A low-noise, single-photon avalanche diode in standard 0.13 μm complementary metal-oxide-semiconductor process

Ryan M. Field,[1] Jenifer Lary,[2] John Cohn,[2] Liam Paninski,[3] and Kenneth L. Shepard[1,a)]
[1]*Department of Electrical Engineering, Columbia University, New York, New York 10027, USA*
[2]*IBM STG, Essex Junction, Vermont 05452, USA*
[3]*Department of Statistics, Columbia University, New York, New York 10027, USA*

(Received 10 September 2010; accepted 29 October 2010; published online 24 November 2010)

We present the design and characterization of a single-photon avalanche diode (SPAD) fabricated with a standard 0.13 μm complementary metal-oxide-semiconductor process. We have developed a figure of merit for SPADs when these detectors are employed in high frame-rate fluorescent lifetime imaging microscopy, which allows us to specify an optimal bias point for the diode and compare our diode with other published devices. At its optimum bias point at room temperature, our SPAD achieves a photon detection probability of 29% while exhibiting a dark count rate of only 231 Hz and an impulse response of 198 ps. © *2010 American Institute of Physics*.
[doi:10.1063/1.3518473]

Fluorescent dyes have become essential markers for observing and quantifying biological processes. Traditionally, fluorescence microscopy has been performed using spectrally discriminated intensity measurements. Recently, fluorescence lifetime imaging microscopy (FLIM), which measures the rate of decay of fluorophore emission after a pulsed excitation, has become a new imaging modality,[1] exploiting the sensitivity of a dye's lifetime to chemical and physical environment, including the proximity of secondary dyes through fluorescence resonance energy transfer.[2] The most common sensors in FLIM are the photomultiplier tube or discrete single-photon avalanche diodes (SPADs). These devices are used in a time-correlated single-photon counting (TCSPC) mode in which arrival time histograms are recorded through time-to-digital conversion of photon-activated pulses from the detectors. Recently, there has been work to integrate SPADs into a complementary metal-oxide semiconductor (CMOS) processing to create arrays of detectors that would allow for higher frame rates with wide-field imaging.[3–7]

SPAD detection limits are determined by noise in the form of the device's dark count rate (DCR). DCR is dominated by avalanche events that are triggered by the thermal generation of carriers from recombination-generation (RG) centers within a diffusion length of the multiplication region of the SPAD. SPADs fabricated in high-voltage processes with local-oxidation-of-silicon-based (LOCOS-based) isolation have achieved DCRs as low as a few hundred hertz.[4,8] At technology nodes below 0.35 μm, the shallow trench isolation (STI) that is used to separate devices creates a relatively defect-rich interface and a significant source of RG traps. SPAD designs in which the STI impinges the detector junction have high DCRs.[9,10] This can be mitigated somewhat by hydrogen passivation, as is done to reduce dark current in imager processes.[11,12]

In our design, we instead isolate the STI from the junction while maintaining the desired structure for a SPAD. We accomplish this without any process modifications, carefully repurposing design layers of an established process. The primary design masks used to create our device are drawn in Fig. 1(a). Our designed SPAD is illustrated in Fig. 1(b) along with a process simulation using Synopsys' Sentaurus in Fig. 1(c). The active layer (RX) limits the presence of STI around the multiplication region. Because typical field-effect transistor (FET) devices define the active area to be a source, drain, or gate, it is necessary to use the BN and BP layers to control the n+ and p+ implants. BN defines the p+ implant region and BP blocks the n+ implants. The PI layer, used to define a triple well for isolated p-type FETs, instead has been used here to generate a p-type guard ring that prevents edge breakdown. The fabricated device has an octagonal photosensitive area with a diagonal of 5 μm. The measured reverse bias breakdown voltage (V_{br}) is -12.13 V and the multiplication region has a simulated width of 115 nm at this bias. A micrograph of the fabricated device structure is shown in Fig. 2(a).

For photon counting, the diode is operated in Geiger mode, biased beyond V_{br} by an overvoltage, V_{ov}, but drawing no current until a free carrier in the multiplication region

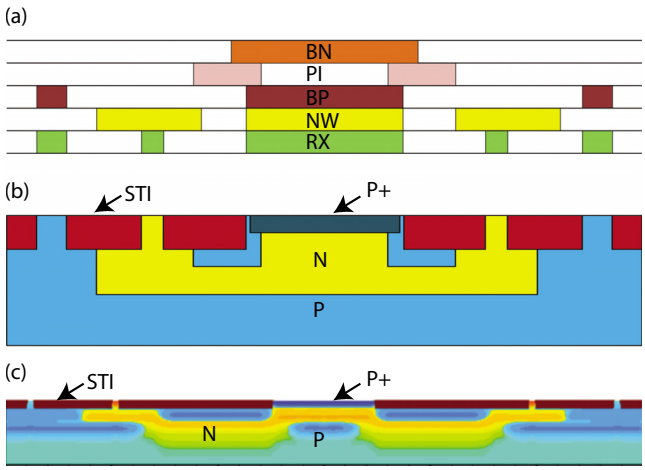

FIG. 1. (Color online) (a) Illustration of the designed mask layers used to make our SPAD. (b) Illustration of the expected pn diode after fabrication. (c) Process simulation results showing the expected diode structure that was fabricated. The multiplication region is the top P^+-N junction in the center of the device.

a)Electronic mail: shepard@ee.columbia.edu.

