

# A 512-Pixel, 51-kHz-Frame-Rate, Dual-Shank, Lens-Less, Filter-Less Single-Photon Avalanche Diode CMOS Neural Imaging Probe

Jaebin Choi<sup>1</sup>, Student Member, IEEE, Adriaan J. Taal, Student Member, IEEE, Eric H. Pollmann, Student Member, IEEE, Changhyuk Lee<sup>2</sup>, Member, IEEE, Kukjoo Kim, Laurent C. Moreaux, Michael L. Roukes, and Kenneth L. Shepard<sup>3</sup>, Fellow, IEEE

**Abstract**—We present an implantable single-photon shank-based imager, monolithically integrated onto a single CMOS IC. The imager comprises of 512 single-photon avalanche diodes distributed along two shanks, with a 6-bit depth in-pixel memory and an on-chip digital-to-time converter. To scale down the system to a minimally invasive form factor, we substitute optical filtering and focusing elements with a time-gated, angle-sensitive detection system. The imager computationally reconstructs the position of fluorescent sources within a 3-D volume of  $3.4 \text{ mm} \times 600 \mu\text{m} \times 400 \mu\text{m}$ .

**Index Terms**—CMOS, compressed sensing, neural probe, single-photon avalanche diode (SPAD), time-gated fluorescence imaging.

## I. INTRODUCTION

**O**PTICAL functional neural imaging has revolutionized neuroscience with optical reporters that enable single-cell-resolved monitoring of neuronal activity *in vivo*. State-of-the-art microscopy methods, however, are fundamentally limited in imaging depth by optical scattering in tissue even with the use of the most advanced multiphoton microscopy techniques [see Fig. 1(a)] [1], [2].

Instead, if a camera itself can be inserted into the brain, then imaging at arbitrary depth is possible. One method to enhance

Manuscript received May 14, 2019; revised August 1, 2019; accepted August 28, 2019. Date of current version October 23, 2019. This article was approved by Associate Editor Pedram Mohseni. This work was supported in part by the National Institutes of Health under Grant U01NS090596, in part by the Defense Advanced Research Projects Agency (DARPA) under Contract N66001-17-C-4012, and in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Contract W911NF-12-1-0594. (Jaebin Choi and Adriaan J. Taal are co-first authors.) (Corresponding author: Kenneth L. Shepard.)

J. Choi, A. J. Taal, and E. H. Pollmann are with the Electrical Engineering Department, Columbia University, New York, NY 10027 USA (e-mail: jc3797@columbia.edu).

C. Lee is with the Brain Science Institute, Korea Institute of Science and Technology, Seoul 02792, South Korea.

K. Kim is with the Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea.

L. C. Moreaux and M. L. Roukes are with the California Institute of Technology, Pasadena, CA 91125 USA.

K. L. Shepard is with Bioelectronic Systems Laboratories, Columbia University, New York, NY 10027 USA.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSSC.2019.2941529

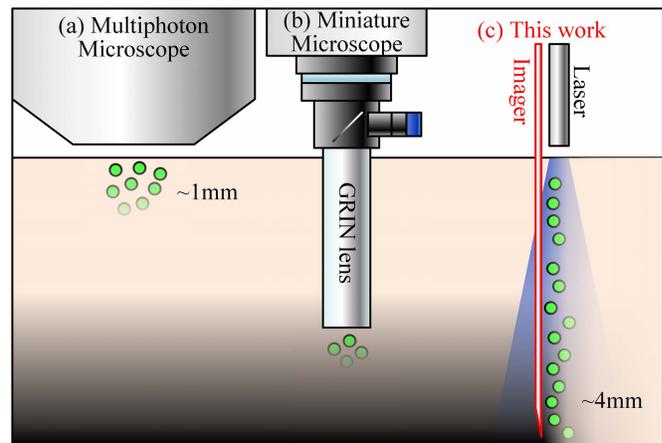


Fig. 1. Size and imaging depth comparison of (a) multi-photon microscope, (b) implantable microscope with inserted GRIN lens, and (c) imaging probe resolving 4.1 mm deep in brain tissue with an external excitation laser. Fluorescent emission is collected by 512 SPAD pixels located on two shanks.

imaging depth from a multiphoton microscope is to implant a relay graded-index (GRIN) lens [3]. While these lenses allow one to couple laser power into deep nuclei at the expense of the cerebral tissue above the interrogation volume and enable functional calcium imaging, their displaced tissue volume is very large ( $> 0.1 \text{ mm}^3$ ) compared with the imaging volume ( $> 0.01 \text{ mm}^3$ ). These methods achieve deep brain imaging at the cost of significant neural network trauma and still require free-space microscopes for readout.

In addition to solving the imaging depth problem, there is an interest in scaling down the size of the epifluorescence microscopes themselves [see Fig. 1(b)] [4], [5]. These efforts have primarily focused on small-form-factor implementations of full microscopes, including lenses and optical filters. This continued reliance on lenses and filters challenges miniaturization.

More recently, there have been significant efforts directed toward lens-less imaging. These approaches have been based on far-field masking (either phase or amplitude) to produce a spatially diverse illumination pattern on the imager that can be used to computationally reconstruct an image [6].

The requirement for far-field positioning of the mask, however, precludes the use of these approaches in shanks, where shank thicknesses want to be maintained at a minimum to prevent tissue displacement. The requirement for optical spectral filters in the scheme of fluorescence microscopy adds to the probe thickness.

In this article, we take an entirely different approach in which the “camera” itself can be inserted into the brain, allowing imaging at arbitrary depth by collecting fluorescence signal very close to the neuron of interest, without its signal having to scatter and attenuate along its insertion depth. In particular, we present a monolithically integrated single-photon imager in the form of a CMOS, shank-based optical image sensor array that can be inserted into the brain [see Fig. 1(c)] [7]. Single-photon avalanche diodes (SPADs) fabricated in CMOS are used as proximal light detectors. Lenses are replaced with near-field diffraction gratings placed on each SPAD, giving each of them a unique angularly modulated field of view (FoV) [8]. The mapping of SPAD counts into a volumetric scene allows the imager to determine the location of light sources. This additional information provides a multi-source localization capability similar to a far-field lens-less imaging approach. An external 480-nm pulsed picosecond laser delivers fluorescence excitation at the insertion point of the shanks by means of a fiber-coupled collimator. Spectral filters are replaced with time-gated operation of the SPADs in which detectors are turned on immediately after the pulsed excitation light has been shut off.

This article is structured as follows. In Section II, we present the system design and explain the post-processing into the shank form factor. The SPAD device and quenching circuit that allows for time-gated, filter-less fluorescence photometry are described in Section III. Section IV describes the implementation of lens-less imaging through angle-sensitive detection, and Section V demonstrates the computational multi-source localization of phantom targets.

## II. SYSTEM OVERVIEW

Fig. 2(a) shows the die photograph of the implantable imaging probe with two shanks, each 4.1-mm long and 120- $\mu\text{m}$  wide, connected to a 2 mm  $\times$  1.2 mm head. Geometrically, this differs from our earlier design [7], extending the shanks by approximately 650  $\mu\text{m}$  in length and reducing the shank separation from 250 to 175  $\mu\text{m}$  to increase the overlap of individual shank FoVs, aiding reconstruction performance. As a result, it is used to accommodate buffers and decoupling capacitors for delivering signal and power along the shank length. Each shank employs two rows of 128 SPADs operating in the Geiger mode at 25.3- $\mu\text{m}$  pitch and 7.7- $\mu\text{m}$  active diameter, delivering a 6.3% fill factor. Each pixel contains a SPAD detector with a quenching circuit, 6-bit memory, and addressing logic. Because the shanks have partially overlapping FoVs, a total imaging volume of 3.4 mm  $\times$  600  $\mu\text{m}$   $\times$  400  $\mu\text{m}$  is achieved. The probe head contains a digital-to-time converter (DTC) for on-chip control of pixel gating and metal–oxide–metal (MOM) capacitors for decoupling the SPAD bias voltage. In the subsequent discussion, we use the x-direction to denote the distance along the length of the shank, the z-direction to denote

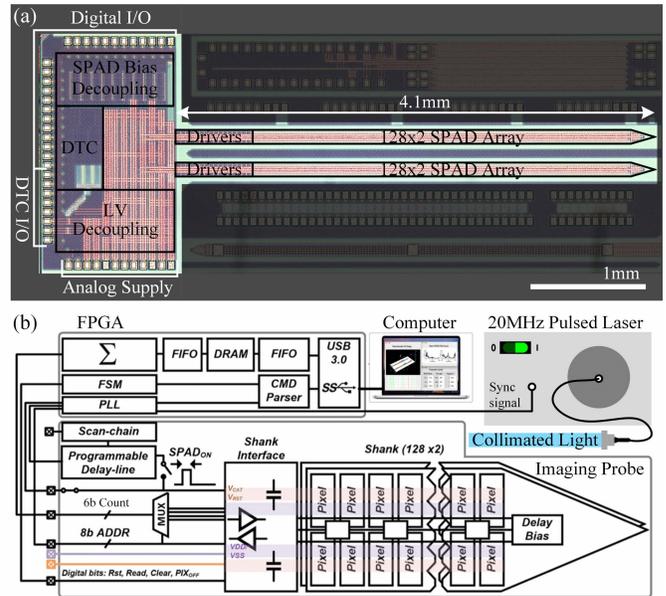


Fig. 2. (a) Chip micrograph with shank IC highlighted. (b) System architecture consisting of optoelectronic IC, readout FPGA, and computer for image reconstruction.

the vertical distance away from the shanks, and the y-direction to denote the distance along the width of the shank.

