

Time multiplexed active neural probe with 678 parallel recording sites

Bogdan C. Raducanu^{1,2}, Refet F. Yazicioglu¹, Carolina M. Lopez¹, Marco Ballini¹, Jan Putzeys¹, Shiwei Wang¹, Alexandru Andrei¹, Marleen Welkenhuysen¹, Nick van Helleputte¹, Silke Musa¹, Robert Puers², Fabian Kloosterman^{1,2,3}, Chris van Hoof^{1,2}, Srinjoy Mitra¹

¹Imec, Heverlee, Belgium, ²KU Leuven, Heverlee, Belgium, ³NERF, Leuven, Belgium

Abstract— We present a high density CMOS neural probe with active electrodes (pixels), consisting of dedicated in-situ circuits for signal source amplification. The complete probe contains 1356 neuron size ($20 \times 20 \mu\text{m}^2$) pixels densely packed on a $50 \mu\text{m}$ thick, $100 \mu\text{m}$ wide and 8mm long shank. It allows simultaneous high-performance recording from 678 electrodes and a possibility to simultaneously observe all of the 1356 electrodes with increased noise. This considerably surpasses the state of the art active neural probes in electrode count and flexibility. The measured action potential band noise is $12.4 \mu\text{V}_{\text{rms}}$, with just $3 \mu\text{W}$ power dissipation per electrode amplifier and $45 \mu\text{W}$ per channel (including data transmission).

Keywords—active neural probes; CMOS; high density component;

I. INTRODUCTION

Due to the need for large-scale recording from individual neurons in multiple brain areas, both high density and high number of electrodes are necessary in neural probes [1]. In order to minimize tissue damage, the ‘shank’ (the implanted part of the probe) has to be as narrow and thin as possible. Various silicon neural probes that have been recently developed, consist of a large number of tiny active electrodes that can locally amplify/buffer the neural signals [1]. However, with such limited space, the CMOS pixel amplifiers (PA) underneath the electrodes are restricted to a bare minimum while most of the signal processing is done in the ‘base’ of the probe. All prior active and passive neural probes [2][3][8] used a dedicated metal line per electrode to send the signal to the base circuitry. Naturally, this limits the number of simultaneous recording electrodes to the number of metal lines fitted in the cross section of the shank (Fig. 1). To overcome this fundamental bottleneck and achieve a denser simultaneous readout, a new architecture is proposed which relies on time division multiplexing and techniques to reduce the associated noise.

II. OPERATION PRINCIPLE

Active neural probes improve recording quality by buffering or amplifying the input signal close to its source (i.e. the electrode). This approach reduces the source impedance and minimizes crosstalk caused by the coupling amongst the long neighboring shank wires [2]. The PA has strict design constraints: the area is limited by the electrode size, the power is limited by the acceptable tissue heating and the noise requirements are imposed by the signal amplitude (as small as tens of μV). Within the limited area, an obvious method to reduce noise is to

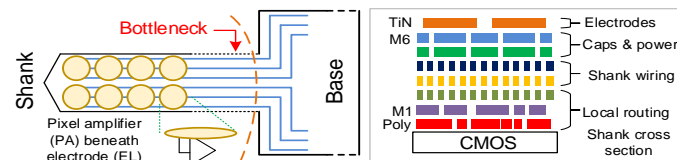


Fig. 1 Neural probe with shank wiring bottleneck and typical cross section, with 6 metal layers.

increase the current consumption of the PA input transistor. This results in the PA having a high bandwidth. Since the neural signal band itself is limited to $\sim 7.5 \text{kHz}$, the PA output can be sampled at $f_s > 15 \text{kHz}$ in the base. Therefore a simple time division multiplexing could be embedded within the shank (Fig. 2a) allowing M number of PA outputs on a single shank wire ($f_{\text{MUX}} = Mf_s$). However, the lack of a traditional anti-aliasing filter limiting the high PA bandwidth increases the in-band noise due to folding. Since it is not possible to fit low pass filters within the limited area of the PA (before the sampling operation), we have employed an alternative method of noise reduction by integrating the signal over a period of time (T_i) (Fig. 2b). The integrate, sample and reset operations strongly attenuate the signal beyond $f_i = 1/T_i$ ($f_i \geq f_{\text{MUX}}$), improving the signal-to-noise ratio.

