A Mechanically Flexible, Implantable Neural Interface for Computational Imaging and Optogenetic Stimulation over 5.4×5.4mm² FoV

Sajjad Moazeni, Member, IEEE, Eric H. Pollmann, Student Member, IEEE, Vivek Boominathan, Member, IEEE, Filipe A. Cardoso, Jacob T. Robinson, Senior Member, IEEE, Ashok Veeraraghavan, Senior Member, IEEE, and Kenneth L. Shepard, Fellow, IEEE

Abstract—Emerging optical functional imaging and optogenetics are among the most promising approaches in neuroscience to study neuronal circuits. Combining both methods into a single implantable device enables all-optical neural interrogation with immediate applications in freely-behaving animal studies. In this paper, we demonstrate such a device capable of optical neural recording and stimulation over large cortical areas. This implantable surface device exploits lens-less computational imaging and a novel packaging scheme to achieve an ultra-thin (250μm-thick), mechanically flexible form factor. The core of this device is a custom-designed CMOS integrated circuit containing a 160×160 array of time-gated single-photon avalanche photodiodes (SPAD) for low-light intensity imaging and an interspersed array of dual-color (blue and green) flip-chip bonded micro-LED (μLED) as light sources. We achieved 60μm lateral imaging resolution and 0.2mm³ volumetric precision for optogenetics over a 5.4×5.4mm² field of view (FoV). The device achieves a 125-fps frame-rate and consumes 40mW of total power.

Index Terms—all-optical neural interface, computational imaging, lens-less imager, SPAD, optogenetics, flexible packaging

I. INTRODUCTION

Recent advancements in optical functional imaging and optogenetics have brought us closer to deciphering the human brain and inventing the next-generation of brain-computer interfaces (BCI). The advent of genetically encoded calcium and voltage indicators (GECI/GEVI) and optogenetic probes has unlocked new capabilities for intracellular recording in-vivo with near single-action-potential sensitivity and for stimulation with cell-type specificity [1]. Combining both optical functional imaging and optogenetics into a single device enables simultaneous, all-optical, neural interrogation with the potential to revolutionize neuroscience studies. However, these methods are relegated today largely to expensive microscopes based on free-space optics. Miniaturizing such microscopes into an implantable form factor remains an elusive goal. As a first step towards developing such a device, a variety of head-mounted “miniscopes” have been demonstrated [2]–[4]. Using conventional lens-based optics, however, these devices require considerable volume (more than 2cm³) to support field of views (FoVs) in the mm² range. Because of the trade-off between the lens size and FoV in lens-based imagers [5], [6], imaging larger FoVs comes at the cost of much larger device size. For instance, achieving a 7.8×4mm² FoV in a miniscope required a volume of over 24cm³ and mass of 33g [3]. Additionally, state-of-the-art miniscopes either lack or have very limited optogenetic capabilities for stimulation with any spatial selectivity [7]. To build a fully implantable miniscope, a more volume-efficient device is necessary that spans a relatively large cortical area while maintaining a minimally-invasive form factor. We have previously laid out a vision of what might be possible with whole new class of implantable optoelectronic devices [8].

In this paper, we demonstrate a new device ultimately capable of such a fully implantable form factor that supports all-optical neural recording and stimulation over a 5.4×5.4mm² FoV at the target cortical depths of up to 200μm (Layer 2/3 in mouse brain). This mechanically flexible, implantable brain-surface device, illustrated in Fig. 1, enables fluorescence imaging in an ultra-thin (< 250μm-thick) form factor by exploiting a recently developed lens-less computational imaging approach [6]. Device functionality is provided by a die-thinned CMOS integrated circuit consisting of a 2D array of SPAD detectors and flip-chip bonded μLED light emitters for both fluorescence excitation and optogenetic stimulation. While the proof-of-concept device developed is wired, the ultimate incarnation of this device will be interfaced wirelessly, as shown in Fig. 1.