As fabricated in a 130-nm high-voltage process, the system is subdivided into four voltage domains, including a 1.5-V digital core domain, a 3.3-V digital input/output (IO) domain, a 5-V domain for the SPAD quenching circuits, and a >16-V domain for biasing the SPAD cathodes.

### A. System Architecture

A system-level block diagram is shown in Fig. 2(b). A field-programmable gate array (FPGA) is used to synchronize time-gated photon counting with an external pulsed laser. A phase-locked loop (PLL) implemented within the FPGA uses a photodetector-produced laser pulse waveform to deliver a synchronization signal to the DTC, which then produces a global time-gated signal with programmable delay and pulse width. Photon counts are accumulated with this global shutter for a desired number of laser pulses and read out by addressing each pixel’s 6-bit memory serially with a 31.25-MHz readout clock.

This readout clock frequency is ultimately limited by the rate at which data can be read from the pixel-level memory. A per-pixel sequential readout rate of 31.25 MHz, combined with a minimum image acquisition time of 3.2  $\mu\text{s}$  when used in synchronization with a 20-MHz laser repetition rate, leads to a maximum possible frame rate of 51 kilo-frames per second (kfps). The 6-bit dynamic range at this frame rate can be maximally utilized at high signal conditions, which corresponds to a 520-nm wavelength irradiance of  $2.2 \times 10^6$  photons/s/ $\mu\text{m}^2$  at the pixel surface. The data are streamed out with a USB 3.0 link, and image reconstruction is performed on an external computer.

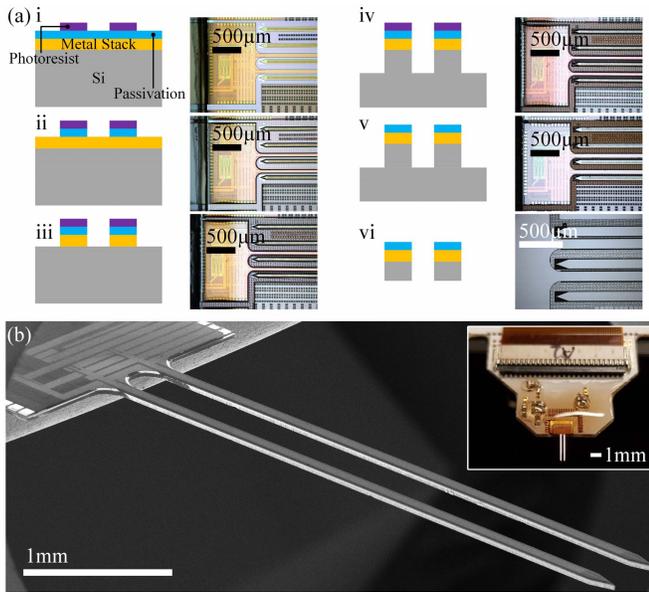


Fig. 3. (a) Post-processing of CMOS die into implantable imaging probe. (b) Scanning electron microscopy image of post-processed imaging probe with 120- $\mu\text{m}$  width and 40- $\mu\text{m}$  thickness, packaged for insertion (inset).

### B. Post-Processing

In order to further miniaturize the imaging probe's volume displacement, excess silicon surrounding the shank is removed to yield a 40- $\mu\text{m}$ -thin structure ready for implantation. Fig. 3(a) shows the fabrication sequence starting with photoresist (AZ P4620, MicroChemicals) patterning, followed by an  $\text{O}_2\text{-CF}_4$  reactive ion etch through the passivation and CMOS stack. A Bosch process etches through 100  $\mu\text{m}$  of silicon substrate, after which photoresist is stripped and the die is mechanically milled (X-PREP, Allied High Tech Products) to remove 250  $\mu\text{m}$  of substrate silicon, fully releasing the shanks [see Fig. 3(b)]. The imaging probe is finally wire-bonded and packaged to a printed circuit board with a flat flexible cable connection and passivated with an epoxy [see the inset in Fig. 3(b)].

## III. TIME-GATED FLUORESCENCE IMAGING

Optical multi-dielectric spectral filters are an important part of any fluorescence microscope. They are required to reject the excitation intensity and allow only the fluorescence response to reach the detector. It is typically accomplished in the spectral domain by three components, an emission filter, an excitation filter, and a dichroic mirror, combined usually in a "cube" to provide an optical density (OD)  $> 6$  rejection of the excitation wavelength. Achieving the same levels of rejection with an integrated filter alone on our shank-based imagers remains difficult and is compounded by the dependence of filter properties on incident angle in the case of thin-film interference filters. Pigment-based absorption filters, as commonly used in color displays and CMOS imagers, do not have adequate rejection ratios [9]. Interference filters are difficult to integrate on CMOS and have problems in the rejection of non-orthogonal incident light [10], [11].

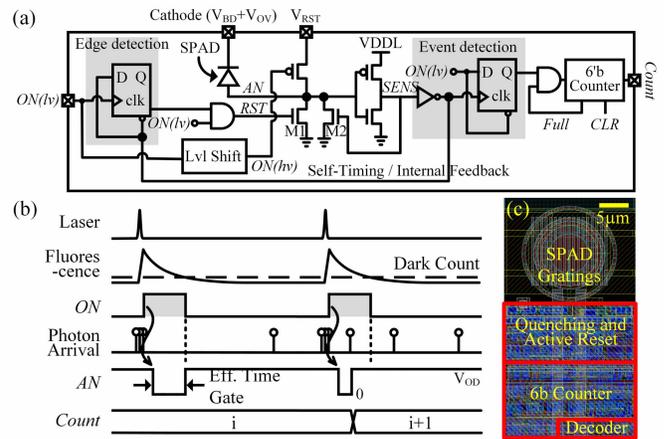


Fig. 4. (a) Circuit diagram of active-reset quenching circuit. (b) Timing diagram of laser-synced time-gated operation. (c) Layout of single pixel.

As an alternative to spectral filters, time-domain rejection is a powerful alternative by time-stamping the excitation source (pulse) and time-gating the detectors such that the fluorescence response is detected after the excitation has been turned off. This relies on the fact that fluorescent reporters have a time decay of fluorescence intensity characterized by their lifetimes. For commonly used fluorescent dyes, these lifetimes are in the 1–10-ns range (4.1 ns for enhanced green fluorescent protein (EGFP), 4.0 ns for fluorescein, and 1.68 ns for Rhodamine B). The signal-to-background ratio (SBR) of a time-gated fluorescence image is determined by the ratio of the integrated photon count coming from the fluorescence emission to that remnant from the excitation source. The exact position of the time-gate relative to the laser turn-off time is chosen to maximize this SBR for the given laser turn-off and lifetime characteristics of the fluorescence.

### A. Pixel Quenching and Reset Circuit

The in-pixel quench-and-reset circuit and pixel layout are shown in Fig. 4(a) and (c). Each of the 512 SPADs is reverse-biased between a global cathode and individual anode (AN). The cathode is held at a constant voltage above breakdown ( $V_{BD}$ ), while the anode (AN) moves between zero and the reset voltage ( $V_{RST}$ ) to put the SPAD into and out of the Geiger mode. The rising edge of the ON signal provided by the DTC triggers edge detection. The large reset NMOS transistor (M1) begins discharging AN until it drops below the input threshold ( $\sim 750$  mV) of the inverter-based comparator. When the inverter output goes high, the internal feedback clears the flip-flop-based edge detector, turning M1 off and setting AN to a high-impedance state. The SPAD is then biased in the Geiger mode and ready for avalanche breakdown at an incoming photon. The self-timed reset, a design change from the earlier version [7], optimizes the pulse width of the RST signal and internally minimizes the reset time for an arbitrary value of  $V_{RST}$  between 1.5 and 5 V without the need of an externally programmed pulse.

The long-channel half-latch transistor (M2) provides high impedance at the AN terminal of the SPAD. During avalanche breakdown, a large current flow causes a buildup of voltage at

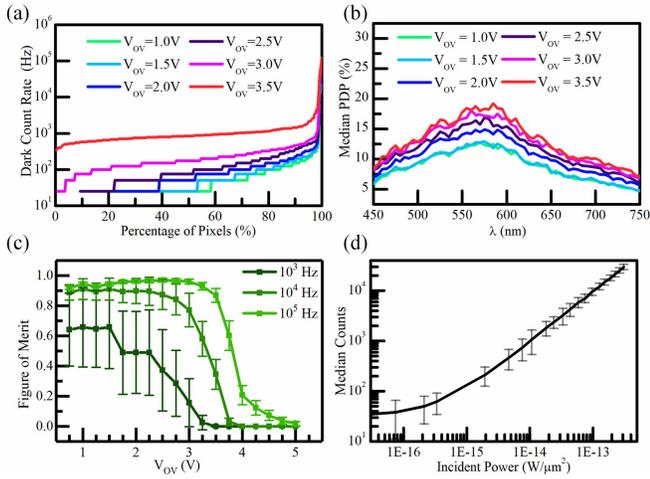


Fig. 5. (a) Dark count rate of 512 pixels. (b) Median PDP versus incident wavelength for 350-nW isotropic illumination. (c) FoM [12] at three different fluences plotted versus  $V_{OV}$ . (d) SPAD count linearity versus incident light intensity. Error bars indicate the standard deviation over 512 pixels.

the input of the comparator. The output of the comparator is flipped and triggers the event detection flip-flop, incrementing the counter by one. Only the events within the  $ON$  signal are detected and at most one photon can be detected per cycle. When the counter reaches 63 counts, the  $Full$  signal is asserted, and the counter stops incrementing. The timing diagram in Fig. 4(b) demonstrates a cycle with zero counts and a cycle with one count. During the zero-count cycle, the internal delay of the self-timing reset mechanism and the falling edge of  $ON$  show the “effective” time-gate. This can be compensated with minor adjustments to the duty cycle of the  $ON$  signal. During the one-count cycle, the avalanche is quenched and the counter is incremented by one.