For this particular design, a multiplexing factor of $M=8$ was enough to overcome the shank-wire bottleneck. To avoid in-band distortion, each channel is oversampled at $f_s = 40 \text{kHz}$ ($> 15 \text{kHz}$), producing a total multiplexing frequency of 320kHz . This, in turn, limits the integration period to a maximum of $3.125 \mu\text{s}$. We have used $T_i = 2.5 \mu\text{s}$ to allow time

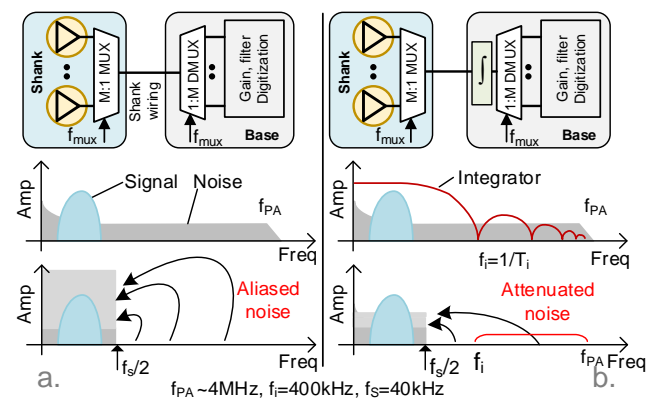


Fig. 2 a: Consequences of multiplexing without filtering. b: filtering signal by integration reduces out-of-band noise.

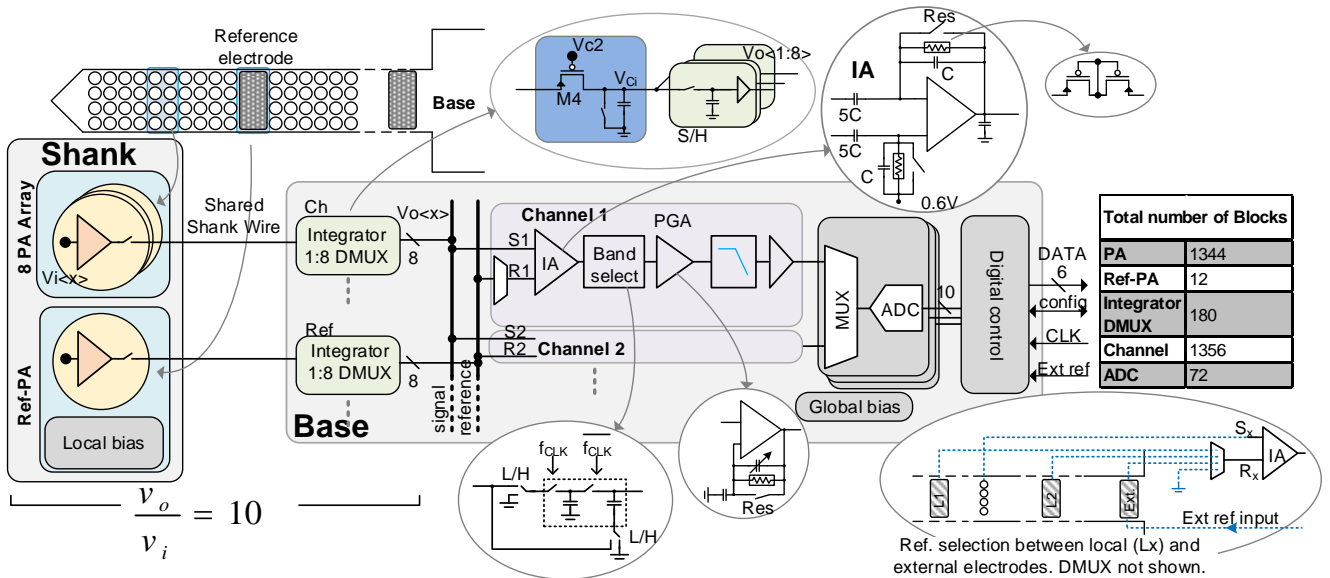


Fig. 3 Architecture of probe showing a pseudo differential signal path from end to end, including internal details and number of block copies on the right.

for transitions. This process essentially results in a low pass operation, strongly reducing the PA bandwidth from ~ 4 MHz to 400 kHz and limiting of the noise folding.