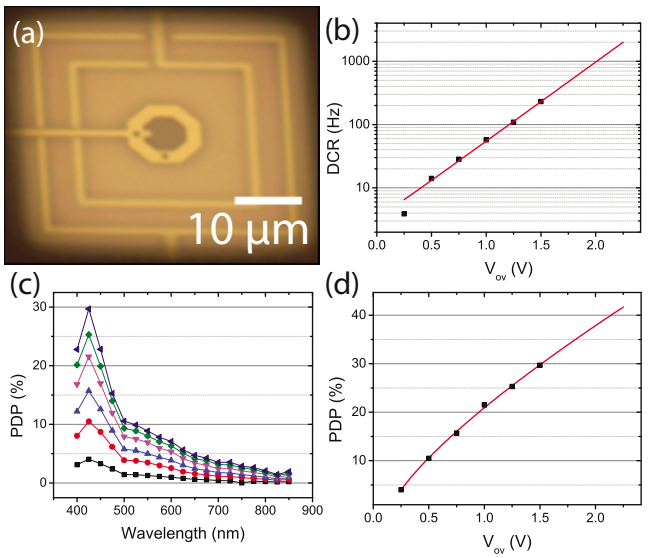

FIG. 2. (Color online) (a) Micrograph of fabricated SPAD structure. (b) A plot of the DCR as a function of V_{ov}. (c) The PDP for the SPAD over the range of visible wavelengths. The PDP measurements were made using a xenon arc lamp, CM110 monochromator, and a 2 in. integrating sphere. The events were recorded using the Agilent 53132A and the incident power was measured using a Thorlabs PM130D energy meter. (d) Plot of PDP as a function of V_{ov}.

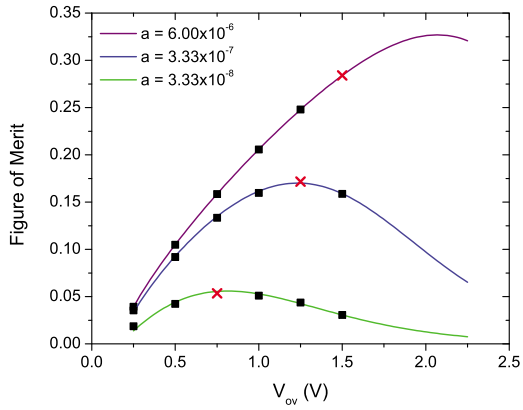

FIG. 3. (Color online) FOM for our device as a function of V_{ov} using the fitted functions in Fig. 2 for different fluorescence intensities. As μ decreases, the optimal bias point shifts to favor a low dark count. The data points are our measured values and the X's indicate the optimum bias point for our device.

triggers an avalanche. This mode of operation requires a quenching circuit, the simplest form of which is a resistor in series with the diode.[3] When an avalanche is triggered, a current flows through the resistor causing a voltage drop, which leads to the voltage across the diode rising above V_{br}, halting the current; the associated RC time constant to return to a reverse bias of $(V_{br}-V_{ov})$ defines the deadtime for the SPAD. In this work, a quenching resistance of 423 kΩ is employed yielding maximum avalanche current levels of 2.36 μA and a deadtime of 15 μs with V_{ov} of 1.0 V. With this quenching circuit, the probability of afterpulsing—a noise event caused by charges that do not clear the multiplication region before the SPAD is reset and retriggers the device—is negligible.

In Fig. 2(b), we present a plot of DCR as a function of V_{ov} at room temperature. For a V_{ov} of 0.25 V, the DCR is 6 Hz, increasing to only 231 Hz at 1.50 V. The DCR has the functional form $A_0 e^{V_{ov}/V_0}+B_0$, where $A_0=3.06$ s^{-1}, $B_0=0.217$ s^{-1}, and $V_0=0.347$ V, for our device at room temperature. In Fig. 2(c), we plot the photon detection probability (PDP) of this same device as a function of photon wavelength for different values of V_{ov}. The PDP peaks at 425 nm at just below 30% at a V_{ov} of 1.50 V. While this peak is at a shorter wavelength than SPADs reported in other technologies, this can be explained by the shallow junction depths ($\sim$300 nm) that result from using the p$^+$ mask for a PFET source and drain implant. In Fig. 2(d), we show the PDP as a function of V_{ov} which has the functional form $A_0|V_{ov}-V_c|^p$, with $p=0.759$, $A_0=23.8$ V$^{-0.759}$, and $V_c=0.156$ V.

