HHS Public Access

Author manuscript

Proc IEEE Int Conf Micro Electro Mech Syst. Author manuscript; available in PMC 2026 March 25.

Published in final edited form as:

Proc IEEE Int Conf Micro Electro Mech Syst. 2026 January ; 2026: 1563–1566. doi:10.1109/mems64181.2026.11419596.

CMOS-EMBEDDED MICROFLUIDICS FOR CHANNEL-ADDRESSABLE PARALLEL READOUT OF SPAD FLUORESCENCE LIFETIME SENSORS

Max S. Ladabaum,
Alexander Di,
Julian M. Bao,
Grigory Tikhomirov,
Jun-Chau Chien
University of California, Berkeley

Abstract

We present the integration of CMOS-embedded subtractive microfluidics with single-photon avalanche diode (SPAD) fluorescence lifetime sensors, enabling independent channel addressability for parallel readout. Microfluidic channels are realized by selectively wet-etching the back-end-of-line (BEOL) metal routing above the SPAD sensors, reducing the analyte-to-active-region spacing to 7 μm while allowing for precision alignment. Simultaneous readout from two SPADs underneath two separate fluidic channels is demonstrated, establishing a pathway toward a scalable, multiplexed lab-on-CMOS biosensing platform.

Keywords

CMOS; SPAD; Fluorescence lifetime; TCSPC; Subtractive microfluidics; Lab-on-Chip; Biosensor

INTRODUCTION

SPADs provide highly sensitive, photon-efficient optical readout and are well suited for fluorescence lifetime detection in compact sensing systems. Fluorescence lifetime measurements offer several advantages over intensity-based methods because lifetime is largely insensitive to excitation power, fluorophore concentration, and variations in optical path. These properties are particularly beneficial for analyte detection in scattering or autofluorescent biological media such as blood, saliva, or tissue.

Compared to traditional fluorescence lifetime detection methods, SPADs offer several key advantages. They are compatible with standard CMOS processes and support monolithic integration with on-chip electronics. They provide picosecond timing resolution, operate at relatively low voltages, and occupy a compact footprint. As a result, SPADs can scale to

CONTACT: M.S. Ladabaum, maxladabaum@berkeley.edu, J-C. Chien, jcchien@berkeley.edu.

high density arrays that are attractive for lab-on-chip portable sensing platforms [1] as well as wearables and implantables [2].

Despite these advantages, most reported SPAD-based biosensing systems use external or overlay microfluidic assemblies that span hundreds of micrometers above the SPAD array [1]. This geometry limits photon collection efficiency and prevents fine spatial control of analytes above individual sensor pixels. As a result, opportunities for on-chip multiplexing and independent channel addressability are constrained. Recent efforts have explored forming microfluidic channels directly within CMOS dies [3][4]. However, these demonstrations have not shown functional fluidic channels combined with on-chip sensing.

In this work, we address these limitations by integrating subtractive microfluidic channels directly above SPAD sensors fabricated in a standard TSMC 180-nm CMOS process. The subtractive microfluidics concept allows fluidic channels to be defined lithographically in the BEOL metal stack [5]. Channels are formed by removing the BEOL metal through a single wet etch step. To our knowledge, this is the first functional demonstration of CMOS embedded subtractive microfluidics combined with SPAD-based fluorescence lifetime sensing and the first demonstration of independently addressable fluidic channels for parallel fluorescence lifetime measurements on a CMOS platform.

METHODS

We designed and tested three P+/N-well-based SPADs at different diameters: 4, 9, and 24 μm (Figure 1). All three designs use a circular geometry to avoid high electric fields associated with sharp corners [6].

The SPADs are operated continuously in Geiger mode with each device reverse-biased above its breakdown voltage (14.9–15.2 V) using an external power supply (Keysight E36312A). Under these conditions, the absorption of a single photon can trigger avalanche breakdown. The resulting charge carriers are accelerated by the high electric field and initiate a cascade of impact ionization events that rapidly increases the number of carriers. This process produces an exponentially rising current pulse, which must be quenched to prevent sustained breakdown.

Each SPAD is wire-bonded to two bond pads on a printed circuit board (PCB). The SPAD's cathode is biased at a positive voltage, V+, above breakdown, and the anode is returned to V− through a quenching resistor (Figure 2). The voltage at the SPAD anode is amplified and then measured using the oscilloscope module from Liquid Instruments Moku:Pro.

Figure 3 shows the measurement setup. The SPADs are illuminated with a 405 nm laser diode (ThorLabs DL5146–101S), positioned a few centimeters above the chip. The laser is driven by a gain-switched laser pulser (ICHaus iC-HS05 iCSY HS3M), producing pulses with pulse-widths between 1 to 4 ns. A periodic TTL signal (10–20 MHz), generated by the arbitrary waveform generator module of the Moku:Pro, is used to trigger the laser pulses.