### B. Pixel Performance

The p-type (n-implant) shallow-junction SPAD was manufactured in a 130-nm process with a  $V_{BD}$  of 15.5 V and its key performance metrics were measured in an integrating sphere [7]. In Fig. 5(a), the cumulative percentage of pixels under a given dark count rate (DCR) is plotted for increasing overvoltage ( $V_{OV} = V_{CATH} - V_{BD}$ ), showing a “hot-pixel” percentage of 2%; hot pixels are defined as those with more than five times the median DCR. The median of the arraywise photon detection probability (PDP) is plotted as a function of wavelength in Fig. 5(b). The SPADs show a peak sensitivity at 575 nm.

Fig. 5(a) and (b) shows an increase in both DCR and PDP, respectively, with increasing  $V_{OV}$ , as measured with a monochromatic light source centered at 520 nm that matches the emission peak of EGFP. We use a figure of merit (FoM; see [12, eq. (2)]) based on photon counting statistics for fluorescence imaging to evaluate the SPAD performance. This FoM selects the highest probability of detecting photon events while avoiding false positives and noise events. In Fig. 5(c), this FoM is plotted for fluences over a range of  $10^3$ – $10^5$  photons/s, equivalent to a photon flux of

TABLE I  
COMPARISON TABLE WITH PRIOR SPAD IMAGERS

	Schwartz [14]	Lee [15]	Perenzoni [16]	Field [17]	This work
Technology	0.35 $\mu\text{m}$	0.18 $\mu\text{m}$	0.35 $\mu\text{m}$	0.13 $\mu\text{m}$	0.13 $\mu\text{m}$
Pixel Pitch	40 $\mu\text{m}$	35 $\mu\text{m}$	50 $\mu\text{m}$	48 $\mu\text{m}$	25.3 $\mu\text{m}$
Pixel Count	64 $\times$ 64	72 $\times$ 60	160 $\times$ 120	64 $\times$ 64	4 $\times$ 128
Power	-	83.8 mW	157 mW	26 W	6.24 mW
Temporal Resolution	350 ps	71 ps	194 ps	62.5 ps	137 ps
DCR Median	1059 Hz @ 2.7 V	400 Hz @ 1.2 V	580 Hz @ 3 V	544 Hz @ 1.5 V	40 Hz @ 1 V
Max PDP	4.8% @ 2.7 V	2.7% @ 1.2 V	-	30% @ 1.5 V	12.4% @ 1 V
Fill Factor	14%	9.6%	21%	0.77%	6.3%
Frame Rate	1144 fps	-	486 fps	100 fps	51 kfps

\*PDP and DCR are reported at respective excess bias voltages. PDP is reported as the maximum value over all wavelengths.

0.4–40  $\text{fW}/\mu\text{m}^2$  at a 520-nm wavelength. Due to the relatively low DCR compared to fluence, the optimal operating value of  $V_{OV}$  is found to be 1 V. Fig. 5(d) shows the linearity of photon counts in response to increasing fluence at 520 nm. The linearity is limited by the dark count at low fluence and by photon pileup at high fluence.

$V_{OV}$  at 1 V also has the benefit of reduced afterpulsing events [13]. Afterpulsing is estimated by measuring DCR while sweeping the active-reset SPAD dead time from 300 to 10 ns. The median 40-Hz dark count holds constant across this range, indicating that afterpulsing is not a significant contributor to DCR.

Table I reports the comparison with prior time-gated SPAD imagers. The imager compares favorably with other time-gated SPAD imagers in conventional planar array formats [14]–[17] while displacing a volume of only 0.03  $\text{mm}^3$  with the implanted shanks.

### C. Time-Gating Circuitry

The system employs a DTC to synthesize the global shutter  $ON$  signal with a tunable delay with respect to the laser pulse. Fig. 6 shows its architecture and timing diagram.  $TRIGGER$  determines the  $ON$  repetition rate. The Johnson counter creates an 8-bit sequence signal,  $CNT$ , where each bit corresponds to an 8  $\times$  down-sampled version of  $CLKIN$ , each 45° delayed and fed into both phase interpolators (PIs). A user-defined 8-bit programmed phase is decoded to select any combination of three phases of  $CNT$ . These are equally summed into  $\Sigma$  [see Fig. 2(b)], low-pass filtered, and buffered [18]. While  $CLK$  is generated in the first PI,  $REF$  is generated from a second PI to flexibly compensate for internal delays, and its rising edge asserts  $READY$  to allow for the next synthesized  $CLK$  to propagate to the output  $ON$ .  $READY$  is pulled low when  $ON$  is activated and asserted again on the rising edge of  $REF$  when  $TRIGGER$  is high. The dynamic range of pulse width and delay is extended through the final tunable counter stage. A serial peripheral interface (SPI) slave programs phase width, and delay from the control FPGA.

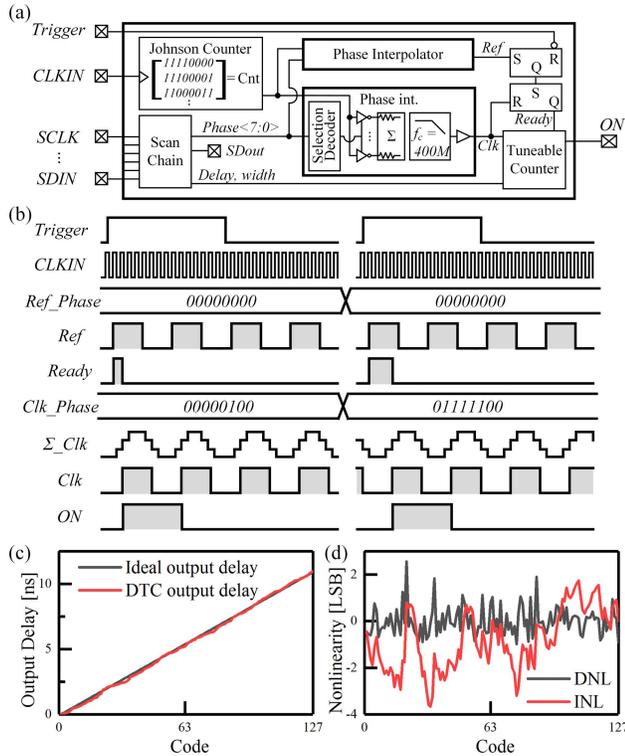


Fig. 6. (a) Circuit diagram of DTC. (b) Timing diagram with  $F_{CLKIN} = 32 \times F_{TRIGGER}$  and tuneable counter width = 1 and delay = 0. (c) Ideal and synthesized  $ON$  signal delays and (d) conversion nonlinearity.

A variable output duty cycle of 20%–50% for  $ON$  allows the SPAD to be moved out of Geiger mode after sufficient decay of the fluorescence signal, avoiding dark count and unnecessary power dissipation. As  $CLK$  and  $REF$  are  $8 \times$  down-sampled, 50% duty cycle, phase-delayed versions of  $CLKIN$ , the  $CLKIN$  frequency determines the duty cycle ( $D$ ) resolution of  $ON$ . The  $CLKIN$  frequency must always be greater than or equal to  $8 \times (F_{TRIGGER}/D)$ , where  $F_{TRIGGER}$  is equal to the laser repetition rate, in order to ensure correct phase and duty cycle for  $ON$ , with higher frequencies allowing higher resolution phase control.

All tests were performed at a 10-MHz  $TRIGGER$  repetition rate. If one keeps the  $REF$  edge constant, the DTC allows  $CLK$  timing to be tuned relative to  $REF$  with an 8-bit resolution in a range of the  $REF$  and  $CLK$  period.