This paper is organized as follows: Section II describes system-level design considerations for an all-optical neural interface. Section III introduces the computational imaging method. Chip architecture and packaging are presented in Section IV and Section V, respectively. Experimental results are...
A. Integrated Light Sources

A fully implantable device must contain all the necessary light emitters for fluorescence excitation and optogenetics that are also capable of reaching the entire FoV with sufficient brightness and spatial selectivity at the target depth. Lasers are typically used in bench-top optical imaging and optogenetics experiments. While continuous-wave (CW) visible light lasers have been previously integrated in a brain implant for optogenetic stimulation [9], poor efficiencies (typically 10%) and the need for an active photonic platform to steer highly focused laser beam to illuminate large FoVs make lasers largely unsuitable for low-power and fully implanted optical neural interfaces. Alternatively, μLED arrays show great promise to create high-density and programmable illumination patterns. However, monolithic GaN [10], [11] are not CMOS compatible, and organic LEDs (OLED) [12], [13] do not yet provide sufficient brightness for fluorescent imaging or stimulation. As a result, we use flip-chip bonded arrays of commercial Cree μLEDs [14] in this work. Despite their relatively larger footprint (220μm×250μm) compared with monolithic GaN and OLEDs, they provide ultrahigh brightness at high efficiency (~30%) and relatively narrow linewidth (~20nm).

Our device incorporates two separate 5×5 arrays of blue (470nm) and green (535 nm) TR2227 Cree μLEDs with 1.08mm pitch for fluorescence excitation and optogenetic stimulation, respectively, as shown in Fig. 1. We will discuss the impacts of missing imager pixels due to μLED placements in Section III. The choice of these wavelengths depends on the target optogenetic and fluorescent probes [1]. In order to estimate the brightness at our target depth (~200μm), we modeled the light propagation of μLEDs in scattering tissue using Monte Carlo methods [15]. We used typical values of optical properties for the brain tissue (scattering and absorption coefficients of μs=10mm−1 and μa=0.01mm−1, respectively, scattering anisotropy of g=0.9, and refractive index of n=1.38 [16]). Fig. 2 shows the simulation results of a 2-by-2 subset of our μLED array. Green μLEDs with maximum radiance of 12mW result in the brightness of ~114mW/mm² at the brain surface. After propagation in 200μm brain tissue, the brightness drops to ~4mW/mm² due to the angular spreading, tissue scattering, and absorption (Fig. 2a). Given the minimum required light intensity of ~0.5mW/mm² required at the location of neuron for most optogenetic probes [17], our μLED array is sufficiently bright and dense to stimulate the entire FoV at the target depth. On the imaging side, blue μLEDs provide a maximum radiance of 30mW equivalent to a power density of ~285mW/mm². This results in a minimum fluorescence excitation light intensity of ~10mW/mm² at the target imaging depth of 200μm while the backscattered excitation light reaches ~10mW/mm² at the surface of device, as the simulation shows in Fig. 2b.

B. Optical Filters

We also need to integrate thin and flexible optical filters with sufficient optical density (OD) to reject the excitation light background over the fluorescent emission. Despite prior efforts to integrate interference filters [18] and plasmonic structures [19] with CMOS image sensors, filtering requirements are particularly challenging for our device due to the small Stokes shift of GECl/GEVI biomarkers (< 30nm) [20] compared with fluorophores in other applications (> 100nm) [21].

In order to estimate the signal-to-background ratio (SBR), we use the light propagation model shown in Fig. 2b. The
emission and 10mW/mm$^2$ interference and absorption filters. Moreover, it also enables a fluorescence conversion efficiency from blue to green fluorescence emission light (e.g., 518nm which is the peak emission of widely used GCaMP6f) is estimated to be $\sim0.17\%$ based on the Beer-Lambert law and GCaMP6f’s brightness ($3700\text{M}^{-1}\text{mm}^{-1}$ [20]) typical concentration ($10^4\text{M}$ [22]), and neuron’s diameter ($\sim20\mu\text{m}$). Assuming isotropic scattered emission and 10mW/mm$^2$ fluorescence excitation light intensity at the target imaging depth, we expect the received fluorescent emission light intensity of $\sim0.02\mu\text{W/mm}^2$ at the device, which results in the peak SBR of $0.2\times10^{-5}$. This normally necessitates using thick (>1mm) and rigid interference filters on glass with ODs of better than 6 in table-top imaging setups and miniscopes. However, this imposes serious challenges to make an ultrathin and mechanically flexible imager. More importantly, computational imaging, adopted in this work, relies on detecting wide-angle photons to provide large effective numerical aperture (NA). On one hand, interference filters are inherently angle-sensitive, providing peak performance at normal incidence. On the other hand, absorption filters are insensitive to angle, but typically cannot provide more than OD3 of rejection due to their more gradual cutoff slope and autofluorescence effects [23]. To achieve the best filter performance, we combine a wide-angle interference filter, an absorption filter (described in Section V), and time gating (TG) to achieve a total peak filter performance of better than OD5.