To measure fluorescence lifetimes, we perform time-correlated single-photon counting (TCSPC), which determines the distribution of photon arrival times following pulsed

excitation. Traditional forward TCSPC starts a timer when the laser pulse is triggered and stops the timer when a photon is detected. In reverse TCSPC, the timer starts at the photon detection event and stops at the next laser pulse. We implement TCSPC using the Time and Frequency Counter module on the Moku:Pro. Because the instrument cannot reset its timer each cycle, we use reverse TCSPC to avoid ambiguity when a photon is detected after multiple excitation pulses. Without reverse timing, these detections would produce periodic peaks in the histogram. After acquiring the reverse histogram, we reconstruct the corresponding forward TCSPC histogram by inverting the time axis. Because the excitation pulse has a finite temporal width, the measured TCSPC histograms reflect the convolution of the true fluorescence decay with the instrument response function (IRF). Figure 4 illustrates an example of a photon detection event and a TCSPC histogram.

In our design, the photon collection efficiency is improved by aligning the fluorescent analyte directly above the SPADs using integrated fluidic channels. This not only increases the detection sensitivity but also enables tight integration that allows for scaling-up the number of fluidic channels. The fluidic channels are formed using the Metal-6 layer from the BEOL metal routing of a TSMC 180-nm CMOS chip as a sacrificial layer. The removal of the sacrificial metal is achieved by submerging the CMOS chip in ammonium persulfate at 90°C for 24 hours.

The etched fluidic channels enable the analytes to flow within 7 μm of the SPAD surface, which enhances the photon collection efficiency (Table 1). Assuming the fluorescent analyte is a point source that emits photons equally in all directions, located in the middle of the fluidic channel (Figure 5):

$$Photon\ Collection\ Efficiency = \eta = \Omega/4\pi \tag{1}$$

$$Solid\ Angle = \Omega = 2\pi(1 - \cos\theta) \tag{2}$$

$$\eta = \frac{1}{2}(1 - z/\sqrt{z^2 + r^2}) \tag{3}$$

In addition to improving the collection efficiency, the etched channels allow each SPAD array to interface with a different fluid, enabling individual addressability and supporting on-chip parallelization, as shown in Figure 6.

Large PDMS microfluidic channels are utilized to interface with the CMOS-embedded fluidic channels. The PDMS is prepared by molding against a 3D printed template ordered from ProtoLabs (MicroFine Green material). After molding, the PDMS layer is manually aligned to the etched microfluidic inlets and outlets on the CMOS chip, where it is sealed against the unetched areas of the surface. Fluid is introduced into the system

using a manually operated syringe, which fills both the PDMS channels and the etched nanochannels. Food dye is added to the injected solution to visualize the wetting process and to verify proper channel filling.

RESULTS

We characterize the dark count rate (DCR) as a function of overvoltage for both etched and unetched SPADs (Figure 7). The unetched SPADs show inherent variability in DCR and breakdown voltage across devices of the same geometry. The etched SPADs exhibit similar variability, with no consistent differences relative to the unetched group. This suggests that the wet etching process does not degrade device performance. The improvement in stability is an advantage over previously reported subtractive microfluidics approaches based on reactive ion etching (RIE), which have shown increased DCR and other effects attributed to plasma-induced damage [3].

To evaluate the performance of the SPAD devices, we first measure the laser pulse-width by analyzing a TCSPC histogram (Figure 8). Using a 20 MHz laser pulse rate, we collect 50,000–100,000 photon counts per second. We perform measurements at different pulse-width settings, varying by a few hundred picoseconds.

In post-processing, we vary the histogram window size to evaluate how integration time impacts the TCSPC data. As the integration window decreases (from 1 s to 50 ms integration time), the extracted pulse-width remains accurate, but the precision degrades due to increased shot noise. This behavior highlights an important advantage of TCSPC over intensity-based fluorescence measurements. Because TCSPC measures a time constant, it is not sensitive to the total number of detected photons. In contrast, intensity-based methods are sensitive to changes in excitation power, detector gain, optical alignment, and sample concentrations, mandating a careful calibration procedure.

To demonstrate independent fluidic addressability, we inject a solution of 50 nM quantum dots (Qdot 605 from Thermo Fisher, lifetime = 31 ns) into a microfluidic channel aligned to a SPAD. The fluorescence lifetimes measured by the sensing device and its neighboring reference SPAD, which is exposed to air, are measured using an identical measurement and signal-processing workflow.

As shown in Figure 9, the QD-addressed SPAD measures a clear lifetime increase with the presence of QDs, while the air-addressed SPAD shows no signal change. The contrast between the two channels shows that each SPAD responds independently to the corresponding fluidic channel with minimal crosstalk. As a control, we replace the QDs with water and observe no detectable lifetime shift.