In the case of fluorescence decay lifetimes on the order of 2–4 ns,  $REF$ -to- $CLK$  delays up to 12.5 ns are required, equivalent to 50% of the available range (7 bits of the DTC).  $CLKIN$  is set to a 320-MHz frequency to synthesize a 40-MHz  $CLK$ , allowing for a 25% duty cycle for  $ON$ . Fig. 6(c) shows how 128 LSB (seven bits) for  $CLK$  results in 97-ps resolution. Fig. 6(d) shows the measured linearity performance of the converter, with DNL staying within  $[-1, +2.5]$  LSB and INL staying within  $[-4, +2]$  LSB, with standard deviations over the nonlinearity curve of 0.6 and 1.3 LSB, respectively. From the measurement data in Fig. 6(c), we calculate the signal-to-noise-and-distortion (SINAD) of the DTC as [19]

$$\text{SINAD} = 20 \log_{10} \left( \frac{T_{fs}}{\sigma_{\epsilon}} \right) = 40.95 \text{ dB} \quad (1)$$

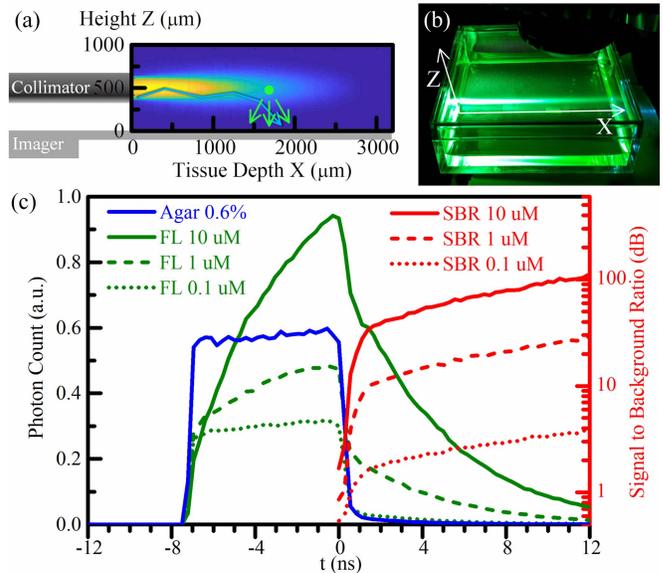


Fig. 7. (a) Diagram of fluorescence measurement setup showing fluorescent and scattered excitation photons reaching the SPAD array located along the  $x$ -axis. (b) Photograph of 10- $\mu\text{M}$  fluorescein in 0.6% agarose, excited by a collimated 488-nm pulsed laser. (c) Time-gated photon counts synchronized to the laser pulse at  $t = 0$  s for pure agarose (blue line) and fluorescein in agarose (FL, green line) plotted against the left axis; respective SBR (Fluorescein:Agarose) is plotted on the right axis (red line).

where  $T_{fs}$  is the full-scale conversion time (12.5 ns), and  $\sigma_{\epsilon} = 0.112$  ns represents the standard deviation of the output error. This leads to an effective number of bits (ENOB) of 6.5 bit or an effective resolution of 137 ps.

#### D. Time-Gating Performance

To demonstrate the effectiveness of time-gated fluorescence imaging, the imaging probe was placed under media with properties similar to a fluorescence-expressing mouse brain [see Fig. 7(a)]. We used a 0.6% agarose gel (A7777, Teknova) to imitate the optical scattering of neural tissue [20]–[22], and a range of fluorescein concentrations was added to it to simulate bulk fluorescence expression [see Fig. 7(b)]. A fluorescein concentration of 10  $\mu\text{M}$  is equivalent in brightness to cytoplasmic expression of GCaMP in a neuron [23]. A picosecond-pulsed laser (Fianium SC450-pp, NKT Photonics) delivers excitation light through a collimated beam 500  $\mu\text{M}$  above the imager with 670- $\mu\text{W}$  average power ( $1.6 \times 10^6$  photon/s), and the time-gated  $ON$  signal with a 10-ns on-time is moved in steps of 280 ps over the range of 12 ns after the laser pulse.

In the scattering medium, the excitation laser light reaches the pixels after traveling along a Mie scattering path, which we find representative of neural tissue without the presence of fluorescence. The resulting photon count [see Fig. 7(c), Agar] is the convolution of the SPADs' 10-ns time-gate with the 50-ps-wide laser pulse, which is denoted as the instrument response function (IRF) of the time-gated quenching circuit. Its sub-nanosecond exponential decay characteristic is the result of minority carriers, created by the high excitation photon flux, diffusing into the multiplication region of the SPAD and creating an avalanche breakdown in the absence of an incoming photon [14]. The time-gated rejection ratio

2 ns after the pulse is 98%, corresponding to an excitation OD of 1.7.

In the case of a fluorescent medium representative of an EGFP-expressing mouse cortex [see Fig. 7(c), FL], direct laser intensity is not observed on the imager because most of the laser light is converted into fluorescence emission before it reaches the imager. Instead, we see a slow rise and 4.7-ns lifetime fall produced by a convolution of the excitation light time-gate and the 4.1-ns fluorescence lifetime of fluorescein.

Time-gated fluorescence imaging of fluorescent markers in a scattering environment requires a time window in which the fluorescence response exceeds that of the scattered excitation light. In the case of 10- $\mu\text{M}$  fluorescein in agarose, the SBR reaches 31.6 dB 2 ns after the laser pulse. SBR and excitation OD continue to increase past 2 ns, reaching 105 dB and 3.2 OD, respectively, at 12 ns; however, this is achieved at the cost of reduced signal yield (5% at 12 ns). While time-gated filtering alone does not match the OD of epifluorescent microscopes, relatively low-OD spectral filters can be added on top of the imager to supplement the rejection from time gating.

When the imager is illuminated with the output-saturating photon flux of  $2.2 \times 10^6$  photons/s/ $\mu\text{m}^2$  and reads out at the maximum frame rate of 51 kfps, operating at maximum power consumption, it consumes a total of 6.24 mW. The charging of the SPAD diode capacitance at quenching events consumes 2.94 mW on the  $V_{\text{RST}}$  and SPAD cathode nodes together. The 1.5-V digital core consumes 3.3 mW. The 3.3-V IO power was supplied through the FPGA. With the majority of the digital logic located on the base of the shank, we estimate a power consumption of  $\sim 3$  mW for the implantable regions of the imager, comparable to levels reported by other implantable CMOS neural recording shanks [24].

### E. Photometry in Scattering Neural Tissue

To test the imaging performance on a fluorescent target with known size and properties, a Monte Carlo photon trajectory simulation [25]–[27] is performed in a tissue model characteristic of EGFP expression in gray matter. The tissue is modeled with anisotropy factor  $g = 0.88$ , refractive index  $n = 1.37$ , scattering coefficient  $\mu_s = 21 \text{ mm}^{-1}$ , and absorption coefficient  $\mu_a = 0.06 \text{ mm}^{-1}$  [28], while EGFP is modeled with a quantum yield of 0.6, an extinction coefficient of  $55000 \text{ M}^{-1}\text{cm}^{-1}$  [29], and a lifetime of 4.1 ns. Using a simulation setup similar to Fig. 7(a), we model an imager implanted along the x-axis with the pixels located between  $x = 0$  and 3.2 mm, in parallel with a collimated pulsed laser light source illuminating the tissue boundary of  $x = 0 \mu\text{m}$ , at a distance  $z = 100 \mu\text{m}$  away from the imager with 700- $\mu\text{W}$  average power. A group of fluorescently labeled somata contained in a 50- $\mu\text{m}$  radius located at height  $z = 100 \mu\text{m}$ , expressing EGFP with an equivalent concentration of 10  $\mu\text{M}$  [23], is swept along the x-axis. The model assumes a duty cycle of 50% and a 12-ns time-gated delay after the excitation pulse, which is the shortest delay at which SBR exceeds one at all depths. At this delay, the filter achieves an OD of 3.2 and the fluorescence emission falls to 5% of its peak intensity.

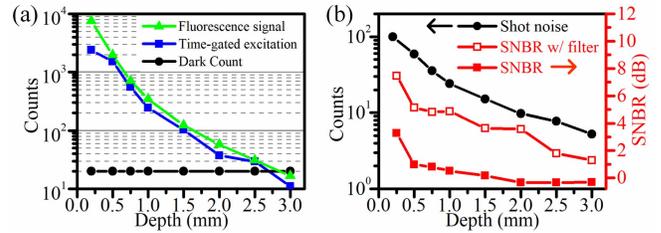


Fig. 8. (a) Fluorescence emission of EGFP-expressing somata in model tissue compared with background and dark count as a function of tissue depth. (b) Combined photon shot noise and dark count shot noise, plotted with SNBR as a function of tissue depth.

We perform these simulations at a frame rate of 1 fps. For a laser repetition rate of 20 MHz, this means that photon counts for  $2 \times 10^7$  laser pulses are integrated. Fig. 8(a) shows the resulting counts for the time-gated excitation source, the fluorescence emission, and the dark count. The fluorescence emission assumes that the group of somata is positioned directly above the sensing SPAD at each depth. Fluorescence emission is found to overcome scattered excitation at all simulated depths, although by a small margin. Fig. 8(b) shows the signal-to-noise-and-background ratio (SNBR), defined as the ratio of fluorescent signal counts to the background, photon-shot-noise, and dark-count-shot-noise. For time-gated delays longer than 12 ns, the drop in signal yield for a given frame rate leads to an increase in shot noise relative to signal. A frame rate of 1 fps is necessary to reduce the shot noise to yield SNBR larger than 3 dB. SNBR is largest at tissue depths below 500  $\mu\text{m}$  because scattering has not yet become significant. When the imager is deeper in tissue, excitation collimation is lost and more blue excitation light scatters into the detector. However, due to the limited acceptance angle of  $60^\circ$  (see Section IV-A), the majority is still rejected, and the excitation background and fluorescence signal reduce proportionally between 500  $\mu\text{m}$  and 2.5 mm.

Beyond 2.5 mm, the exponential extinction of the excitation light places the detector into dark-count-limited operation. In the absence of better excitation rejection, the SNBR of our imager is sufficient up to a depth limit, and deep brain imaging becomes a problem of light delivery. Potential solutions have been demonstrated, including implantable waveguides [30]–[32], densely integrated GaN  $\mu\text{LEDs}$  [33], [34], or multimode fiber endoscopes [35].

Another solution is to incorporate spectral filters that work together with time gating. In Fig. 8(b), we repeat the same analysis with the addition of a filter that is able to block the excitation wavelength at an OD of only 0.6 [36], but which acts to further reduce the background and photon shot noise associated with this background. The SNBR, in this case, is sufficient to allow imaging at full shank depths. Imaging at the hardware limit of 51 kfps would be possible given high enough illumination. The frame rate is in direct tradeoff with shot-noise-limited SNBR.