C. Image Sensor

Miniscopes and miniaturized cameras use commercial imagers with arrays of photos diodes (PDs) fabricated in CMOS imager (CIS) processes. Instead, we design and optimize a custom SPAD-based image sensor to enable TG fluorescence imaging through single-photon counting. In gated operation, the imager is only activated after each excitation pulse, as illustrated in Fig. 3a. Gating the sensor provides additional background rejection of scattered excitation light complementing the interference and absorption filters. Moreover, it also enables a fluorescence life-time imaging (FLIM) mode [24], [25].

While TG imaging can, in principle, provide arbitrarily large excitation rejections, the efficacy is limited in practice by the turn-off time of the excitation pulse (the more abrupt the turn-off, the better), the fluorescence life-time (longer lifetimes provide easier background rejection), and the impulse-response function (IRF) of the SPAD detectors (the faster the IRF, the better). Despite the advantage of having no read noise, SPAD-based imagers suffer from lower photon detection efficiencies (PDE) when compared to conventional PD imagers. Additionally, TG imagers typically consume larger power compared with PD or passive SPAD imagers due to the extra dynamic power required for gating the pixels. In order to contrast the overall performance of a SPAD-based TG imager with a conventional PD-based imager, we have modeled and calculated SNR [26], [27] for both imager types. The PD imager parameters (quantum efficiency (QE) of 0.6, well capacity of 52ke, dark current of 0.15fA, and read noise of 10e) are based on a commercial Aptina CMOS image sensor (MT9V021) [2], while we assumed a conservative dead-time of 25ns and dark-count-rate of 26cps for the SPADs. We look at two PDE possibilities for the SPAD imager, 1% and 10%, as determined by both fill factor (FF) and photon detection probability (PDP).

Both SPAD and PD imagers contain photon shot noise due to the background and fluorescence counts as well as dark-counts, while the PD imager has additional read noise. Shot noise limit curves in Fig. 3 indicate the theoretical SNR limit of imaging fluorescence with no background and the noise floor is entirely determined by photon shot noise.

At high intensities, saturation is determined by full well capacity in the PD imager and by the dead-time in the SPAD imager. The dynamic range of the PD imager is extended through multiple sampling in which the frame window is divided into $k$ samples at $1/2^k$ fractions of the starting integration time, such that well saturation is avoided for the shortest integration-time sample. The total frame time (sum of all integration time of the k samples) is set to $1/f$, where $f$ is the frame rate. The resulting SNR of the PD imager in the well-saturated regime is estimated using the multiple sampling scheme [27]. At our target frame-rate of $f=125$fps (as required for emerging GEVI reporters which can detect sub-ms action potentials [28]), Fig. 3 shows the SNR as a function of received fluorescence power for three imagers under two filtering scenarios: optical filtering of OD5 (Fig. 3b), and optical filtering of OD3 (Fig. 3c-d). Notice that imagers can become saturated due to the relatively large excitation background intensity. Overall, the TG SPAD-imager can achieve a higher SNR compare with a PD imager for low-OD (OD < 3) filtering, and the improvement is larger at higher TG excitation background rejections (TG of 1× and 25× are plotted in Fig. 3c-d). This results from the challenges PD imagers have in managing the background light. Notice that even increasing the dynamic-range of a PD imager cannot solve this issue as the excitation background and SNR will be limited by the optical filtering OD.
At OD 5, however, PD imagers outperform the SPAD-based imagers, mainly because of the lower PDE of SPADs. This can be rectified in the future with larger FF SPAD designs, demonstrated in part by the more comparable performance achieved at 10% PDE for the SPAD-based design in Fig. 3b.

Another critical design consideration is the total power consumption to minimize the heating in brain tissue. As explained in Section VI, our device consumes 45mW total power, estimated to keep tissue heating below ~1.5°C [29]. While it is within the range to enable continuous operation without heating concerns, further low-power optimization can lower the heating to even below the ideal range of 1°C.

III. COMPUTATIONAL IMAGING

Lens-based microscopes suffer from a fundamental trade-off between device size and performance; as lenses become smaller, they must either collect less light or image a smaller FoV [6]. This makes them unsuitable for implantable fluorescence imagers. In this work, we exploit computational imaging [5], [30] to enable a lens-less, compact and ultrathin device. In a lens-based system, the lens collects and focuses light from the scene pixels to the sensor pixels with a one-to-one mapping, but in a computational lens-less system, unfocused light from a single point on the scene is spatially modulated (by the mask) and mapped to multiple sensor pixels. Computational imaging masks can be realized with either a coded-aperture mask [6], phase mask [31] or an arbitrary diffuser mask [30]. While an amplitude binary mask with overall 50% transmission was used in this work, phase masks can also be used to achieve a higher SNR [31], [34].