While we observe a clear increase in fluorescence lifetime when the channel is filled with QDs, the extracted lifetime does not match the expected value from the datasheet. This discrepancy is likely due to the shift in the IRF between different SPADs and packaging variations.

CONCLUSION

This work demonstrates the integration of SPADs with subtractive microfluidics for fluorescence lifetime detection in a standard CMOS process. Benefiting from the direct integration, the device enables channel-addressable sensing in a compact chip-scale form factor. Measurements with quantum dots verify that the platform can detect fluorophores at tens of nanomolar concentrations, and experiments with two adjacent fluidic channels demonstrate independent fluidic addressability without noticeable crosstalk. Together, these results validate subtractive microfluidics as a practical route toward scalable, multiplexed fluorescence lifetime sensing in CMOS. Future work will expand the number of channels, integrate on-chip processing circuitry, and explore applications in point-of-care sensing.

ACKNOWLEDGEMENTS

The authors thank TSMC University Shuttle Program, NIH Trailblazer Award R21EB036154, NIH R01EB038309, NSF CAREER Award 2442713, and DARPA MICA Award HR001125CE031.

REFERENCES

[1]. Canals J et al. , A Point-of-Care Device for Molecular Diagnosis Based on CMOS SPAD Detectors with Integrated Microfluidics. Sensors. 2019; 19(3):445. [PubMed: 30678225]

[2]. Choi J et al. , “Fully Integrated Time-Gated 3D Fluorescence Imager for Deep Neural Imaging,” in IEEE Transactions on Biomedical Circuits and Systems, vol. 14, no. 4, pp. 636–645, Aug. 2020. [PubMed: 32746353]

[3]. Zikrallah A et al. , “Integration of a SPAD on a CMOS nanofluidic platform for fluorescence-labeled single biomolecules detection and imaging”, Proc. SPIE 13316, Optical Diagnostics and Sensing XXV: Toward Point-of-Care Diagnostics, 133160I (20 March 2025).

[4]. Meng H, “CMOS Nanofluidics.” MIT Thesis, 2018.

[5]. Weng W-Y et al., “Subtractive Microfluidics in CMOS,” 2024 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2024, pp. 1–4.

[6]. Bose S et al. , “Parametric Study of p-n Junctions and Structures for CMOS-Integrated Single-Photon Avalanche Diodes,” in IEEE Sensors Journal, vol. 18, no. 13, pp. 5291–5299, July 1, 2018.

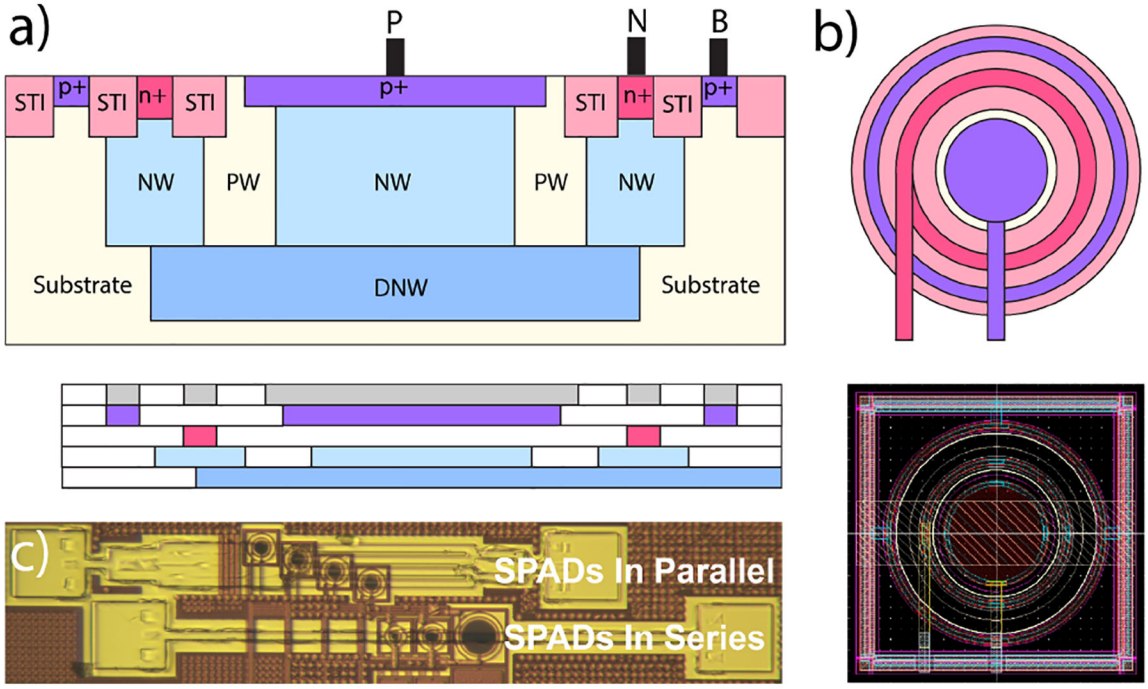

Figure 1:

a) SPAD cross-section illustration. b) SPAD top view (top) and its layout (bottom). c) A microscope image of SPADs with co-integrated subtractive microfluidic channels. The top structure has a 9-μm SPAD beneath each fluidic channel. The bottom structure has a series of three different sized SPADs situated along a single fluidic channel.