## IV. LENS-LESS IMAGING USING NEAR-FIELD ANGLE-SENSITIVE DETECTION

The lack of refractive focusing optics poses two challenges in image formation. The first is low photon yield, which is

addressed in part through the use of sensitive SPAD detectors. The second is the lack of spatial resolution. Simply relying on the limited numerical aperture of the detectors in the array itself (which is an acceptance cone of approximately  $60^\circ$  single-sided) would allow an  $x$ - or  $y$ -spatial resolution of only  $\sim 450 \mu\text{m}$  to be achieved  $200 \mu\text{m}$  from the detector. Instead, we rely on near-field diffraction gratings on each pixel to provide an angle sensitivity diversity that can be used for computational image reconstruction [16]. The use of these gratings comes at a cost in transmission efficiency, which averages only 3% of incident photons. This can be improved to 29% using non-blocking dielectric phase gratings to achieve the same diffraction gratings instead of metal wires [37].

The inverse imaging problem, in this case, is formulated as a linear system from the scene  $x$  to the array photon count  $y$

$$y = Ax + \epsilon. \quad (2)$$

The sensing matrix  $A$  compresses the scene into a much lower dimensionality determined by the pixel count. Dark count and background  $\epsilon$  are assumed to be spatially and temporally uncorrelated to the sources, and afterpulsing was estimated to be negligible. Image quality is improved by constructing a sensing matrix  $A$  that maximizes information extraction from the scene  $x$ . Such an optimized  $A$  allows for the best reconstruction of the inverted image  $\hat{x}$ . In particular, we seek  $A$  that has maximally incoherent columns [38], indicating that each location in space is compressed onto the imager's response in a maximally distinct way.

### A. Angle-Sensitive Pixel Design

$A$  is determined by a diversity of angle sensitivity introduced into each pixel with near-field diffraction gratings, as shown in Fig. 9. These gratings, formed in two layers of back-end metal separated by  $1.14 \mu\text{m}$  [see Fig. 9(a)], constitute 16 angular-sensitivity variations (which take up a total distance along the length of the shank of  $200 \mu\text{m}$ ) consisting of orthogonal combinations of two angular modulation frequencies, two rotations, and four quadrature phases [8]. The gratings are employed across both shanks in a repeating pattern of  $2 \times 8$  pixels to ensure that a source located  $100 \mu\text{m}$  above the center axis of the shank and positioned (in  $x$ ) at the midpoint of the 16-pixel group is situated in the  $60^\circ$  FoV of each pixel variation.

The angular response for each grating combination was measured by illuminating the imager with a collimated monochromatic light source at  $520 \text{ nm}$  and varying the angles of incidence along  $\theta$  and  $\varphi$ , two azimuthal angles pivoting around the  $x$ - and  $y$ - axes, respectively, using a rotational and tilting stage under the imager [see Fig. 9(b)]. Two examples of grating structures, each with a low and high angular frequency, are shown in Fig. 9(c), along with the parametric fit.

The higher angular modulation amplitudes at extreme angles in Fig. 9(b) reflect the fact that the Talbot self-image is not generated at the exact height of the secondary grating. At even higher incident angles, the CMOS metal interconnect starts occluding photons for incident angles larger than  $45^\circ$

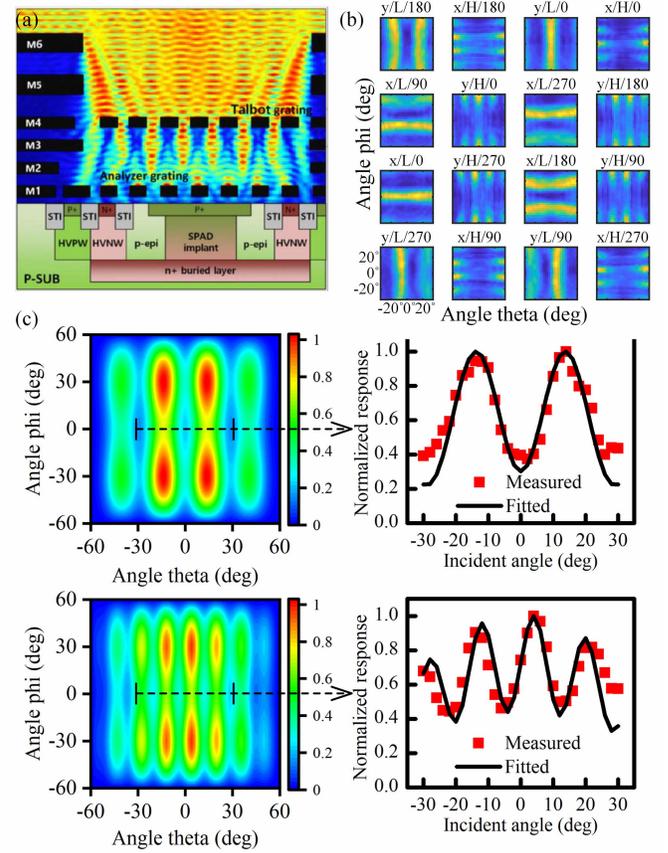


Fig. 9. (a) Illustration of an angular-sensitive SPAD implemented in CMOS. (b) Measured angular modulation of 16 different grating structures in direction ( $x$  and  $y$ ), frequency (high low), and quadrature phase ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ ), for  $520\text{-nm}$  illumination. (c) Simulated angular responses of grating combinations  $y/L/180^\circ$  (top left) and  $y/H/90^\circ$  (bottom left), with respective  $\varphi = 0^\circ$  cross sections measured and parametrically fit (right).

and rejects completely at  $60^\circ$ . In addition, the PDP at green wavelengths is lower for obliquely incident photons due to a deeper effective multiplication layer, shifting peak PDP away to a higher wavelength.

### B. Point Spread Function

The point spread function (PSF) describes the system response to a point source, and its width determines how close two point sources can be placed while still being individually reconstructed. The PSF of this imaging system can be computed as a pseudoinverse backprojection of a single voxel

$$x_{\text{PSF}} = (A^+ A) x_{\text{source}} \quad (3)$$

where  $A^+$  is the left pseudoinverse of  $A$  and  $x_{\text{source}}$  is an  $n_{\text{voxel}}$ -by-1 array that is 0 except at a single voxel.  $n_{\text{voxel}}$  is  $\sim 2$  million in the case of using a voxel grid of  $5 \mu\text{m}$ . The PSF of a single voxel located at Cartesian coordinates of  $[2000, 0, 200 \mu\text{m}]$  is plotted in Fig. 9(a)–(c), showing the correlated nature of voxels in this highly underdetermined system. Fig. 10(d) shows the resolution in the form of full-width at half-maximum (FWHM) of the 1-D PSF profile. For continuity of the plot, data points in which the local minima

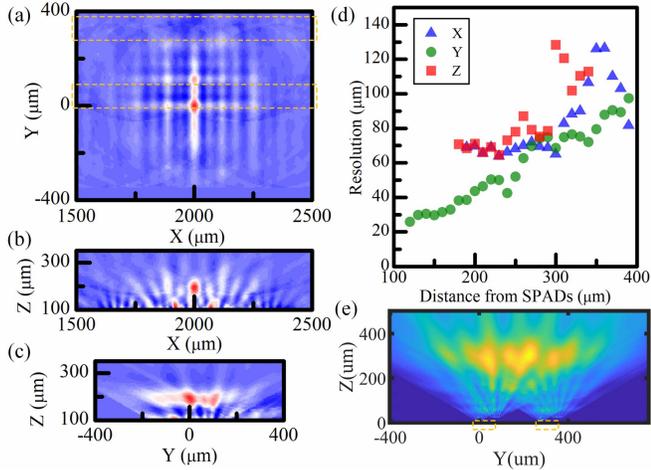


Fig. 10. PSF of a single point source located at coordinates [2000, 0, 200 μm] portrayed as (a) xy, (b) xz, and (c) yz plane cross sections of a 3-D pseudoinverse backprojection; voxels are 5 μm across all dimensions. (d) PSF FWHM in x, y, and z with respect to the increasing distance of point source from imager. (e) xz plane cross section of imager FoV. Dotted yellow lines in (a) and (e) mark the location of the imager.

closest to the peak did not fall below the half-maximum were omitted.

The broadening of the FWHM with increasing source distance results from an approximately constant angular resolution translating to a larger spatial spread at further distances. The minimum resolution is measured to be 64, 26, and 65 μm in the x-, y-, and z-directions, respectively. In comparison to a GRIN lens-based miniature microscope [3], while the imager resolution is 1–2 orders of magnitude worse, it images a brain volume larger by the same factor. The imaging volume can be sculpted further by restricting the illumination profile to a volume of interest, drastically reducing the image background.

## V. IMAGING PERFORMANCE

### A. Single Point Source Imaging

To mimic a cluster of neuron somata, a diffuser tip of 100-μm radius (Thorlabs CFDSB20), coupled to a monochromatic green LED (300 μW and  $7.9 \times 10^{14}$  photons/s), was placed 200 μm above the imager and translated along its length at three different positions (see Fig. 11, *p1*, *p2*, and *p3*). The image was acquired with a 0.8-fps frame rate and a photon accumulation time of 200 ms, yielding a maximum of 41k counts at the pixel directly under the light source. The repeating pattern of 16 gratings is responsible for the periodicity of raw counts in Fig. 11(a) and contributes to localization of the point source. Pseudoinverse backprojections ( $x_{\text{proj}} = A^+y$ ) sectioned at  $z = 200$  μm show the imager's ability to find centers of brightness with a 30-μm standard deviation [see Fig. 11(b)].

