nm. A complete calibration of the detection efficiency spectrum of these detectors has not yet been performed, but they are expected to behave similarly in this respect to other InGaAs/InP SAGM devices, in which η_{SPAD} falls off rapidly at optical wavelengths longer than the $In_{0.53}Ga_{0.47}As$ absorption edge.[10] As regards the short wavelength limit, a substantial increase in timing jitter to ~1 ns observed for $\lambda_{PL}<1000$ nm is attributed to the onset of absorption in the *n*-type InP contact layer, characteristic of backillumination of a SAGM device.

ACKNOWLEDGMENTS

This work is funded by the UK Engineering and Physical Sciences Research Council (ref GR/L81895) and by the European Commission under the EQUIS project (IST-1999-11594). The authors would also like to thank Dawn Marshall and Alan Miller of St. Andrews University for supplying test samples and Paul Townsend of Corning Research Center (Ipswich, UK) for invaluable discussions.

[1] D. V. O'Connor and D. Phillips, *Time-Correlated Single Photon Counting* (Academic, London, 1984).
[2] A. Lacaita, S. Cova, F. Zappa, and P. A. Francese, Opt. Lett. **18**, 75 (1993).
[3] G. S. Buller, S. J. Fancey, J. S. Massa, A. C. Walker, S. Cova, and A. Lacaita, Appl. Opt. **35**, 916 (1996).
[4] R. Takahashi, Y. Kawamura, T. Kagawa, and H. Iwamura, Appl. Phys. Lett. **65**, 1790 (1994).
[5] A. D. Güçlü, C. Rejeb, R. Maciejko, D. Morris, and A. Champagne, J. Appl. Phys. **86**, 3391 (1999).
[6] P. A. Hiskett, G. S. Buller, J. M. Smith, A. Y. Loudon, I. Gontijo, A. C. Walker, P. D. Townsend, and M. J. Robertson, Appl. Opt. **39**, 6818 (2000).
[7] J. M. Smith, P. A. Hiskett, and G. S. Buller, Opt. Lett. (accepted).
[8] Z. I. Alferov, A. B. Zuravlev, E. L. Portnoi, and N. M. Stel'makh, Sov. Tech. Phys. Lett. **12**, 452 (1996).
[9] F. B. McCormick, F. A. P. Tooley, T. J. Cloonan, J. L. Brubaker, A. L. Lentine, R. L. Morrison, S. J. Hinterlong, M. J. Herron, S. L. Walker, and J. M. Sasian, Appl. Opt. **31**, 5431 (1992).
[10] A. Lacaita, F. Zappa, S. Cova, and P. Lovati, Appl. Opt. **35**, 2986 (1996).
[11] A. L. Lacaita, S. Cova, A. Spinelli, and F. Zappa, Appl. Phys. Lett. **62**, 606 (1993).
[12] A. Spinelli and A. L. Lacaita, IEEE Trans. Electron Devices **44**, 1931 (1997).
[13] T. A. Louis, G. Ripamonti, and A. Lacaita, Rev. Sci. Instrum. **61**, 11 (1990).
[14] J. S. Massa, G. S. Buller, A. C. Walker, J. Simpson, K. A. Prior, and B. C. Cavenett, Appl. Phys. Lett. **64**, 589 (1994).
[15] A similar treatment of the SNR has been used for analysis of TCSPC in a time-of-flight ranging system; see S. Pellegrini *et al.*, Meas. Sci. Technol. **11**, 712 (2000).
[16] The parameters used in this estimate are radiative recombination coefficient $=10^{-10}$ cm^3 s^{-1}; nonradiative recombination lifetime=10 ns; sample volume detected $=3\times10^{-11}$ cm^3; refractive index of semiconductor=3.5; objective lens NA=0.4; and SPAD detection efficiency=8%.