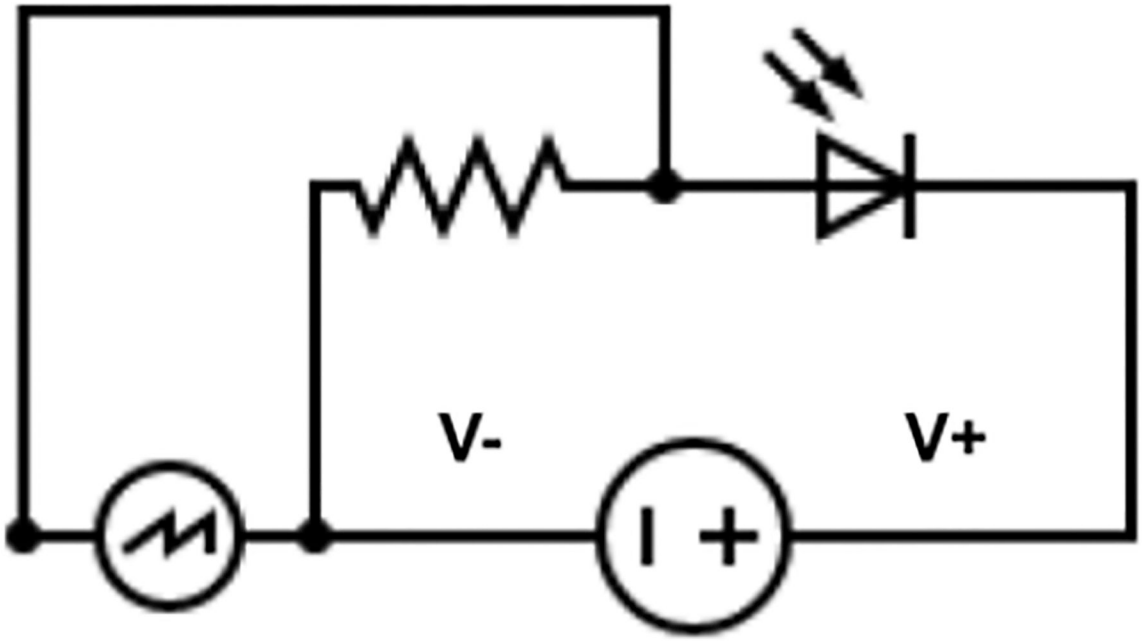

Figure 2:
Schematic of the passive quenching scheme.