### B. Multi-Source Phantom Imaging

A multi-source phantom fluorescent target was constructed by depositing fluorescent microspheres (F8836 10 μm, Thermofisher) on a cover glass, which was then placed

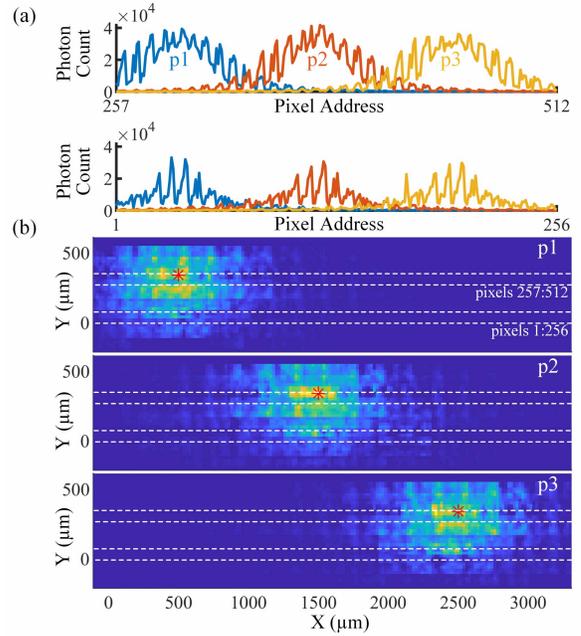


Fig. 11. Monochromatic light source hovering 200 μm above the imager was placed at x coordinates of 500 μm (*p1*), 1500 μm (*p2*), and 2500 μm (*p3*). (a) Raw 512-pixel counts along two shanks and (b) pseudoinverse backprojection in plane  $z = 400$  μm with source locations (\*) and probe (-) overlaid.

400 μm away from the imager. A picosecond-pulsed excitation (Fianium SC450-pp, NKT Photonics) centered at 480 nm and carrying 700 μW of average power ( $1.7 \times 10^{15}$  photons/s) with a 20-MHz repetition rate was delivered in a direction parallel to the shank with a 500-μm-diameter collimator. This parallel incident angle for the excitation power minimizes the image background produced by incomplete time-gated rejection. Fig. 11(a) and (b) show two distinct scenes with a difference in the size and arrangement of the imaging target. The fluorescence microscope image is displayed alongside the reconstructed pseudoinverse backprojections of SPAD counts. Photon counts were accumulated for 200 ms at 1 fps with a time-gated delay of 550 ps after the laser pulse, resulting in a maximum SBR of 15 dB. Voxel sizes are 20 μm in x, y, and z dimensions to optimize the computation speed.

To show the efficacy of time-gated fluorescence imaging, a third scene [see Fig. 12(c)] is constructed with two microsphere clusters, one composed of fluorescent spheres and another with scattering latex spheres. Despite the visibility of both clusters under brightfield imaging, the imager selectively detects only the fluorescent microspheres.

### C. Source Localization by $L1$ -Norm Minimization

Due to the highly compressed nature of the imaging system which maps  $\sim 10^5$  voxels to 512 pixels when constructing a 3-D image with 20-μm voxels, a direct backprojection image will inevitably have limited spatial resolution and appear blurred. Furthermore, high frequency noise is amplified into the image through high spatial sampling frequencies, associated with low magnitude singular values of the sensing matrix *A*. By incorporating a sparsity constraint, the brightest few

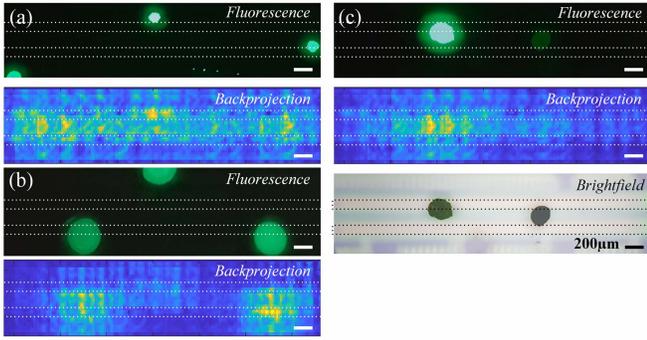


Fig. 12. (a) 10- $\mu\text{m}$ -diameter microsphere clusters deposited on cover glass placed 400  $\mu\text{m}$  above imager (dotted line overlay). Top: epifluorescence microscopy image. Bottom: pseudoinverse backprojection. (b) Same with larger clusters. (c) Two sphere types, fluorescent (left) and scattering (right), showing selective imaging of fluorescent target. Top: fluorescent. Middle: backprojection. Bottom: brightfield. All scale bars are 200  $\mu\text{m}$ .

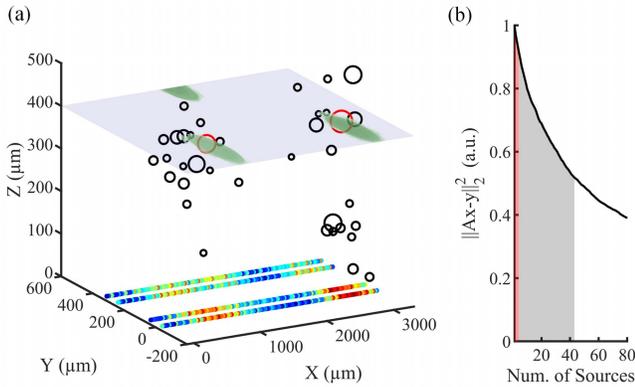


Fig. 13. (a) 3-D reconstruction of microsphere clusters in Fig. 12(b); two most significant solutions are marked in red circles and the following 40 in black, with diameters depicting amplitude estimate; the overlaid ground-truth image indicates the location of microspheres 400  $\mu\text{m}$  above the chip. (b)  $L_2$ -norm of residual error decreasing with added sources, red and gray corresponding to first 2 and next 40 voxels in the order of highest amplitude estimate.

source locations can be solved for in a least-mean-squares manner. Performing this basis pursuit denoising optimization allows an estimation of best-fit source locations while penalizing the number of total sources with a weighting parameter  $\lambda$  [39]. This optimization problem employs a cost function that includes the  $L_1$ -norm of sources ( $f_{L1}$ ) and the  $L_2$ -norm of the residual error ( $f_{L2}$ )

$$\hat{x} = \arg \min_x \left( \lambda \|x\|_1 + \frac{1}{2} \|Ax - y\|_2^2 \right) = \arg \min_x (f_{L1} + f_{L2}). \quad (4)$$

To solve this optimization problem, we use the in-crowd algorithm [40] to solve for sparse locations within the imager's FoV in real time. Fig. 13(a) shows the 3-D reconstructed image of the microsphere arrangement in Fig. 12(b) solved with 20- $\mu\text{m}$  voxel sizes. Raw imager data are drawn as a heat map at the plane of the imager ( $z = 0$   $\mu\text{m}$ ). The largest magnitude contributors to the solution ( $\hat{x}$ ) are displayed in circles, while circle diameters indicate the magnitude estimate. The two voxels with the largest magnitude, colored red, show

the imager's ability to locate the two microsphere clusters at their correct volumetric locations. The next highest 40, colored black, are also found to gravitate around the two clusters with decreasing amplitude estimates. As more voxels are added to the solution  $\hat{x}$ , the  $L_2$ -norm of the residual error ( $f_{L2}$ ) declines as shown in Fig. 13(b), but each with a diminishing contribution to residual error minimization. By combining this information from residual error with the appearance of clusters of estimated sources, we can heuristically infer sparsity and location without relying on an external verification of ground truth.

#### D. Limit of Reconstruction Accuracy

We perform a Monte Carlo simulation to determine the maximum number of sources that the imaging probe can resolve simultaneously. Sparse reconstruction theory states that the maximum number of detectable sources is the number of pixels  $M$ . However, for  $L_1$ -minimization solvers that assume no constraints on the sensing matrix  $A$ , phase transition occurs around  $\sqrt{M}$  [41], meaning that a sharp decline in reconstruction reliability occurs above that density. In the case of the imaging probe, the columns of  $A$  have a repetitive structure determined by the angular-sensitive gratings. As a result, we expect the maximum allowable sparsity to be below  $\sqrt{M}$ .

We place  $K$  sources randomly in the imaging volume of 3.2 mm  $\times$  200  $\mu\text{m}$   $\times$  150  $\mu\text{m}$  above a single shank at 250- $\mu\text{m}$  depth. Photon shot noise is applied to the simulated fluorescence data, and additional excitation background is added with the expected SBR of 3.5 dB. Source placement and subsequent localization are repeated  $N = 1000$  for each value of  $K$ . To correctly compare the localization result against the ground truth, an exhaustive search method links found sources with their closest true locations, where computational tractability limits the analysis to  $K = 10$ . We analyze the localization error in terms of mean square error (MSE), allowing a decomposition into the average error squared and variance of the error

$$\text{MSE}(x, \hat{x}) = \frac{1}{N} \|x - \hat{x}\|_2^2 = \left( \frac{1}{N} \sum (x - \hat{x}) \right)^2 + \sigma_{x - \hat{x}}^2. \quad (5)$$

Fig. 14 shows how localization performance is affected by the source cardinality,  $K$ . Fig. 14(a) shows the distribution of the error over all iterations and source cardinalities in each dimension. Large outliers are essentially ground-truth locations that are not found, and the larger imaging volume in the  $x$ -direction permits larger outliers along that axis. The error is distributed with 98% of sources found within 130  $\mu\text{m}$ , or around two times the largest FWHM at 100- $\mu\text{m}$  distance, of the true location. The compressive sensing method has a 2% likelihood for introducing false positives and negatives, which need to be accounted for in the subsequent imaging target analyses.