III. CIRCUIT DESCRIPTION

A. Architecture

Fig. 3 shows the block level architecture of the complete probe. The output from an array of 8 multiplexed pixel amplifiers is sent to the base through a shared shank wire. The signal is fed to an integrator (in the base) whose output is demultiplexed (DMUX block) using 8 sample-and-hold circuits ($V_o < 1:8 >$). Each V_o then goes to its corresponding channel block where the signal is further amplified and filtered, keeping only the band of interest. The outputs of 20 channels are multiplexed and digitized with the help of a 10-bit successive approximation register (SAR) analog to digital converter (ADC). The digital control block is responsible for generating the clocks for the ADCs and the MUX/DMUX blocks. It also serializes the parallel data from all the ADCs to only 6 data lines. With the aid of daisy-chained shift registers, all the channels, PAs and bias parameters are configured. The chip contains 1344 small electrodes ($20 \times 20 \mu\text{m}^2$) and 12 larger reference electrodes ($40 \times 80 \mu\text{m}^2$). A 13th Ref-PA block is used to amplify the external reference and is placed at the neck of the shank. A total of 180 Integrator-DMUX blocks drive the 1440 channels which are digitized by 72 ADCs. The extra channels are provided for the external reference and testing purposes. The global bias block contains a band-gap reference and other circuits to generate all the required voltages and currents for the chip.

B. Pixel amplifier

The integrator architecture described in Section II is split in two parts. Within the limited area of the pixel, the PA acts as a voltage to current converter (Fig. 4). The current is then integrated for a fixed time ($T_i = 2.5 \mu\text{s}$) over a capacitor ($C_i = 15 \text{ pF}$) in the base shared among 8 channels. After T_i , the voltage on C_i is sampled and then it is discharged for the next cycle. The integration capacitor, and sample and hold (S/H)

circuits forming the de-multiplexer are located in the less area restricted base. A flipped voltage follower buffer following the S/H circuit allows the Ref DMUX to drive multiple channels.

The PA employs an open loop, AC coupled, transconductance (gm) stage ($M1$). This produces an overall small signal gain of 10, given by:

$$A = \frac{v_o}{v_i} = gm \frac{T_i}{C_i} \quad (1)$$

The cascode transistor $M2$ reduces the clock feedthrough from the switches A and B to the gm stage. These switches are overlapping in order to ensure an always ON current through $M1$. These aspects are of crucial importance to maintaining stability, as the gate (G) of $M1$ is a high impedance node. The pseudo-resistor $M3$ and C_i set the high pass corner of the PA ($< 1 \text{ Hz}$), necessary to reject the relatively high DC level (hundreds of mV) produced by the electrode-tissue interface. During normal operation, the cascode transistor $M4$ located between the current source (i.e. the PA) and the integrating capacitor (C_i) ensures that the shank wire connected at the source of $M4$ is at a constant voltage equal to the supply rail ($V_s \sim 1.2 \text{ V}$). By keeping all shank wires at a constant voltage,

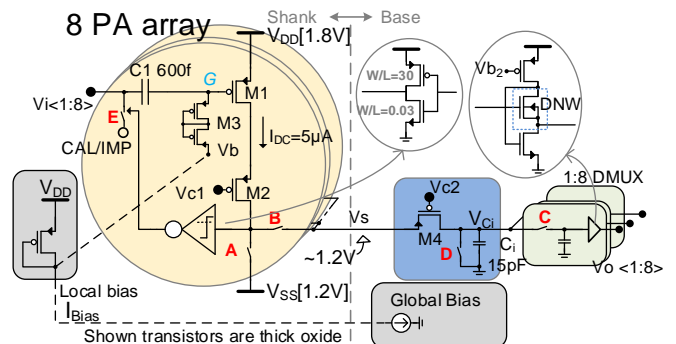


Fig. 4 PA architecture: $M1$ works as gm stage. The cascode transistor $M2$ isolates $M1$ from the clock feedthrough at the output and overlapped A/B switches enable $M1$ to always have an ON current. Both these methods ensure the stability of high impedance node G . The S/H circuit uses flipped voltage follower buffer with a deep N-well NMOS.

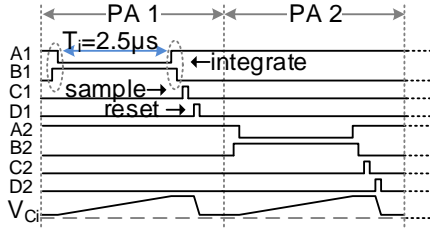


Fig. 5 Detailed timing diagram of 2 PAs indicating the corresponding changes in voltage V_{Ci} . The overlaps between A & B are marked.

this approach reduces the crosstalk amongst channels caused by capacitive coupling of the long shank lines.

The supply rails of the PA (1.8/1.2 V) are defined by multiple factors. The power budget ($I_{DC}(V_{DD}-V_{SS})$) coupled with minimal noise requirement, induces a trade-off between the current through $M1$ and its V_{DS} . However, the chosen operating point must account for the drop in the power supply lines across the shank, thus 0.6 V supply voltage is optimal ($V_{DD}-V_{SS}$). By using 1.2 and 1.8 V rails, the current can be directly integrated over C_i (within the range of 0 to 1.2 V), eliminating the need for a negative supply. Furthermore, since the 1.2 V rail is used by the following stages, the current from the unselected pixels (switch A closed) can be reused to power the blocks in the base, reducing overall consumption. The power dissipation limit of the PA is given by the physiologically safe limit of $< 1^\circ\text{C}$ tissue heating, including the dissipation in the power lines.

Additionally, through switch E , the PA allows for gain calibration (CAL) and electrode impedance characterization (IMP). Applying a known voltage (via the CAL/IMP port) while the electrode is floating (not connected to sample/solution) allows for measurement and calibration of the end to end gain. Similarly, applying a known current while the probe is submerged in saline solution allows for the characterization of the electrode-tissue interface impedance. Since this measurement requires the connection of a single PA input to the shared signal injection CAL/IMP port, the selection is of that particular switch E is done by temporarily lowering the cascode voltage V_{c2} when its switch B is ON. This triggers a high-threshold inverter within that specific PA, thus setting the switch E , while allowing normal operation of the PA. This method of using the output line simultaneously as a select signal eliminates the need for any control registers within the PA. The transistors shown in Fig. 4 are thick-oxide transistors in order to reduce gate leakage, except for the high-threshold inverter.

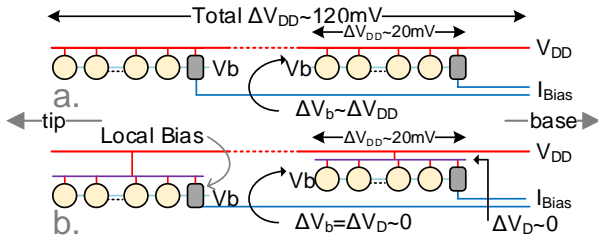


Fig. 6 a: Large supply drop changes the bias voltage, ΔV_b , due to high current in the supply rail consumed by all PAs. b: a tree-like power supply ensures that ΔV_b stays constant within a "branch", due to the lower current in the local rail. Each branch contains its dedicated local bias generator.

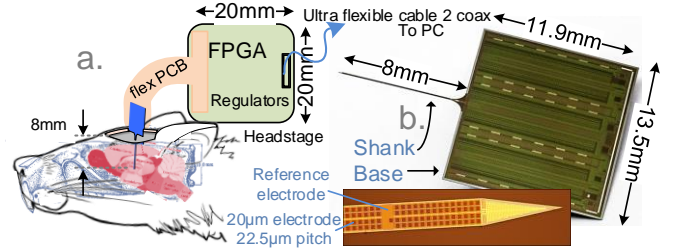


Fig. 7 a: Envisioned system diagram. A small headstage insures communication and power through an ultra-thin cable. b: chip micro photographs.

The extremely high aspect ratio of the shank along with the limited area for supply routing (due to the large number of signal wires) results in a high voltage drop of $\sim 120\text{ mV}$ across the shank power supply (Fig. 6 a). Even after locally generating the gate bias V_b , this voltage drop creates enormous differences among the PA bias voltages ($\Delta V_b \sim \Delta V_{DD}$) and severely affects operation performance. To mitigate this problem, a tree structure for the supply line is implemented by splitting the shank in 12 branches (113 PA each). Here, each branch experiences a much lower drop ($\Delta V_b \sim 0\text{ V}$) resulting in a more controlled bias current.

In order to maintain proper operation under these voltage drops while also accounting for supply ripple on the high-resistive lines, only 6 (random) branches can be turned ON simultaneously without additional penalty on noise. This allows recording from 6 arbitrary sections on the 8 mm shank ($\sim 0.7\text{ mm}$ each) with the reported performance, suitable for covering multiple regions on a rat brain. Moreover, the design also allows simultaneous complete readout from all electrodes on the shank (1356) with increased noise and power.

C. Channel

Each channel receives a signal (S_x) and reference (R_x) line, from the corresponding DMUX (Fig. 3) that feeds the instrumentation amplifier (IA). The reference (REF) line can be selected from i) one of the local Reference PA (Ref-PA), ii) few locally averaged Ref-PAs or iii) an external signal. Furthermore, single ended operation is possible, which along with the readout of the reference channels, enables software referencing. This may result in an improved signal quality [9]. In order to preserve circuit symmetry and avoid distortions, each Ref-PA is de-multiplexed to 8 outputs, such that for each channel the 2 inputs of the IA are de-multiplexed (i.e. sampled) simultaneously.

By providing a gain of 10, the integrator also relaxes the noise budget of the IA. The IA is implemented using an AC-coupled folded-cascode OTA with the bandwidth limited to approximately 15 kHz. This prevents aliasing from the subsequent switched capacitor (SC) band-select filter. The SC filter (operating at 80 kHz) can be configured as high pass, low pass or disabled to select the action potential (AP: 300 Hz – 7.5 kHz), the local field potential (LFP: 1 Hz – 1 kHz) or the full band (1 Hz – 7.5 kHz), respectively. A programmable gain

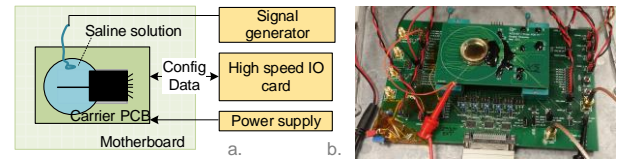


Fig. 8 a: Test setup schematic. b: test setup picture.

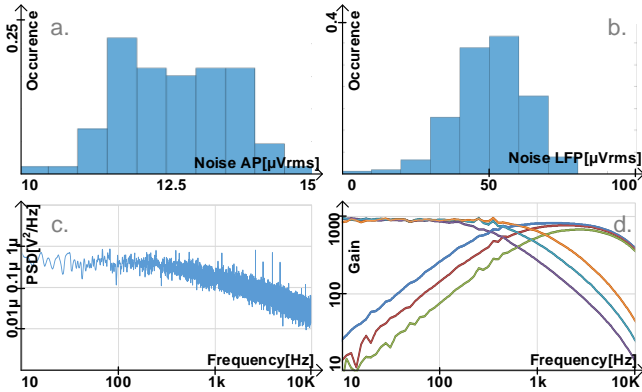


Fig. 9 Measurement results. a: distribution of noise in AP and b: LFP band. c: noise density in AP band, d: filter configuration for a fixed gain of 1000.

amplifier (PGA) follows the filter and provides 8 configurable gains between 1 and 50. After the PGA, the signal passes through an anti-aliasing filter and is buffered in order to be multiplexed and drive the ADC. A class-AB ADC driver is used to reduce the static power consumption. Each channel allows for independent band selection, gain configuration, reference selection, calibration selection and power down through a chain of shift registers distributed across the chip.

IV. TEST RESULTS

The chip was fabricated using a 6-metal layers 0.13 μm Al CMOS process, followed by biocompatible TiN electrode deposition and wafer thinning down to 50 μm (Fig. 7). Measurements were performed in a dark Faraday cage, using phosphate buffered saline solution. The total power consumption is 31 mW for 678 channels, with 2.3 mW dissipated in the shank (3 $\mu\text{W}/\text{PA}$) and 28.7 mW in the base, including data transmission with 4 pF loading. For the optimal (high-performance) setting of 6 random electrode branches ($113 \times 6 = 678$), the input referred noise is $12.4 \pm 0.9 \mu\text{Vrms}$ in the AP band and $50.2 \pm 12 \mu\text{Vrms}$ in the LFP band (Fig. 9). The entire probe (12 groups, 1356 electrodes) can be simultaneously turned on for low fidelity scanning purposes in order to select the region of interest. The crosstalk across the full signal chain is -63 dB at 1 kHz, with the measurement being limited by the noise floor. Fig. 10 shows a pre-recorded neural signal that has been fed into the saline solution and captured by the probe along with the separation within the two signal bands.

This work demonstrates at least 2 times increase in the number of simultaneous recording channels with respect to the state of the art [8] active neural probes while having comparable power and noise performance. TABLE I. compares this work with prominent passive and active neural probes.

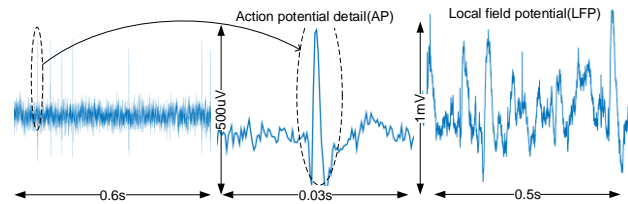


Fig. 10 Pre-recorded neuronal signals played back and captured by the probe showing action potential (left) and local field potentials (right).

V. ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's 7th Framework Programme (FP7/2007-2013) under grant agreement n°600925, NeuroSeeker.

TABLE I. COMPARISON TABLE WITH PRIOR ART

Parameter	Measured Values							
	[3]	[4]	[6]	[5]	[7]	[2]	[8]	This work
Probe Shank								
No. Electrodes	8	64	--	334	--	455	966	1356
Electrode Pitch [μm]	100	24	--	30	--	35	20	22.5
CSAC [μm^2]	127.5	30.55	--	11.98	--	10.99	3.65	3.7
Total Power/El [μW]	--	--	--	--	--	3.6	4.7	3
Crosstalk [dB]	--	-84	--	--	--	-44.8	-64	-63
Probe Base (Recording System)								
No. recording channels with specified performance	8	64	100	16	96	52	384	768
No. recording channels for intermittent observation	N/A	N/A	N/A	N/A	N/A	N/A	N/A	1356
Gain	1000	194	400/600	--		30-4000	50-2500	50-2500
HP Corner [Hz]	300	1.3	0.25	--	300	0.5/200/300/500	0.5/300/500/1000	0.5/300/500/1000
LP Corner [Hz]	10000	6400	2500-10000	--	10000	200/6000	1000/10000	300/500/1000/8000
ADC Resolution [b]	5	--	9	--	10	10	10	10
Sampling Rate (kS/s)	160 (8 Ch)	--	200 (10 Ch)	--	31/Ch	120 (4 Ch)	390 (13 Ch)	400 (20 Ch)
Full probe								
Total Power/Ch [μW]	94.5	351.6	0.94**	--	67	27.84	49	45
Total Area/Ch [mm^2]	0.625	0.45	0.25	--	0.26	0.19	0.12	0.12
Input Noise (μV)	9.2	2	3.2	--	2.2	3.2	6.36	12.4
** IO digital power not included in this number.								

REFERENCES

- [1] G. Buzsáki, E. Stark, A. Berényi, D. Khodagholy, D. R. Kipke, E. Yoon, K. D. Wise "Tools for Probing Local Circuits: High-Density Silicon Probes Combined with Optogenetics" *Neuron*, 2015.
- [2] C. M. Lopez, A. Andrei, S. Mitra, M. Welkenhuysen, W. Eberle, C. Bartic, R. Puers, R. F. Yazicioglu, G. Gielen, "An Implantable 455-Active-Electrode 52-Channel CMOS Neural Probe" *JSSC*, 2014.
- [3] R. Olsson, K. Wise "A Three-Dimensional Neural Recording Microsystem with Implantable Data Compression Circuitry" , R.H. Olsson and K.D. Wise, *JSSC*, 2005.
- [4] J. Du, T. J. Blanche, R. R. Harrison, H. A. Lester, S. C. Masmanidis, "Multiplexed, High Density Electrophysiology with Nanofabricated Neural Probes", *PLoS One*, 2011.
- [5] A. Sayed Herbawi, F. Larramendy, T. Galchev, T. Holzhammer, B. Mildnerberger, O. Paul, and P. Ruther, "CMOS-based neural probe with enhanced electronic depth control" *Transducers*, 2015.
- [6] D. Han, Y. Zheng, R. Rajkumar, G. S. Dawe, M. Je, "A 0.45 V 100-Channel Neural-Recording IC With Sub-/Channel Consumption in 0.18 CMOS", *TBCAS*, 2013.
- [7] R. M. Walker, H. Gao, P. Nuyujukian, K. Makinwa, K. V. Shenoy, T. Meng, B. Murmann, "A 96-Channel Full Data Rate Direct Neural Interface in 0.13 μm CMOS", *VLSI*, 2011.
- [8] C. M. Lopez, S. Mitra, J. Putzeys, B. Raducanu, M. Ballini, A. Andrei, S. Severi, M. Welkenhuysen, C. Van Hoof, S. Musa, R. F. Yazicioglu, "A 966-Electrode Neural Probe with 384 Configurable Channels in 0.13 μm SOI CMOS", *ISSCC* 2016
- [9] K. A. Ludwig, R. M. Miriani, N. B. Langhals, M. D. Joseph, D. J. Anderson, D. R. Kipke, "Using a Common Average Reference to Improve Cortical Neuron Recordings From Microelectrode Arrays", *J Neurophysiol.* 2009