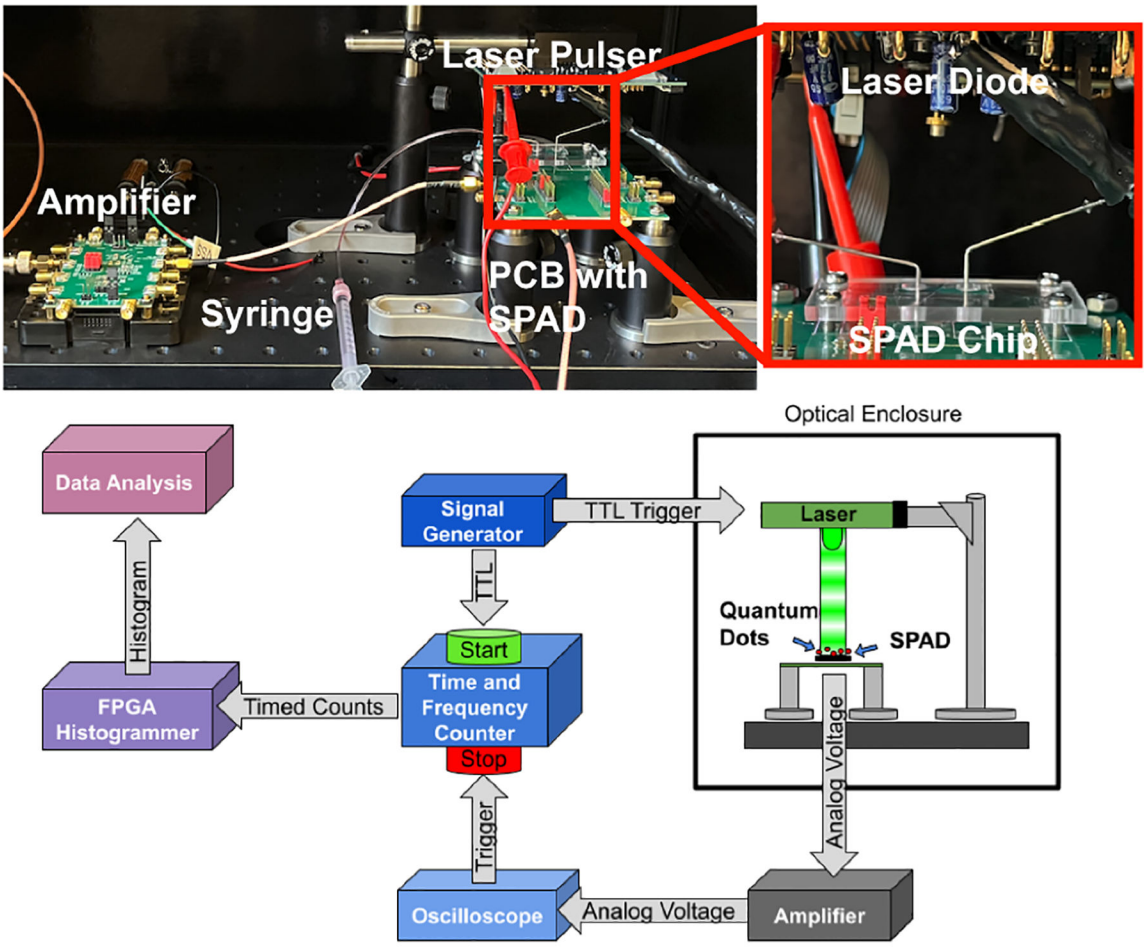

Figure 3:

Measurement setup (top) and the corresponding system diagram (bottom). The testing setup sits within an optical enclosure. A syringe is used to inject quantum dots into the etched fluidic channels.

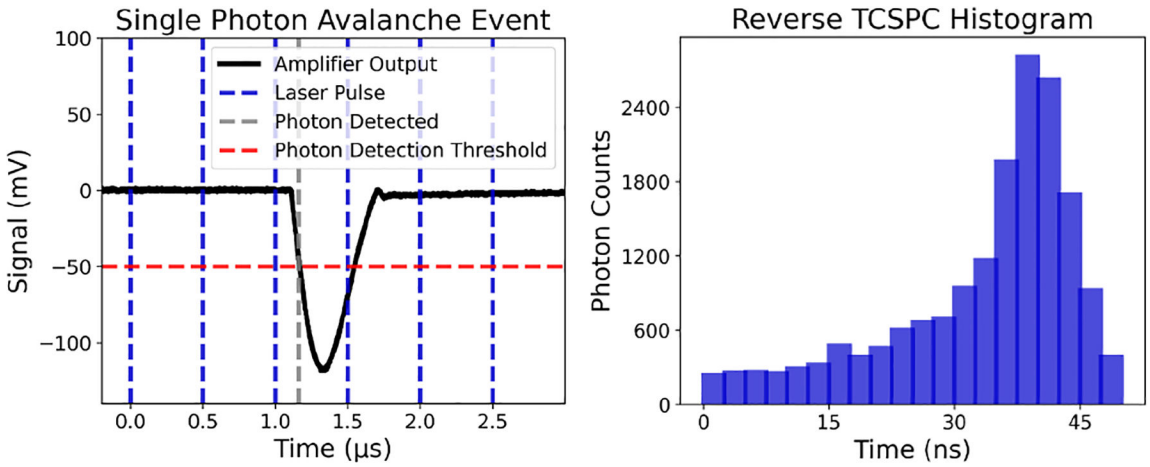

Figure 4:

A single photon avalanche event with overlaid laser pulses (left). A reverse TCSPC histogram generated by repeatedly measuring the time between photon detection events and the next laser pulse (right).