A. Texas two-step (T2S) Model

The T2S model [6] requires a coded-aperture mask to spatially encode the unfocused light from the scene comprised of incoherent micron-scale fluorescent sources (e.g., neurons) as shown in Fig. 4. This approach greatly simplifies both the calibration and image reconstruction process by using a separable mask pattern composed of the outer product of two random binary vectors. Despite a simple mask design and fabrication steps (Section V), it improves the computational tractability of reconstruction over other methods using arbitrary diffuser masks [30]. While an amplitude binary mask with overall 50% transmission was used in this work, phase masks can also be used to achieve a higher SNR [31], [34].

B. Image Reconstruction

Figure 4 illustrates the T2S model assuming a single fluorescent bead as the scene. Due to the mask separability, local spatially varying point spread function can be written as the superposition of two independent terms: The first term models the effect of the coding (mask) and the second term models the effect of a hypothetical “open” mask (with no apertures). In order to image a 2D (planar) scene $X$, the raw capture at the imager ($Y$) can be expressed as:

$$ Y = Y_0 + Y_c = P_0 X Q_0^T + P_c X Q_c^T $$

(1)

where $P_0$ and $P_c$ operate only on the rows of $X$, and $Q_0$ and $Q_c$ operate only on the columns of $X$ (the subscripts o and c refer to “open” and “coding,” respectively). In order to image a 3D volumetric scene from a single shot, we can extend equation (1) to multiple depths as:

$$ Y = \sum_{d=1}^{D} P_{od} X_d Q_{od}^T + P_{cd} X_d Q_{cd}^T $$

(2)

where the subscript $d$ represents the depth assuming the entire target 3D volume is discretized over $D$ planar samples. By knowing the $P_{o,c}$ and $Q_{o,c}$ matrices, we can reconstruct the scene from any raw capture by solving a regularized least-squares reconstruction optimization problem (Fig. 5a). The one-time
computational imaging by modifying equation (1) as follows:

\[ Y = M \odot (P_1XQ_1^T + P_2XQ_2^T) \]  

(3)

where \( M \) is a 160×160 binary matrix with zeros where ever pixels are missing and \( \odot \) denotes the element-wise (Hadamard) product. Thus, we used the modified image reconstruction optimization (for a single depth) as shown in Fig. 5c.

IV. CHIP ARCHITECTURE

The all-optical neural interface chip was designed and implemented in a 0.13-μm high-voltage CMOS process. Fig. 6 depicts the chip architecture and imaging timing diagrams. Details of each major sub-block circuitry are described below.

Fig. 6. Block diagram of the all-optical neural interface chip with the timing diagrams of control signals for rolling-shutter imaging and stimulation included (all the control and enable signals are thermometer coded).

A. Imager Architecture and Optogenetics Backend

The imager operates in the photon-counting mode with a configurable, sliding time-gate to provide additional background rejection as described in Section II. In order to reduce power, area, and readout data-rate requirements, the imager employs a rolling-shutter which is combined with selective powering of the excitation μLEDs to reduce power and also help to reduce excitation background. The entire imager is comprised of a 5×5 array of macros, each consisting of 16 blocks in a 4×4 configuration. Of these blocks, 14 are 8×8 SPAD arrays, one is for the excitation μLED driver, and one is for the optogenetic μLED driver. During the selection of each sub-frame, which consist of a 5×1 macro column, only excitation μLEDs in the given column are activated. The SPAD array blocks are enabled column-wise, where each pixel remains active for 1024 excitation pulses of the μLED, and the detected photon counts are stored in shared-row 10-bit counters. Overall, the imager achieves a frame-rate of 125fps with a 40MHz reference clock. The data transmitter (Tx) block serializes the counter values and sends them off-chip to a control FPGA.

A. Imager Architecture and Optogenetics Backend

For optogenetics, stimulation signals are generated on-chip with tunable repetition-rates (5-40Hz) and pulse-widths (with 0.1ms LSB precision). The optogenetic μLED columns are also time-multiplexed to reduce the peak current and required on-chip decoupling capacitance. The stimulation pattern can be configured by enabling individual μLEDs, each one illuminating ~0.2mm² brain regions covering the entire FoV with 2D spatial selectivity as simulated in Fig. 2.

B. Micro-LED Drivers

There are two different types of μLED drivers for optogenetic and fluorescence excitation arrays. Due to slow optogenetic signals, the programmable control signals are generated in the chip periphery and then distributed across each column. Then, the μLED high-voltage drive head is implemented using a level-
shifter to apply up to 3.3V to the optogenetic μLEDs. Each high-swing μLED driver is also equipped with ~0.1nF of on-chip decoupling capacitance. The driver for the excitation μLEDs (shown in Fig. 7) is also based on the same high-voltage drive head with additional capabilities to generate sub-ns pulses locally by driving the μLED cathode to CVDD for turn-off. In doing so, we also push the μLED into reverse bias, reducing the turn-off time and, consequently, improving the effectiveness of time gating [36].

C. Active Quench and Clock-based Reset Circuit (AQC)

The SPADs are implemented with a 7.5-μm-diameter active area using a custom implant. Each pixel has a 30μm pitch with an AQC and 5% effective FF. The AQC circuit and its timing diagram is illustrated in Fig. 7. The excitation μLED driver generates ultrashort light pulses at every positive edge of the excitation clock (Exc. Col). Next, a delayed clock (EN) is used to activate the AQCs for photon detection. On the rising edge of EN, a level-shifter turns-off M1 and triggers a minimum duration one-shot reset pulse to the gate of M2. When ANODE is discharged to GND, the inverter-based comparator flips its output leaving the SPAD in a high-impedance Geiger-counting mode. Avalanche detection of a photon causes the comparator to flip, which quenches the avalanche current and triggers the buffered output to the row-shared 10b counter.

V. INTEGRATION AND PACKAGING

Micrographs of the chip and major sub-blocks are shown in Fig. 8. The entire chip area is 8×8mm². In this section, we elaborate on the fabrication and post-processing steps of this CMOS chip to make it mechanically flexible, while containing all the crucial components including optical filters and μLEDs.

A. The Imaging Mask and Optical Filters Fabrication

The binary amplitude mask is fabricated by patterning a 100-nm-thick chromium layer with 15μm feature sizes on a 100-μm-thick flexible absorption filter that is bonded onto the chip. The mask pattern and feature sizes are optimized for a mask-sensor separation and the imaging distance of 100μm and 200μm, respectively. Fig. 9 shows the cross-section of integrated optical filters and the imaging mask along with some of their micrographs. We employed both a long-pass absorption filter and a custom-designed wide-angle interference filter to reject excitation light scattered from the tissue. The interference filter consists of multiple thin-film dielectric layers with a total thickness of ~10μm, which is directly deposited on the CMOS chip. This filter is optimized to reject the 470nm excitation light relative to fluorescence emission at 520nm with OD > 3 for angles up to 45°. During the deposition process, bond pads are
protected by photoresist and tape, making them accessible after the filter integration for bonding μLEDs.

Unlike angle-sensitive interference filters, the absorption filter accepts wide-angle photons and can be used as the 100μm spacer required for the computational mask without increasing the overall thickness of the system. The gelatin-based filter in this work provides ~26dB (OD2.6) and it is also laser cut to create openings for the μLEDs.

B. Flexible Packaging Flow

Future implantable neural interfaces should be mechanically flexible to conform to brain curvatures [37]. This is particularly critical for large surface devices, covering large cortical areas. Here, we describe the packaging approach developed to make our device mechanically flexible, which is achieved by die thinning the CMOS chip down to below 20μm total thickness. It has been shown previously that die-thinned CMOS chips can be bent to sub-2-cm radii of curvature with gate delay and drain current changes of less than 7% [38], [39]. While flexible thin-film transistor platforms have been demonstrated recently for biomedical applications [40], [41], they lack the versatility and scalability of CMOS processes.

Fig. 10a shows the full packaging flow developed in this work. After depositing the interference filter on the CMOS chip, the die is flip-chip bonded onto a polyimide flexible PCB with thickness of ~50 μm with lead-free solder bump (Step 1). The PCB has a cut-out to provide a viewing window for the chip while maintaining sufficient overlap with chip boundary for underfill epoxy. Next, the μLEDs are pick-and-place flip-chip bonded using solder bumps (Step 2). In Step 3, the CMOS chip is temporary bonded on a glass carrier face-down for mechanical backside grinding and polishing via a micro-grinder tool. This in-house post-processing step results in a silicon substrate thickness of below 7.5μm with ~1μm standard deviation across entire chip area. After releasing the die-thinned chip from glass carrier, we epoxy a polyimide support substrate (~50μm thick) to the backside of the chip, epoxy the optical filter (with imaging mask) to the frontside of the chip, and finally parylene coat the entire device for bio-compatibility [42]. Fig. 10b shows the cross-section of final flexible package with <250μm thickness.

VI. EXPERIMENTAL RESULTS

Measurements and device characterizations are presented in this section. All the measurements are done with a 40MHz reference clock leading to the 125fps frame-rate. Total chip power consumption is 40mW with 3mW and 2mW budgeted for excitation and optogenetics (for 1ms pulse-width at a 20Hz rate setting) illumination, respectively.

A. Sub-block Characterizations

Characterization of the imager’s optical elements and time-gating is shown in Fig. 11. First, we characterized the spectral response of the SPAD sensor which achieves 12% photon-detection probability (PDP) at 520nm (target peak emission) with a median dark-count rate (DCR) of 26cps. The optical absorption filter provides ~400× (=26dB) rejection between 470nm and 520nm as discussed in Section V. The transmission of custom-designed interference filter at multiple angles is shown in Fig. 11b, which shows OD > 3 for angles below 45°. We have also characterized the nonlinear relationship between SPAD counts and photon flux due to photon pile-up [43], [44]. While this effect is negligible in our application since the imager will only operate in the low photon-flux regime, we linearize the sensor counts to enhance imaging quality and extend the dynamic-range for other applications.
for fluorescence excitation and optogenetic stimuli (providing ~0.5 mW/mm² because of area and power constraints).

Time-to-digital converters [46], which were not implemented for SPADs produces a longer-tail IRF to approaches based on achieving time-correlated single-photon counting (TCSPC) in practice. II. The μLED temporal characteristics of short excitation optical pulses are measured via the SPAD detectors on the chip itself (Fig. 12b). Estimated turn-off time is ~0.6 ns given that the TG response of the SPADs with an estimated full width at half values ~2.3 ns [45] due to the relatively long tail in the IRF. We can also extract the beads’ fluorescence lifetime (~3.6 ns), which is longer than previously reported values ~2.3 ns [45] due to the relatively long tail in the IRF response of the SPADs with an estimate full width at half maximum (FWHM) of ~1.1 ns. It is known that time-gating to achieve time-correlated single-photon counting (TCSPC) in SPADs produces a longer-tail IRF to approaches based on time-to-digital converters [46], which were not implemented because of area and power constraints.

Finally, Fig. 12 shows the measured optical power density of blue and green μLEDs and the estimated target operating points for fluorescence excitation and optogenetics (providing ~0.5 mW/mm² at 200 μm in brain tissue as explained in Section II). The μLED temporal characteristics of short excitation optical pulses are measured via the SPAD detectors on the chip itself (Fig. 12b). Estimated turn-off time is ~0.6 ns given that the TG imager’s IRF is convolved with optical pulses in this plot. Applying a 2 V reverse voltages (CVDD) using proposed μLED driver reduced the turn-off time by ~0.3 ns (Fig. 7).

B. Imager Calibration and Characterization

The experimental test setup for imaging and calibration (to estimate the separable transfer functions as explained in Section III) is shown in Fig. 13, where we used a commercial green (520 nm) LED array and pass the light through a 30-μm cross-hair line slit mounted to an 80° wide angle diffuser. This mimics an isotropic, fluorescent line source. The one-time calibration process is performed at the target image distances of 200 μm to 500 μm with a 50-μm step size. After calibration, the resolution is measured by imaging a double-line slit mask with 30 μm width and 60 μm spacing (Fig. 14). The cross-section of the reconstructed double-line slits proves that we can achieve better than 60-μm lateral resolution.

Next, a sparse sample of Fluoresbrite YG beads with 45 μm diameter on a glass cover slip is imaged with our device and a confocal microscope simultaneously. Fig. 15 shows the results under ~0.5 mW/mm² excitation intensity at the beads plane. Ground truth images are the overlap of confocal microscope image and actual bead locations. In the first experiment one sample is placed at a 200 μm distance from the mask. In the second experiment, we placed two planar samples at a 300 μm and 500 μm distance from the mask to verify the 3D imaging capability. Beads are fully imaged and reconstructed in both

<table>
<thead>
<tr>
<th>Reference</th>
<th>Papageorgio, et al. [32]</th>
<th>Lee, et al. [33]</th>
<th>This Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMOS Technology</td>
<td>0.18 μm</td>
<td>0.18 μm</td>
<td>0.13 μm</td>
</tr>
<tr>
<td>Chip Area (mm²)</td>
<td>4.7×2.25</td>
<td>2.64×3</td>
<td>8×8</td>
</tr>
<tr>
<td>Supply Voltage (V)</td>
<td>1.8</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Pixel Pitch (μm)</td>
<td>55</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Excitation Wavelength (nm)</td>
<td>450</td>
<td>385</td>
<td>470</td>
</tr>
<tr>
<td>Emission Wavelength (nm)</td>
<td>705</td>
<td>540</td>
<td>520</td>
</tr>
<tr>
<td>PDP (or QE)</td>
<td>5%</td>
<td>2.7%</td>
<td>12%</td>
</tr>
<tr>
<td>Fill-factor</td>
<td>28%</td>
<td>14.4%</td>
<td>5%</td>
</tr>
<tr>
<td>Resolution (μm)</td>
<td>220</td>
<td>1000</td>
<td>&lt; 60</td>
</tr>
<tr>
<td>FoV (°)</td>
<td>4.4×1.98</td>
<td>2.1×2.5</td>
<td>5.4×5.4</td>
</tr>
<tr>
<td>Frame-rate (fps)</td>
<td>20</td>
<td>N/R</td>
<td>125</td>
</tr>
<tr>
<td>Power (mW)</td>
<td>3.5</td>
<td>83.8</td>
<td>40</td>
</tr>
<tr>
<td>Array Size</td>
<td>80×36</td>
<td>72×60</td>
<td>160×160</td>
</tr>
<tr>
<td>Pixels Acceptance Angle</td>
<td>±18°</td>
<td>±10°</td>
<td>±70°</td>
</tr>
<tr>
<td>Read-out Data-rate (MB/s)</td>
<td>N/A</td>
<td>N/R</td>
<td>40</td>
</tr>
<tr>
<td>Integrated Light Source</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* Estimated from figures/data, ** Illumination power not included, + Not including off-chip ADCs

Fig. 14. Imager’s resolution characterization with vertical and horizontal double-line slits (insets show median intensity across the lines).

Fig. 15. Imaging sparse samples of fluorescent beads and cross-validation with confocal microscope images.

![Fig. 16. (a) Photo of the mechanically flexible device wrapped around a 10 mL beaker for testing, (b) functional μLEDs pulsing after bending, and (c) performance comparison of CMOS after die-thinning and bending.](image-url)
experiments with some minor artifacts. These artifacts are due to the insufficient SBR requiring further improvements to the optical filters and TG efficiency in future implementations.

The raw captures in Fig. 14 and Fig. 15 clearly show the effects of pixel “gaps” due to the placement of μLEDs on the imager. However, we can still reconstruct the scene as explained in Section III. While reconstructed image shows some variations (e.g., peak intensities in Fig. 14) depending on the portion of missing information from the raw capture, such artifacts can be reduced in future by minimizing gap sizes (e.g., by using smaller μLEDs). Compared to the double-line slit reconstruction (Fig. 14) which shows a few artifacts, the sparse sample in Fig. 15 is fully reconstructed without artifacts and more accurately mimics the sparsity of neural activity.

C. Mechanically Flexible Device

Finally, we wrapped the mechanically flexible device around a 10mL lab beaker with 1.25cm radius of curvature (Fig. 16) to show and evaluate the device flexibility. The flexible packaging allows for the chip and all the μLED and I/O bond pad connections to bend without breaking, as shown in Fig. 16, where the chip is functioning and lighting up the optogenetic μLEDs. In order to measure the impact of die-thinning and mechanical stress on electronics after bending, we measured the delay of a CMOS logic gate chain (with 16ns nominal delay), which is distributed across the chip curvature. The bar-chart shows the normalized CMOS logic speed (estimated from 1/total delay) at each state with maximum 6% changes in the curved state. Similarly, changes in analog block (including drivers, AQC, and clocking) performance were negligible. We expect that the proposed calibration and reconstruction approach can be extended to the bended imager case with a simple geometric mapping since radius of curvature is much larger than mask feature size.

VII. Conclusion

Measurement’s summary and comparison with prior fluorescence imagers are presented in Table. I. We have achieved the highest resolution and largest FoV with fully integrated light-source, thanks to the computational imaging technique adopted and modified in this work.

We have demonstrated a mechanically flexible, low-power, implantable, and lens-less device for all-optical neural stimulation and recording that achieves better than 60μm resolution over a large 5.4×5.4mm² FoV in this work. Our approach is the first step toward a fully implantable all-optical recording and stimulation brain interface device that can achieve better than cm² FoVs. Future generations will reduce pixel pitch and increase sensor FF by sharing SPAD’s guard-rings and simplifying in-pixel circuitry (e.g., by operating at standard VRST voltage-range which eliminates the level-shifter and thick-oxide devices). Moreover, background rejection can be further improved using semiconductor-based absorption filters and metal-dielectric interference filters. Additional power and data wireless capabilities will also simplify the flexible packaging and make these devices more suitable for in-vivo applications.

ACKNOWLEDGMENT

We gratefully acknowledge TSMC for chip fabrication and support for the use of SPADs. This work was supported in part by DARPA NESD program under Contract N66001-17-C-4012 and by the National Science Foundation under Grant 1706207.

REFERENCES


**Sajjad Moazeni** (M’13) received the B.S. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2013, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of California at Berkeley, Berkeley, CA, USA, in 2016 and 2018, respectively.

From 2018 to 2020, he was a Post-Doctoral Research Scientist in Bioelectronic Systems Lab at Columbia University, New York, NY, USA. He is currently an Assistant Professor of Electrical and Computer Engineering Department, at University of Washington, Seattle, WA, USA. His research interests are designing integrated systems using emerging technologies, integrated photonics, neuro and bio photonics, and analog/mixed-signal integrated circuits.

**Eric H. Pollmann** (S’19) received the B.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, in 2017 and the M.S. degree in electrical engineering from Columbia University, New York, NY, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering at Columbia University, New York, NY. His research interests are in implantable CMOS fluorescence imagers for applications in biology and neuroscience.

**Vivek Boominathan** (M’19) received the bachelor’s degree in electrical engineering from the Indian Institute of Technology Hyderabad, India, in 2012, and the M.S. and Ph.D. degrees in electrical and computer engineering from Rice University, Houston, TX, USA, in 2016 and 2019, respectively. He is currently a Postdoctoral Associate at Rice University, Houston, TX, USA. His research interest lied in the areas of computer vision, computational imaging, wave optics, and machine learning.

**Filipe A. Cardoso** (M’11) has studied Physics Engineering at Instituto Superior Técnico (IST). During his PhD at IST and INESC-MN, he has developed magnetoresistive biosensors to detect magnetically labeled molecules. After his PhD, Filipe has continued the development of magnetoresistive sensor for both biosensors and non-destructive testing. Filipe has further co-founded a startup company, Magnomics, aiming to push magnetoresistive biosensor technology into market, in particular animal health market. Filipe served as CTO of the company. In 2018, he has joint Prof. Ken Shepard group at Columbia University (New York, USA) as a Research Scientist where he is currently developing powering, communication and biosensing technologies for implantable devices.

**Jacob T. Robinson, Jr.** is an Associate Professor in Electrical & Computer Engineering and Bioengineering at Rice University, and an Adjunct Associate Professor in Neuroscience at Baylor College of Medicine. His research group uses nanofabrication technology to create devices that can manipulate and monitor neural circuit activity. He received a B.S. in Physics from UCLA in 2003 and a Ph.D. in Applied Physics from Cornell University in 2008. After completing his Ph.D. studying silicon nanophotonics he began postdoctoral research in the Department of Chemistry and Chemical Biology at Harvard University. In 2012, he joined the ECE and BioE departments at Rice where he currently develops technologies...
to better stimulate and record neural activity. Dr. Robinson is the recipient of the DARPA Young Faculty Award, the Materials Today Rising Star Award, and is a Senior Member of IEEE. He has served as co-chair of the IEEE Brain Initiative and is currently a core member of the IEEE Brain Neuroethics working group.

Ashok Veeraraghavan received the bachelor’s degree in electrical engineering from the Indian Institute of Technology, Madras, Chennai, India, in 2002 and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA, in 2004 and 2008, respectively. He is currently an Associate Professor of Electrical and Computer Engineering, Rice University, Houston, TX, USA. Before joining Rice University, he spent three years as a Research Scientist at Mitsubishi Electric Research Labs, Cambridge, MA, USA. His research interests are broadly in the areas of computational imaging, computer vision, machine learning, and robotics. Dr. Veeraraghavan’s thesis received the Doctoral Dissertation Award from the Department of Electrical and Computer Engineering at the University of Maryland. He is the recipient of the National Science Foundation CAREER Award in 2017. At Rice University, he directs the Computational Imaging and Vision Lab.

Kenneth L. Shepard (M’91–SM’03–F’08) (Fellow, IEEE) 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 a 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 CMOS bioelectronics and power electronics. Dr. Shepard has been an Associate Editor for 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.