To establish a metric by which SPAD devices can be compared and optimized at a given set of experimental conditions, we consider the case of a FLIM application in which a pulsed laser excites an ensemble of fluorophores with a monoexponential lifetime, τ. Following some simulation-based models,[13] we describe the fluorescence emission from a single fluorophore as a nonhomogeneous Poisson point process. We considered the PDP as constant in time and the DCR as a Poisson process with a rate constant given by the experimentally measured DCR. We assume that the afterpulsing probability is negligible and that the detector electronics are able to quench and reset the device in time for the next laser repetition and are not a limiting factor. We then compute three characteristic probabilities for the device: the probability of detecting an actual photon arrival, the probability of recording a miss when no photons are incident on the device, and the probability that a photon triggered an event given that an event has occurred. We define a figure of merit (FOM)[14] as the product of these three probabilities, which selects the device with the highest probability of properly recording photon events while avoiding noise events,

$$\begin{aligned}\mathrm{FOM} = {} & P(\text{detecting a hit} \mid \geq 1 \text{ photon arrives}) \\ & \times P(\text{detecting a miss} \mid 0 \text{ photons arrive}) \\ & \times P(\text{photon hit} \mid \text{event occurred}). \end{aligned} \tag{1}$$

These probabilities can be determined analytically, consistent with the assumptions above, to yield

$$\begin{aligned}\mathrm{FOM} = {} & \left\{1+\frac{\exp[-(\mathrm{DCR}\cdot T+\mu)]-\exp[-(\mathrm{DCR}\cdot T+\mu\cdot \mathrm{PDP})]}{1-\exp(-\mu)}\right\}\exp(-\mathrm{DCR}\cdot T) \\ & \times\left\{\frac{[1-\exp(-\mu)]-\exp(-\mu\cdot \mathrm{PDP})\times\{1-\exp[-\mu\cdot(1-\mathrm{PDP})]\}}{[1-\exp(-\mu)]-\exp(-\mu\cdot \mathrm{PDP})\times\{1-\exp[-\mu\cdot(1-\mathrm{PDP})]\}+[1-\exp(-\mathrm{DCR}\cdot T)]}\right\}. \end{aligned} \tag{2}$$

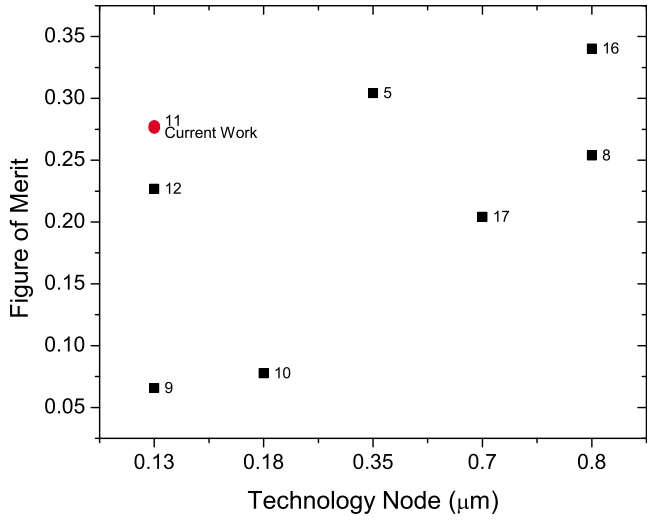

FIG. 4. (Color online) A comparison of the results of this work to other published results at both STI and LOCOS technology nodes based on our FOM. The appropriate reference is noted.

In this expression, T is the time window over which counts are recorded, limited by the period of laser repetition, and μ is the integrated photon count over T (as determined by the intensity of the fluorescence being detected). In the limit that $\mu \ll 1$ and $\mathrm{DCR}\cdot T \ll 1$, this becomes

$$\mathrm{FOM} = \mathrm{PDP}\cdot\left(\frac{\mu\cdot\mathrm{PDP}}{\mu\cdot\mathrm{PDP}+\mathrm{DCR}\cdot T}\right). \tag{3}$$

If we write the incident photon rate as being determined by the monoexponential fluorescence decay, then $\mu = A\cdot a\cdot\tau\cdot(1-e^{-T/\tau})$, where a is the maximum incident photon flux and A is the SPAD device area. For $T/\tau \gg 1$, this becomes $\mu = A\cdot a\cdot\tau$.

In Fig. 3, we plot the FOM as a function of V_{ov} for our device for $a = 3.33\times10^{-8}$, 3.33×10^{-7}, and 6.00×10^{-6} $\mu\mathrm{m}^{-2}\,\mathrm{s}^{-1}$ with $T = 20$ ns and $\tau = 3$ ns, which are representative of typical FLIM experiments. We have highlighted the optimum bias point, determined by the maximum FOM for each curve. To compare our device with other published results, we take the PDP and DCR for our device at the optimal V_{ov} for $a = 6.00\times10^{-6}$ $\mu\mathrm{m}^{-2}\,\mathrm{s}^{-1}$ in Fig. 3 of 1.5 V. Using the same representative T and τ values, this value of a is the largest that can be used while assuring that no device in the comparison has a $\mu\cdot$PDP product that results in a photon being detected in more than 1% of measurement windows, consistent with limiting pulse pile-up.[15] For other published results, we use the PDP and DCR given at the bias point chosen by the authors in reporting their results,[5,8–12,16,17] which may not be optimal for maximizing the FOM, but, nonetheless, forms a basis for comparison. In Fig. 4, we present the results of this analysis, which shows that devices in LOCOS technology nodes are capable of outperforming the STI based devices due to their lower DCRs, but that our device is one of the best performing designs in 0.13 μm technology.

In order to utilize our device in an application that requires precise timing, it must have a narrow impulse response. We measure the impulse response by exciting the SPAD with a 408 nm laser having a 45 ps full width at half maximum (FWHM) pulse and histogramming the resulting detector response with a Tektronix TDS7404 oscilloscope. The impulse response of the entire measurement system including the oscilloscope, laser, and SPAD is 198 ps FWHM.

In this work, we present a sensitive low-noise SPAD designed in a 0.13 μm CMOS process without modifications. This device creates the opportunity to pair a low-noise, highly sensitive SPAD with compact and fast timing circuits that will lead to more compact array-based imagers with short TCSPC image acquisition times. In addition, we have provided a framework with which the device's proper bias point can be established and SPADs from different fabrication processes can be compared.

This work was supported in part by the National Institutes of Health under Grant No. R33-HG003089. RMF was supported by the National Science Foundation Graduate Research Fellowship Program and National Defense Science and Engineering Graduate Fellowship.

[1] A. Kusumi, A. Tsuji, M. Murata, Y. Sako, A. C. Yoshizawa, S. Kagiwada, T. Hayakawa, and S. Ohnishi, Biochemistry **30**, 6517 (1991).
[2] L. Stryer, Annu. Rev. Biochem. **47**, 819 (1978).
[3] S. Tisa, F. Zappa, A. Tosi, and S. Cova, Sens. Actuators, A **140**, 113 (2007).
[4] E. Charbon, J. Phys. D **41**, 094010 (2008).
[5] D. Stoppa, D. Mosconi, L. Pancheri, and L. Gonzo, IEEE Sens. J. **9**, 1084 (2009).
[6] D. E. Schwartz, P. Gong, and K. L. Shepard, Biosens. Bioelectron. **24**, 383 (2008).
[7] M. Marwick and A. Andreou, CISS, 2007, pp. 741–744.
[8] A. Rochas, M. Gani, B. Furrer, P. A. Besse, R. S. Popovic, G. Ribordy, and N. Gisin, Rev. Sci. Instrum. **74**, 3263 (2003).
[9] C. Niclass, M. Gersbach, R. Henderson, L. Grant, and E. Charbon, IEEE J. Sel. Top. Quantum Electron. **13**, 863 (2007).
[10] H. Finkelstein, M. Hsu, S. Zlatanovic, and S. Esener, Rev. Sci. Instrum. **78**, 103103 (2007).
[11] J. A. Richardson, L. A. Grant, and R. K. Henderson, IEEE Photon. Technol. Lett. **21**, 1020 (2009).
[12] M. Gersbach, J. Richardson, E. Mazaleyrat, S. Hardillier, C. Niclass, R. Henderson, L. Grant, and E. Charbon, J. Solid State Electrochem. **53**, 803 (2009).
[13] M. Repich, D. Stoppa, L. Pancheri, and G. Dalla Betta, Proc. SPIE **7355**, 73550O (2009).
[14] See supplementary material at http://dx.doi.org/10.1063/1.3518473 for a detailed presentation of the FOM and its justification.
[15] C. M. Harris and B. K. Selinger, Aust. J. Chem. **32**(10), 2111 (1979).
[16] S. Tisa, F. Zappa, and I. Labanca, Tech. Dig. - Int. Electron Devices Meet. **2005**, 815.
[17] L. Pancheri and D. Stoppa, IEEE ESSDERC, 2007, pp. 362–365.