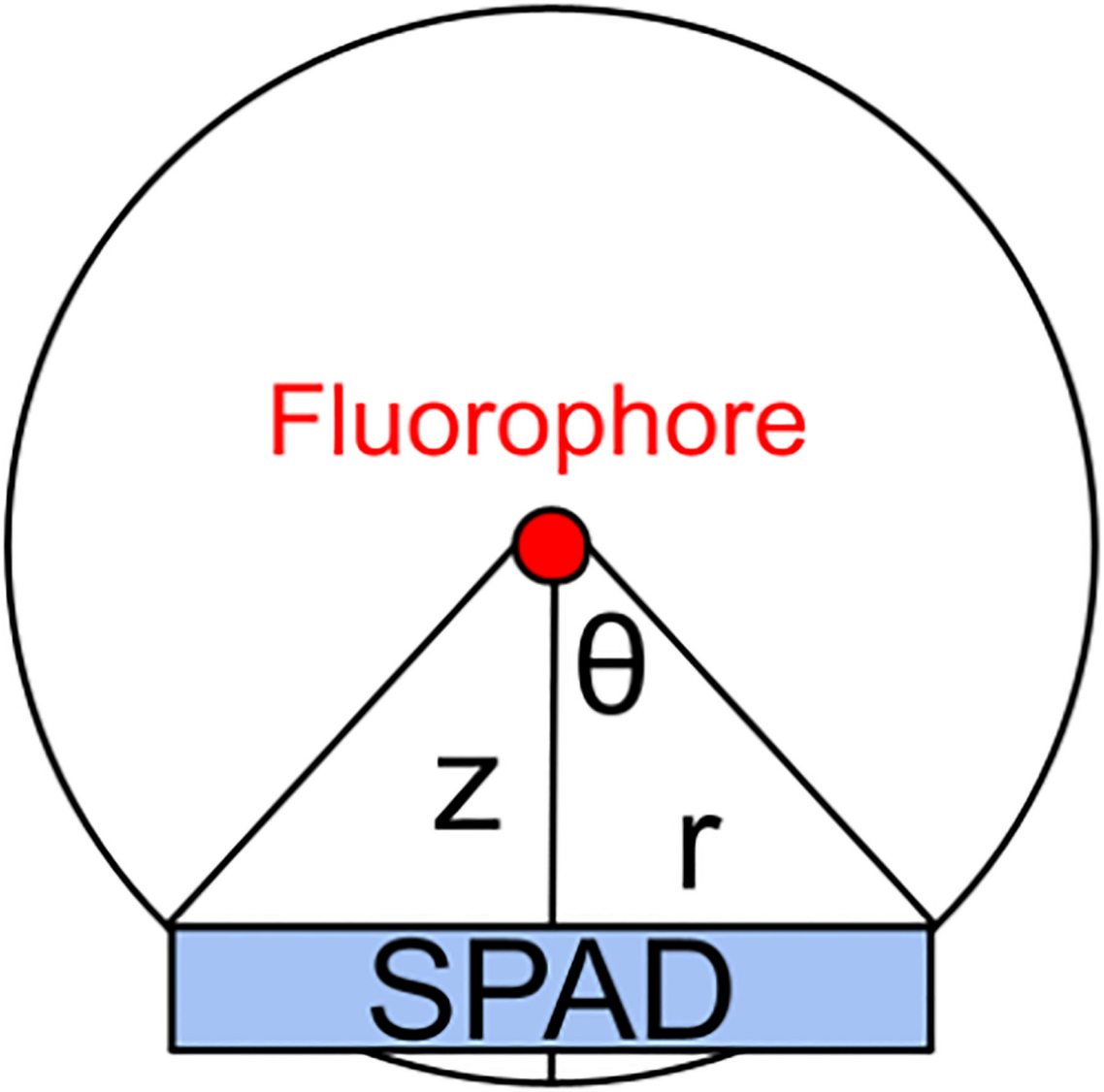

Figure 5:
Schematic of a fluorophore as a point source.

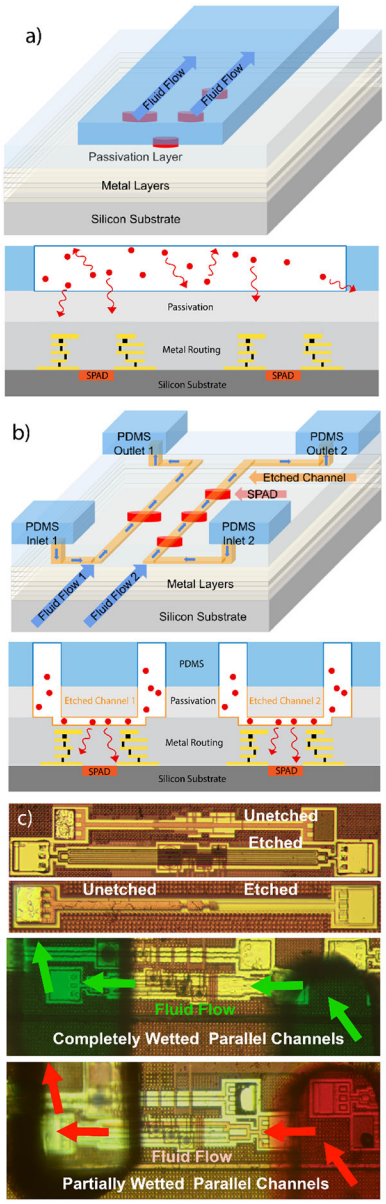

Figure 6:

a) Illustration showing a conventional fluidic assembly with a single, large PDMS fluidic channel placed on top of the chip. b) Illustration showing our fluidic assembly with PDMS inlets and outlets feeding into etched fluidic channels. c) Microscope images showing etched versus unetched channels (top) and fluid flow through etched channels (middle, bottom).

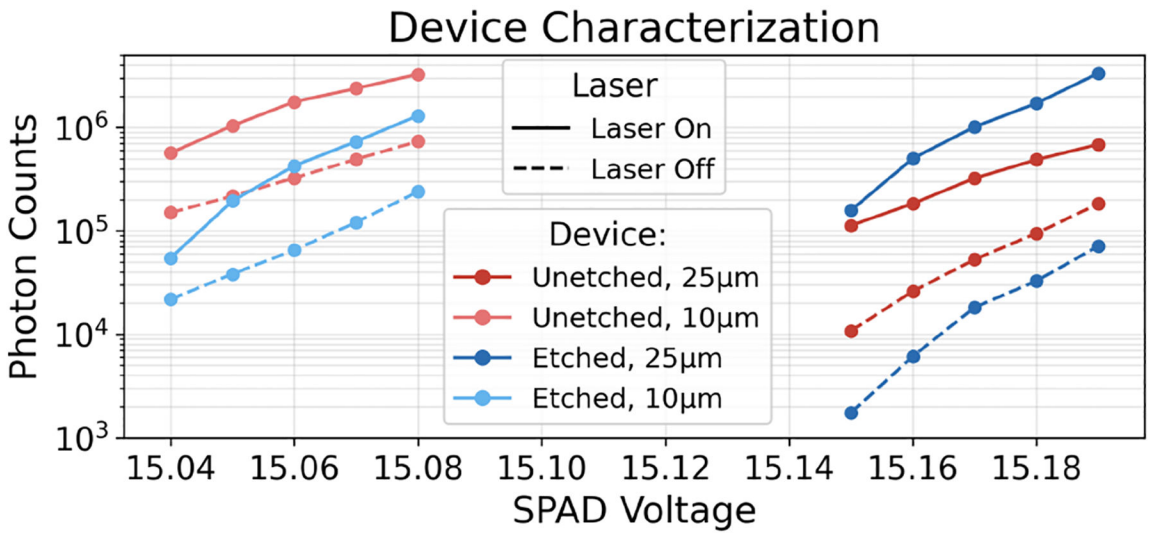

Figure 7:
Device characterization showing dark counts as a function of SPAD overvoltage.

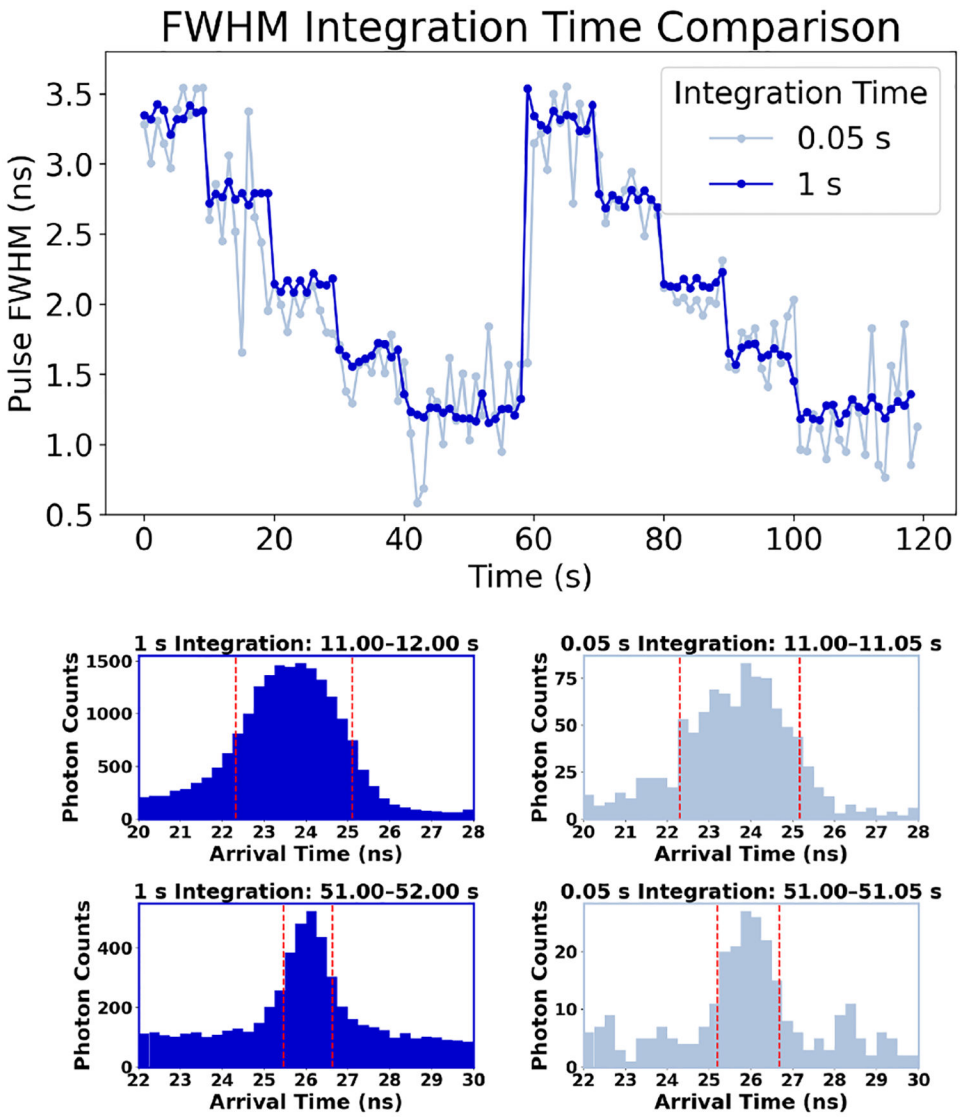

Figure 8:

The full width at half maximum (FWHM) of a periodic (20 MHz) laser pulse for various pulse-widths (top). Each FWHM value is extracted from a TCSPC histogram generated by collecting many photons over a specified integration time (bottom).

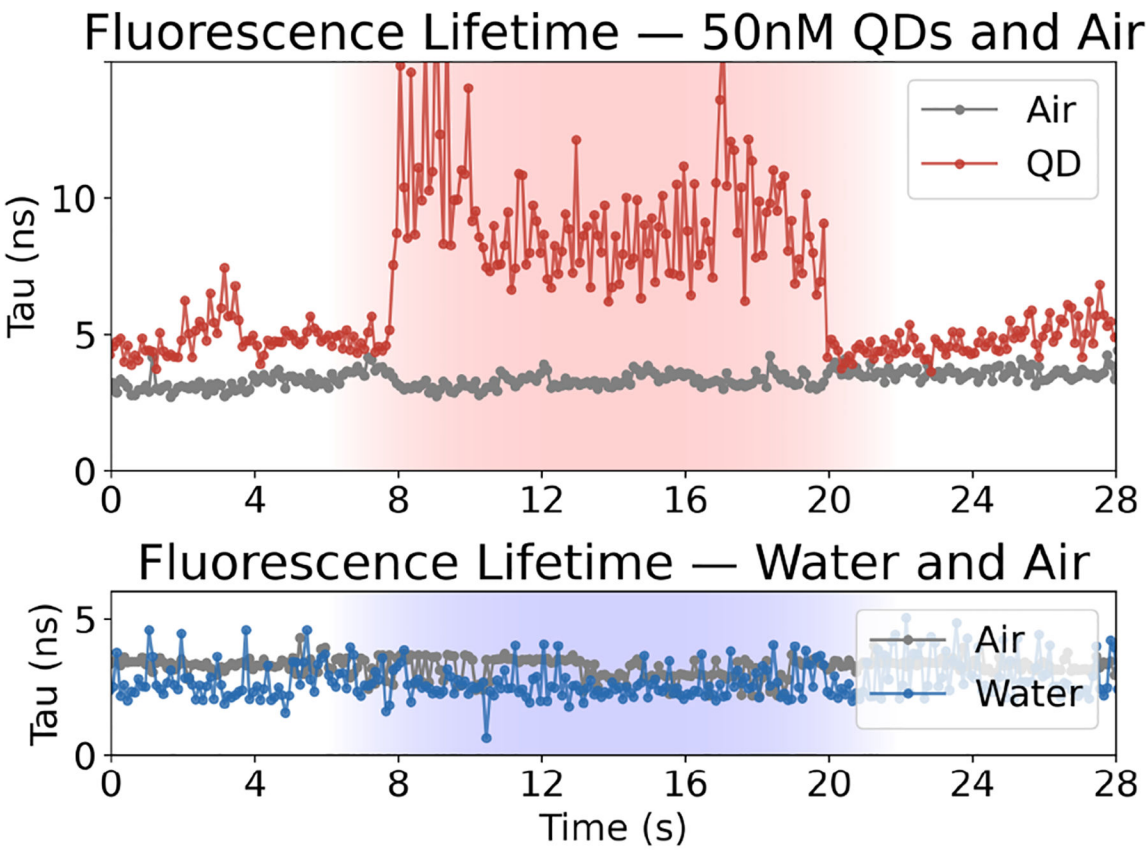

Figure 9:
Simultaneous fluorescence lifetime measurements from adjacent SPADs, with QDs above one SPAD and air above the other (top). A control experiment with water in place of QDs (bottom).

Table 1:

Photon collection efficiency as a function of distance between SPAD and fluorescent analyte.

Analyte-to-SPAD Distance	SPAD Radius	Collection Enhancement
200 μm (coverslip)	2 μm	1x
10 μm (PDMS on chip)	2 μm	335x
7 μm (this work)	2 μm	770x