The total ensemble standard deviation of all errors in all directions reduces to 17  $\mu\text{m}$  [see Fig. 14(b)] when corrected for these outliers. The error increases with more sources due to increased noise sensitivity at higher source densities. Both the standard deviation and the bias of the error

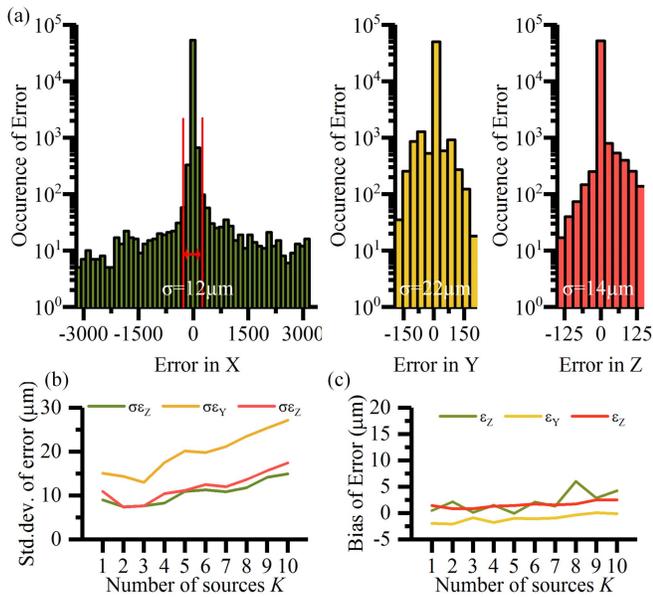


Fig. 14. (a) Distribution of localization error in spatial coordinates. (b) Standard deviation when accounted for 2% false positive/false negative outliers. (c) Bias of localization error.

[see Fig. 14(c)] are smaller than the pixel pitch in all directions x, y, and z. The sign of the data point indicates an estimation bias in the respective direction.

Even at the reported large variance, a low error bias indicates the imager's strength at localizing volumetric centers of brightness with micrometer accuracy, where precision improves with the repeated measurements and averaging over multiple reconstructions. We, therefore, expect that a single shank with 256 pixels can localize up to ten point sources simultaneously with 98% certainty at SBR levels comparable with those estimated for an *in vivo* environment.

## VI. CONCLUSION

In this article, we present an implantable neural shank-based imager with a minimal tissue displacement of only  $0.03 \text{ mm}^3$  while monitoring a volume of  $0.4 \text{ mm}^3$ . The minimally invasive imager is able to localize the fluorescent objects at depths beyond those of conventional imaging systems. It is capable of distinguishing multiple fluorescent objects with a resolution of 64, 26, and  $65 \mu\text{m}$  in the x-, y-, and z-directions, respectively, without the use of spectral filters and refractive focusing optics. Time gating adds two orders of magnitude of excitation light rejection, adding to what can be provided by thin-film absorptive optical filters. The imager's modest pixel count of 512 trades off spatial resolution with fast image acquisition; its high maximum frame rate (51 kfps), synchronized with an external pulsed excitation source, allows the investigation of sparsely expressed biomarkers for *in vivo* neural imaging.

## ACKNOWLEDGMENT

The authors would like to thank the TSMC Foundry for their full support in test chip fabrication and B. Ahn for the helpful discussion on image reconstruction.

## REFERENCES

- [1] S. Weisenburger *et al.*, "Volumetric  $\text{Ca}^{2+}$  imaging in the mouse brain using hybrid multiplexed sculpted light microscopy," *Cell*, vol. 177, no. 4, pp. 1050–1066 e14, May 2019.
- [2] T. Wang *et al.*, "Three-photon imaging of mouse brain structure and function through the intact skull," (in English), *Nature Methods*, vol. 15, no. 10, pp. 789–792, Oct. 2018.
- [3] M. J. Levene, D. A. Dombeck, K. A. Kasichke, R. P. Molloy, and W. W. Webb, "In vivo multiphoton microscopy of deep brain tissue," (in English), *J. Neurophysiol.*, vol. 91, no. 4, pp. 1908–1912, Apr. 2004.
- [4] O. Skocek *et al.*, "High-speed volumetric imaging of neuronal activity in freely moving rodents," (in English), *Nature Methods*, vol. 15, no. 6, pp. 429–432, Jun. 2018.
- [5] G. Corder, B. Ahanonu, B. F. Grewe, D. Wang, M. J. Schnitzer, and G. Scherrer, "An amygdalar neural ensemble that encodes the unpleasantness of pain," (in English), *Science*, vol. 363, no. 6424, pp. 276–281, Jan. 2019.
- [6] J. K. Adams *et al.*, "Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope," (in English), *Sci. Adv.*, vol. 3, no. 12, Dec. 2017, Art. no. e1701548.
- [7] C. Lee *et al.*, "11.5 A 512-Pixel 3kHz-frame-rate dual-shank lensless filterless single-photon-avalanche-diode CMOS neural imaging probe," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2019, pp. 198–200.
- [8] A. Wang, P. Gill, and A. Molnar, "Light field image sensors based on the Talbot effect," (in English), *Appl. Opt.*, vol. 48, no. 31, pp. 5897–5905, Nov. 2009.
- [9] H.-S. Koo, M. Chen, and P.-C. Pan, "LCD-based color filter films fabricated by a pigment-based colorant photo resist inks and printing technology," (in English), *Thin Solid Films*, vol. 515, no. 3, pp. 896–901, 2006.
- [10] Y. Yu, L. Wen, S. Song, and Q. Chen, "Transmissive/reflective structural color filters: Theory and applications," (in English), *J. Nanomater.*, vol. 2014, p. 6, Apr. 2014.
- [11] Y. Gu, L. Zhang, J. K. Yang, S. P. Yeo, and C.-W. Qiu, "Color generation via subwavelength plasmonic nanostructures," *Nanoscale*, vol. 7, no. 15, pp. 6409–6419, Apr. 2015.
- [12] R. M. Field, J. Lary, J. Cohn, L. Paninski, and K. L. Shepard, "A low-noise, single-photon avalanche diode in standard  $0.13 \mu\text{m}$  complementary metal-oxide-semiconductor process," *Appl. Phys. Lett.*, vol. 97, no. 21, 2010, Art. no. 211111.
- [13] D. Bronzi, S. Tisa, F. A. Villa, S. Bellisai, A. Tosi, and F. Zappa, "Fast sensing and quenching of CMOS SPADs for minimal afterpulsing effects," *IEEE Photon. Technol. Lett.*, vol. 25, no. 8, pp. 776–779, Apr. 15, 2013.
- [14] D. E. Schwartz, E. Charbon, and K. L. Shepard, "A single-photon avalanche diode array for fluorescence lifetime imaging microscopy," *IEEE J. Solid-State Circuits*, vol. 43, no. 11, pp. 2546–2557, Nov. 2008.
- [15] M. Perenzoni, N. Massari, D. Perenzoni, L. Gasparini, and D. Stoppa, "A  $160 \times 120$ -pixel analog-counting single-photon imager with Sub-ns time-gating and self-referenced column-parallel A/D conversion for fluorescence lifetime imaging," (in English), *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 155–167, Jan. 2016.
- [16] C. Lee, B. Johnson, T. Jung, and A. Molnar, "A  $72 \times 60$  Angle-sensitive SPAD imaging array for lens-less FLIM," *Sensors*, vol. 16, no. 9, p. 1422, Sep. 2016.
- [17] R. M. Field, S. Realov, and K. L. Shepard, "A 100 fps, time-correlated single-photon-counting-based fluorescence-lifetime imager in 130 nm CMOS," (in English), *IEEE J. Solid-State Circuits*, vol. 49, no. 4, pp. 867–880, Apr. 2014.
- [18] A. Tsimpos, G. Souliotis, A. Demartinos, and S. Vlassis, "All digital phase interpolator," (in English), in *Proc. 10th Int. Conf. Design Technol. Integr. Syst. Nanoscale Era*, Apr. 2015, pp. 1–6.
- [19] Y. Cao, W. De Cock, M. Steyaert, and P. Leroux, "1-1-1 MASH  $\Delta \Sigma$  time-to-digital converters with 6 ps resolution and third-order noise-shaping," (in English), *IEEE J. Solid-State Circuits*, vol. 47, no. 9, pp. 2093–2106, Sep. 2012.
- [20] B. W. Pogue and M. S. Patterson, "Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry," *J. Biomed. Opt.*, vol. 11, no. 4, Jul. 2006, Art. no. 041102.
- [21] R. Pomfret, G. Miranpuri, and K. Sillay, "The substitute brain and the potential of the gel model," *Ann. Neurosci.*, vol. 20, no. 3, p. 118, Jul. 2013.

- [22] N. Sindhvani, O. Ivanchenko, E. Lueshen, K. Prem, and A. A. Linninger, "Methods for determining agent concentration profiles in agarose gel during convection-enhanced delivery," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 626–632, Mar. 2011.
- [23] J. Akerboom *et al.*, "Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design," *J. Biol. Chem.*, vol. 284, no. 10, pp. 6455–6464, Mar. 2009.
- [24] J. J. Jun *et al.*, "Fully integrated silicon probes for high-density recording of neural activity," *Nature*, vol. 551, no. 7679, pp. 232–236, Nov. 2017.
- [25] L. Wang, S. L. Jacques, and L. Zheng, "MCML—Monte Carlo modeling of light transport in multi-layered tissues," *Comput. Methods Programs Biomed.*, vol. 47, no. 2, pp. 131–146, Jul. 1995.
- [26] A. K. Pediredla *et al.*, "Deep imaging in scattering media with selective plane illumination microscopy," (in English), *J. Biomed. Opt.*, vol. 21, no. 12, Dec. 2016, Art. no. 126009.
- [27] Q. Fang, "Mesh-based Monte Carlo method using fast ray-tracing in Plücker coordinates," *Biomed. Opt. Express*, vol. 1, no. 1, pp. 165–175, Aug. 2010.
- [28] G. Yona, N. Meitav, I. Kahn, and S. Shoham, "Realistic numerical and analytical modeling of light scattering in brain tissue for optogenetic applications," *eNeuro*, vol. 3, no. 1, Feb. 2016.
- [29] P. J. Cranfill *et al.*, "Quantitative assessment of fluorescent proteins," (in English), *Nature Methods*, vol. 13, no. 7, pp. 557–562, May 2016.
- [30] A. N. Zorzos, E. S. Boyden, and C. G. Fonstad, "Multiwaveguide implantable probe for light delivery to sets of distributed brain targets," *Opt. Lett.*, vol. 35, no. 24, pp. 4133–4135, Dec. 2010.
- [31] E. Shim, Y. Chen, S. Masmanidis, and M. Li, "Multisite silicon neural probes with integrated silicon nitride waveguides and gratings for optogenetic applications," *Sci. Rep.*, vol. 6, Mar. 2016, p. 022693.
- [32] J. W. Reddy and M. Chamanzar, "Low-loss flexible Parylene photonic waveguides for optical implants," (in English), *Opt. Lett.*, vol. 43, no. 17, pp. 4112–4115, Sep. 2018.
- [33] E. Klein, C. Gossler, O. Paul, and P. Ruther, "High-density  $\mu$ LED-based optical cochlear implant with improved thermomechanical behavior," *Frontiers Neurosci.*, vol. 12, p. 659, Sep. 2018.
- [34] F. Wu, E. Stark, P.-C. Ku, K. D. Wise, G. Buzsáki, and E. Yoon, "Monolithically integrated  $\mu$ LEDs on silicon neural probes for high-resolution optogenetic studies in behaving animals," *Neuron*, vol. 88, no. 6, pp. 1136–1148, Dec. 2015.
- [35] S. A. Vasquez-Lopez *et al.*, "Subcellular spatial resolution achieved for deep-brain imaging *in vivo* using a minimally invasive multimode fiber," *Light: Sci. Appl.*, vol. 7, p. 110, Dec. 2018.
- [36] Z. Li, S. Butun, and K. Aydin, "Large-area, lithography-free super absorbers and color filters at visible frequencies using ultrathin metallic films," (in English), *ACS Photon.*, vol. 2, no. 2, pp. 183–188, Feb. 2015.
- [37] S. Sivaramakrishnan, A. Wang, P. Gill, and A. Molnar, "Design and characterization of enhanced angle sensitive pixels," (in English), *IEEE Trans. Electron Devices*, vol. 63, no. 1, pp. 113–119, Jan. 2016.
- [38] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
- [39] S. B. Chen and D. Donoho, "Basis pursuit," (in English), in *Proc. Conf. Rec. Asilomar C*, 1994, pp. 41–44.
- [40] P. R. Gill, A. Wang, and A. Molnar, "The in-crowd algorithm for fast basis pursuit denoising," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4595–4605, Oct. 2011.
- [41] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," Mar. 2013, *arXiv:1303.6672*. [Online]. Available: <https://arxiv.org/abs/1303.6672>



**Jaebin Choi** (S'14) received the B.S. degree in engineering physics from Cornell University, Ithaca, NY, USA, in 2011, and the M.S. degree in electrical engineering from Columbia University, New York, NY, USA, in 2012, where he is currently pursuing the Ph.D. degree in electrical engineering.

He was a Research Scientist with the Korea Institute of Science and Technology, Seoul, South Korea, from 2014 to 2017. His current research interest includes optoelectronic neural interfaces.



**Adriaan J. Taal** (S'12) received the B.S. and M.S. degrees in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with Columbia University, New York, NY, USA.

His research interests are optoelectronic circuit design, computer vision, and developing imaging applications for neuroscience.



**Eric H. Pollmann** (S'19) received the B.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2017, and the M.S. degree in electrical engineering from Columbia University, New York, NY, USA, in 2018, where he is currently pursuing the Ph.D. degree in electrical engineering.

His research interests include implantable CMOS fluorescence imagers for applications in biology and neuroscience.



**Changhyuk Lee** (S'08–M'14) received the B.S. degree in electrical and computer engineering from Korea University, Seoul, South Korea, in 2005, and the M.S. and Ph.D. degrees from Cornell University, Ithaca, NY, USA. His Ph.D. work focuses on the light-field (angle-sensitive) photon counting CMOS image sensors, neural interface, and radio frequency integrated circuit (RFIC).

He is currently a Senior Research Scientist with the Brain Science Institute, Korea Institute of Science and Technology (KIST), Seoul. He worked at Samsung, Changwon, South Korea, from 2004 to 2006. He was with Cornell University. He was also with Com2us, Seoul, from 2007 to 2008. His current research interests include optoelectronic neural interfaces, implantable biomedical devices, computational neuroscience, and behavioral studies.

Dr. Lee was a recipient of the Qualcomm Innovation Fellowship from 2013 to 2014.



**Kukjoo Kim** received the B.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008 and 2014, respectively. His thesis focuses on solution-processed graphene and its electronic applications.

He continued as a Post-Doctoral Scholar at the Laboratory of Prof. Kenneth Shepard, Columbia University, New York, NY, USA, in 2016. He is currently a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon. His research interests include flexible/stretchable electronics and neural interfaces.



**Laurent C. Moreaux** received the degree in optoelectronics from Polytech Paris-Sud, Orsay, France, in 1997, the M.S. degree in optics and photonics from the Institut d'Optique, Orsay, in 1998, and the Ph.D. degree in physics from the ESPCI Paris, Paris, France, in 2002, with Dr. J. Mertz and Dr. S. Charpak focusing on the development of novel, non-linear optical microscopy tools and methods for biological imaging.

He continued as a Post-Doctoral Scholar at the Laboratory of Dr. Gilles Laurent, Division of Biology, Caltech, Pasadena, CA, USA, in 2002. He was a permanent Research Scientist with CNRS in 2004, where he continued working on the development of optical techniques to probe neuronal activity in several model systems. He returned to Caltech in 2009, where he has held positions at the Division of Biology and Bioengineering and the Division of Physics, Mathematics and Astronomy. He is currently a permanent Research Scientist with the French National Center for Scientific Research (CNRS), Paris, and a Senior Research Scientist with the California Institute of Technology. He has contributed his technical expertise to many projects in neuroscience at Caltech, including developing methods for whole-cell recordings combined with extracellular arrays and two-photon microscopy in the hippocampus of awake and behaving mice. With the current depth limitation of free-space optical microscopy, he believes that the use of integrated photonics will allow us to directly bring the microscope into the brain to probe the activities of an unprecedented number of neurons. With this vision in mind, his most recent efforts have focused upon fundamental conceptualization, device design, and initial characterization of novel integrated neurophotonic probes.



**Michael L. Roukes** received the B.A. degree in physics and chemistry from the University of California at Santa Cruz, Santa Cruz, CA, USA, in 1978, and the M.S. and Ph.D. degrees in physics from Cornell University, Ithaca, NY, USA, in 1985.

He was a member of Technical Staff/Principal Investigator with Bell Communications Research, Red Bank, NJ, USA. Since 1992, he has been with the California Institute of Technology (Caltech), Pasadena, CA, USA, where he is currently a Frank J. Roshek Professor of physics, applied physics, and bioengineering. His research has spanned mesoscopic electron and phonon transport at ultralow temperatures, physics of nanoelectromechanical systems (NEMS), biomolecular detection and imaging, ultrasensitive NEMS-based mass spectrometry, and ultrasensitive calorimetry.

Dr. Roukes is a fellow of the American Physical Society. He was a recipient of the NIH Director's Pioneer Award. He was named the Chevalier (Knight) of the Ordre des Palmes Académiques by the Republic of France.



**Kenneth L. Shepard** (M'91–SM'03–F'08) received the B.S.E. degree from Princeton University, Princeton, NJ, USA, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1988 and 1992, respectively.

From 1992 to 1997, he was a Research Staff Member and the Manager with the VLSI Design Department, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA, where he was responsible for the design methodology for IBM's G4S/390 microprocessors. He was the Chief Technology Officer with CadMOS Design Technology, San Jose, CA, USA, until its acquisition by Cadence Design Systems in 2001. Since 1997, he has been with Columbia University, New York, NY, USA where he is currently the Lau Family Professor of electrical engineering and biomedical engineering and the Co-Founder and the Chairman of the Board of Ferric, Inc., New York, which commercializes technology for integrated voltage regulators. His current research interests include power electronics, carbon-based devices and circuits, and CMOS bioelectronics.

Dr. Shepard has been an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE-SCALE INTEGRATION SYSTEMS, the IEEE JOURNAL OF SOLID-STATE CIRCUITS, and the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS.