# 256-Channel Neural Recording and Delta Compression Microsystem With 3D Electrodes

Joseph N. Y. Aziz, *Member, IEEE*, Karim Abdelhalim, Ruslana Shulyzki, Roman Genov, *Member, IEEE*, Berj L. Bardakjian, *Member, IEEE*, Miron Derchansky, Demitre Serletis, and Peter L. Carlen

***Abstract*—A 3D microsystem for multi-site penetrating extracellular neural recording from the brain is presented. A $16 \times 16$-channel neural recording interface integrated prototype fabricated in 0.35 $\mu$m CMOS occupies 3.5 mm $\times$ 4.5 mm area. Each recording channel dissipates 15 $\mu$W of power with input-referred noise of 7 $\mu V_{rms}$ over 5 kHz bandwidth. A switched-capacitor delta read-out data compression circuit trades recording accuracy for the output data rate. An array of 1.5 mm platinum-coated microelectrodes is bonded directly onto the die. Results of *in vitro* experimental recordings from intact mouse hippocampus validate the circuit design and the on-chip electrode bonding technology.**

***Index Terms*—Multi-channel recording, microelectrodes, extracellular recording, electrode array, implantable, brain, hippocampus, delta compression, neural amplifier.**

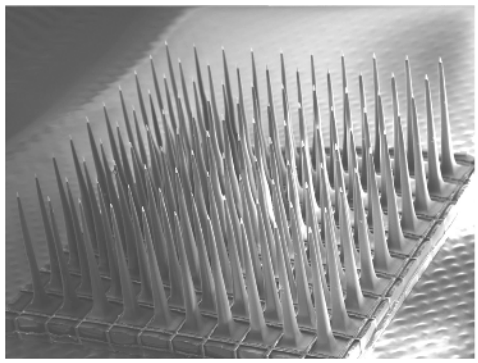

Fig. 1. Utah electrode array (UEA).

## I. Introduction

THE human brain is a collection of billions of interconnected neurons. Each neuron employs electrochemical reactions to receive, process and transmit information. Monitoring bio-electrical signals in the brain helps understand the nature of various neurophysiological behaviors.

Electrical recording from the brain is commonly performed by one of several electrophysiological methods. Intracellular recording and patch-clamp recording yield very accurate information but require precise placement of an electrode or an electrically conductive glass pipette to penetrate or abut the neuron cell membrane. They are typically limited to recording from a single neuron. Extracellular recording employs an electrode placed in the extracellular space near one or several neurons without penetrating a neuron cell body. The electrode captures the cumulative electrical activity of the surrounding neurons. As precise placement of the electrode is not required, extracellular recording is well suited for multi-site recording of neural activity using microelectrode arrays such as the Utah electrode array depicted in Fig. 1. Multi-site neural recordings capture and utilize the heterogeneity across biological ensembles in the brain.

The amplitude of extracellular neural activity typically ranges between several microvolts and a few millivolts, with most of the signal power concentrated in the frequency band between sub-hertz and a few kilohertz. Recording of neural activity has been traditionally performed using high-accuracy bench-top biomedical instrumentation equipment. These instruments are generally stationary, bulky, limited to one or a few acquisition channels, and prone to excessive interference noise due to electrode wiring. Integrated neural recording microsystems, fabricated on a single semiconductor substrate, lack these drawbacks. They offer a small, low-power, low-noise, and cost-effective chronically implantable alternative to commercial bench-top instruments [1]–[9].

The recording channel in an integrated neural interface typically performs signal acquisition, band-pass filtering, amplification, and in some instances quantization. An efficient design of the neural amplifier has been a focus of much research. As neural signals typically have weak amplitudes, low-noise operation is of critical importance. When a neural recording microsystem is implanted in the brain, a tight power density budget is imposed in order to prevent damage to the surrounding tissue due to excessive dissipated heat [10]. Channel integration area is often limited by form-factor considerations such as the recording site pitch as well as the technology cost and yield. Low-noise performance of a neural amplifier typically comes at the cost of spending power and integration area.

The conflicting trade-offs among noise, power dissipation and integration area of a neural amplifier impose a set of stringent

Manuscript received April 03, 2007; revised September 08, 2008. Current version published February 25, 2009. This work was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and Krembil Fund.

J. N. Y. Aziz was with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. He is now with Broadcom, Irvine, CA 92617 USA.

K. Abdelhalim, R. Shulyzki, and R. Genov are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: roman@eecg.utoronto.ca).

B. L. Bardakjian is with the Institute of Biomaterials and Biomedical Engineering, Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A4, Canada.

M. Derchansky, D. Serletis, and P. L. Carlen are with the Toronto Western Research Institute and the Department of Physiology, University of Toronto, Toronto, Canada.

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

Digital Object Identifier 10.1109/JSSC.2008.2010997

design constraints. A number of CMOS transconductance amplifier topologies have been utilized and their designs optimized to satisfy these constraints. The design in [1] employs a current-mirror transconductance amplifier with some transistors biased in weak inversion region. The neural amplifier in [2] is comprised of a five-transistor differential amplifier with a second common-source gain stage. The neural recording interface in [3] employs a folded-cascode amplifier topology. Although a variety of amplifier topologies have been investigated, their noise is typically reduced at the cost of increased power and integration area. We employ a transconductance amplifier with the telescopic topology in the first stage of a recording channel. It has fewer transistors which generate noise and require power for their biasing. The amplifier dissipates less power and occupies less area than the topologies in [1], [2], and [3] and maintains an acceptable RMS noise level of 7 $\mu$V over 5 kHz bandwidth.

As the bandwidth of neural signals can reach 5 kHz, neural interfaces recording on many channels yield large amount of data which can not be efficiently stored, processed or transmitted. Spike detection has been a data reduction method of choice in a number of neural recording microsystem designs [4], [5]. Conventional spike detection algorithms perform thresholding of the amplitude of the recorded signal in order to capture the majority of action potentials, but discard other important neurophysiological information such as neural activity below or above a threshold level, or activity in an abnormal neurological state. We perform on-chip analog delta compression of neural data. Temporal derivative of a neural signal is computed and its value transmitted only if it rises above a certain threshold. The choice of the threshold sets a flexible trade-off between recording accuracy and output data rate. The output data rate is proportional to the mere information rate of the neural data but not to the dimensions of the electrode array or the sampling rate.

In applications such as neural prostheses, animal studies and high throughput drug screening, recordings from within a volume of neural tissue are required. Spatial proximity of multiple recording sites and interference noise considerations suggest fabrication of multiple electrodes directly on or near the die. Recording microsystems with three-dimensional (3D) electrode arrays of various configurations fabricated near the die have been reported such as with electrodes co-planar with the die [4], [6]. Implementations with 3D electrode arrays bonded directly to the surface of the chip have been reported for 16 recording sites [7] and proposed for 100 recording sites [5]. Implementations with a higher number of channels have been reported without electrodes and at the cost of increased circuit noise [8]. We previously reported a 256-channel microsystem with 100-$\mu$m on-chip electrodes for recording of neural activity from acute brain slices of mice [9], [11]. We now present a neural recording microsystem with on-chip array of longer electrodes for volume recording from brain tissue. The electrode array is attached directly to the surface of the die by means of a low-cost bonding method.

This paper presents a 256-channel neural recording and delta compression microsystem implementation first reported in [12], and addresses in detail the design challenges and corresponding solutions identified above. The rest of the paper is organized as follows. Section II details the design and VLSI implementation of the low-power densely-integrated neural amplifier and other circuits in the recording channel. Section III describes the peripheral circuits performing on-chip delta compression of the recorded neural data. Section IV presents the hybrid microsystem comprised of a neural recording interface prototype integrated with an array of on-chip microelectrodes.

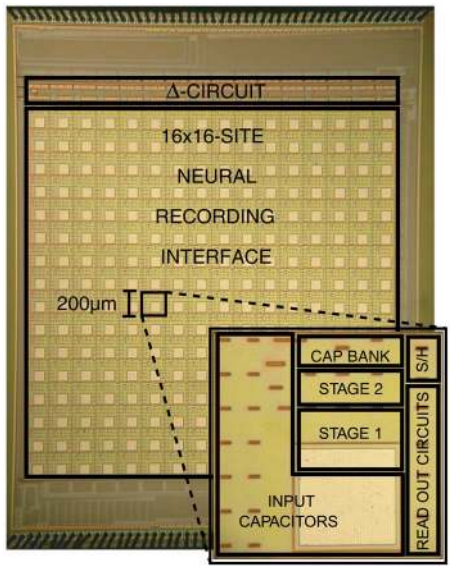


Fig. 2. Micrograph of the 256-channel integrated neural interface. The $3.5 \times 4.5$ mm$^2$ die was fabricated in a 0.35 $\mu$m CMOS technology.

## II. Recording Channel

### A. Channel Architecture

The microsystem records neural activity simultaneously on 256 channels from a volume of brain tissue. The signal acquisition circuits are arranged in a $16 \times 16$ array as shown in the micrograph of one of the prototypes in Fig. 2. As recording electrodes are bonded directly onto the surface of the die, the 200-$\mu$m cell pitch is dictated by spatial resolution requirements of extracellular recording in the hippocampus of mice. Electrodes are interfaced to the recording channels through non-passivated electrode bonding pads, each $80 \times 80$ $\mu$m$^2$ in size.

To preserve channel-to-channel temporal correlation of recorded neural signals, each channel is implemented as an independent voltage amplifier with in-channel memory as depicted in Fig. 3. Continuous-time implementation of the amplification stages avoids input signal contamination by digital signals. As a high closed-loop gain is required, the channel employs two stages of amplification. This maintains capacitor sizes within the recording cell pitch requirement.

The first stage of the channel is a low-noise difference amplifier. Electrochemical effects occurring at the electrode-tissue interface cause a random DC voltage offset that is several orders of magnitude larger than the neural signal amplitude. To avoid amplifier saturation, the channel inputs are capacitively coupled to the first stage operational transconductance amplifier. This ensures DC input rejection of the amplifier. Amplification is performed by means of capacitive feedback. The mid-band gain is determined by the ratio $(C_{\text{in}})/(C_f)$, where $C_{\text{in}} = 5$

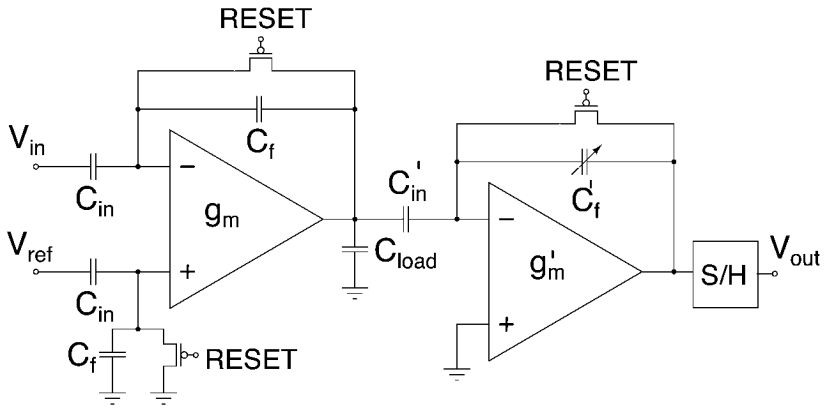


Fig. 3. Block diagram of the recording channele.

pF and $C_f = 100$ fF for the nominal value of 34 dB. The DC biasing point at the inputs of the amplifier is set by a resistive element in the negative feedback. Its value sets the cut-off frequency of a resulting high-pass filter. The spectral content of neural signals in the brain extends down to sub-hertz frequencies. To capture these frequencies, a large resistor, in the order of giga-ohms, should be utilized. As a large linear resistor can not be densely integrated, a MOS transistor biased in the subthreshold region of operation [1], [13] is employed. Its bias voltage sets the high-pass filter cut-off frequency which can be tuned between 0.01 Hz and 70 Hz. The amplifier also performs anti-aliasing filtering. The low-pass filter cut-off frequency is approximately equal to $(g_m)/(2\pi C_{\rm load})(C_{\rm in})/(C_f)$ and is tunable in the range from 500 Hz to 5 kHz as set by the bias current of the operational transconductance amplifier. Any DC drift due to junction leakage can be removed by periodic resetting the amplifier to the unity gain configuration [3]. In order to relax the trade-offs among noise, power and area, a telescopic operational transconductance amplifier is employed. As a wide dynamic range is not needed in the first stage, the telescopic topology yields savings in power dissipation and integration area, as discussed in detail in Section II.B.

The second stage of the channel is a single-ended capacitively-coupled voltage amplifier. Mid-band gain programmability is achieved by configuring a programmable bank of capacitors in its feedback. The second stage requires a large output dynamic range and employs a wide-swing current-mirror operational transconductance amplifier topology. As the noise contribution of the second stage is insignificant compared to that of the first stage, the power dissipation of the second stage can be reduced with little effect on the recording channel noise performance. The details of this design are given in [11] and [14].

A high-resistance MOS transistor is also employed in the second stage feedback to set a DC voltage at its input. Undesirable nonlinear distortion can be caused by changes in the feedback resistance of the second stage when the output voltage swing is large. This effect can be reduced by biasing the second stage at a lower cut-off frequency than that of the first stage or by using a lower setting of the second stage gain. Other second stage implementations such as connecting the feedback MOS resistor to a DC voltage [15] or using a purely resistive feedback [16] can eliminate these constraints.

The output of the second stage amplifier is sampled by a switched-capacitor sample-and-hold circuit. The sample-and-hold circuit contains two analog memories in order to store two consecutive signal samples as shown in Section II.C. This is necessary to compute temporal difference in delta compression of neural data as discussed in Section III.

### *B. Low-Power Neural Amplifier*

A critical component in the design of the recording channel is the first stage operational transconductance amplifier, as it contributes most of the circuit noise. Previously reported designs utilize several operational transconductance amplifier topologies with low-noise performance and a small power budget. A wide-swing current-mirror operational transconductance amplifier is proposed in [1], where the noise performance was optimized by increasing power consumption. A fully-differential folded-cascode architecture was reported in [3] where degeneration resistors considerably reduce the noise contribution of certain transistors and lead to an improvement of the overall noise performance without a power overhead. It is observed, however, that in these two common architectures a wide-swing operational transconductance amplifier topology has a dynamic range much larger than what is needed in the first stage of amplification in a multi-stage recording channel. These topologies can be replaced by a low-swing topology in the first stage of a channel with less power consumption and without any compromise in the noise performance.

The circuit diagram of the telescopic operational transconductance amplifier employed in the first stage of the recording channel is shown in Fig. 4. As will be shown next, this low-swing topology yields a factor of five reduction in power dissipation compared to the design in [1] for the same noise. This is due to fewer DC branches and fewer noise sources in the signal path. Transistor biasing conditions are critical for achieving low-noise performance while dissipating little power. At low frequencies it is possible to model the overall noise of an amplifier as a single voltage source in series with the input [17].

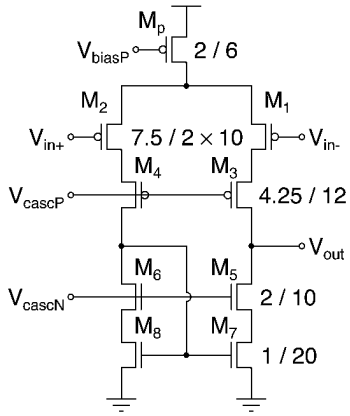


Fig. 4. Telescopic operational transconductance amplifier in the first stage of the recording channel (all transistor sizes are in units of micron).

The thermal component of the voltage noise of the telescopic operational transconductance amplifier can be shown to be

$$\overline{V_{n,\text{thermal}}^2} = 2\left[\frac{4kT}{g_{m1}}\left(\frac{2}{3}\right)\left(1+\frac{g_{m7}}{g_{m1}}\right)\right]\Delta f. \quad (1)$$

The thermal noise effect of transistors $M_{3\text{–}6}$ is negligible. It is clear that by biasing the transistors $M_{1,2,7,8}$ such that $g_{m7,8} \ll g_{m1,2}$ the noise contribution of $M_{7,8}$ is minimized. Thus, by making $W/L|_{7,8} \ll W/L|_{1,2}$ such that $M_{7,8}$ are biased in strong inversion and $M_{1,2}$ are biased in weak inversion, the thermal noise contribution of $M_{7,8}$ is minimized and the overall thermal noise performance is improved. Further noise reduction can be achieved by increasing the biasing current at the cost of higher power dissipation.

The $1/f$ noise of the transistors is accounted for by placing a voltage source given by $(K/(C_{\text{ox}}WL))1/f$ in series with each gate terminal. The $1/f$ input-referred noise power of the telescopic operational transconductance amplifier can be shown to be

$$\overline{V_{n,1/f}^2} = 2\left[\frac{K_P}{C_{\text{ox}}W_1L_1} + \frac{K_N}{C_{\text{ox}}W_7L_7}\left(\frac{g_{m7}}{g_{m1}}\right)^2\right]\frac{\Delta f}{f}. \quad (2)$$

The $1/f$ noise contribution of $M_{7,8}$ is scaled by a small factor of $(g_{m7,8}/g_{m1,2})^2$ making it insignificant compared to that of $M_{1,2}$. The $1/f$ noise contribution of $M_{1,2}$ is greatly reduced by employing large pMOS transistors in the input differential pair as they exhibit less $1/f$ noise than nMOS transistors [18]. The total input-referred noise is the sum of the two noise contributions in (1) and (2).

Table I summarizes the simulated characteristics of the the first-stage operational transconductance amplifier. Figs. 5(a) and (b) show its measured noise spectrum and amplitude response of the first stage. Fig. 6 depicts neural activity experimentally recorded by two neural amplifiers on the chip connected to off-chip recording electrodes. The shown neural activity corresponds to an epileptic seizure-like event chemically-induced in an intact hippocampus of a mouse. A characteristic seizure development is clearly seen.

TABLE I
SIMULATED TELESCOPIC OPERATIONAL TRANSCONDUCTANCE AMPLIFIER ELECTRICAL CHARACTERISTICS

| | |
|---|---|
| DC Gain | 88dB |
| Unity Gain Frequency (24pF) | 106.4kHz |
| Slew Rate (24pF) | 58.3V/ms |
| Output Voltage Swing | $0.5V_{pp}$ |
| Total Bias Current | $1.4\mu A$ |
| Input-Referred Thermal Noise | $52.7nV_{rms}/\sqrt{Hz}$ |
| Supply Voltage | 3V |

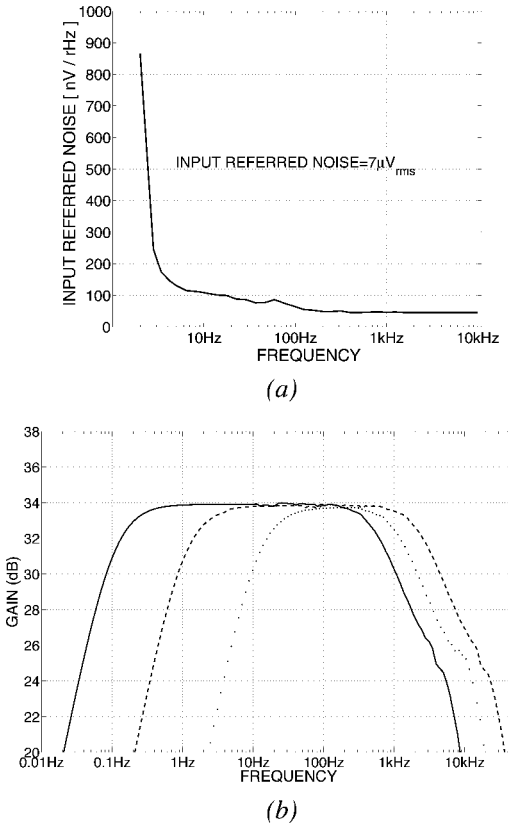


Fig. 5. Experimentally measured neural amplifier (the first stage) input-referred noise, (a); and amplitude frequency response for three sets of BPF programmable cut-off frequencies, (b).

Table II compares this neural amplifier with the previously reported designs in [1], [2], and [3]. The noise efficiency factor (NEF) [19] is comparable to the best one in [1] (15 percent more) but the area of the channel is significantly smaller (the first stage occupies half of the channel area) as dictated by the recording channel cell pitch requirement. This improvement in integration density is a result of several design choices. The telescopic topology offers a better trade-off between noise, power and area. This allows for a smaller amplifier active area with a comparable NEF. The minimum capacitor values are smaller due to the area constraints. Cascading stages reduces capacitor

TABLE II
COMPARISON OF LOW-POWER NEURAL AMPLIFIER CHARACTERISTICS

| | OTA Topology | Technology [$\mu m$] | Noise [$\mu V$] | Bandwidth [Hz] | Power [$\mu W$] | Voltage [V] | NEF | Gain [dB] | Area [$mm^2$] |
|---|---|---|---|---|---|---|---|---|---|
| [1] | Current mirror | 1.5 | 2.2 | 0.025-7.2k | 80 | 5 | 4.0 | 40 | 0.16 |
| [2] | Two-stage | 1.5 | 7.8 | 0.1-10k | 114.8 | 3 | 18.7 | 40 | 0.107 |
| [3] | Folded cascode | 0.6 | 11.7 | 0.1 - 100k | 160 | 5 | 8.1 | 20 | 0.062 |
| This work | Telescopic | 0.35 | 7 | 1 - 5k | 4.2 | 3 | 4.6 | 34 | 0.02 |

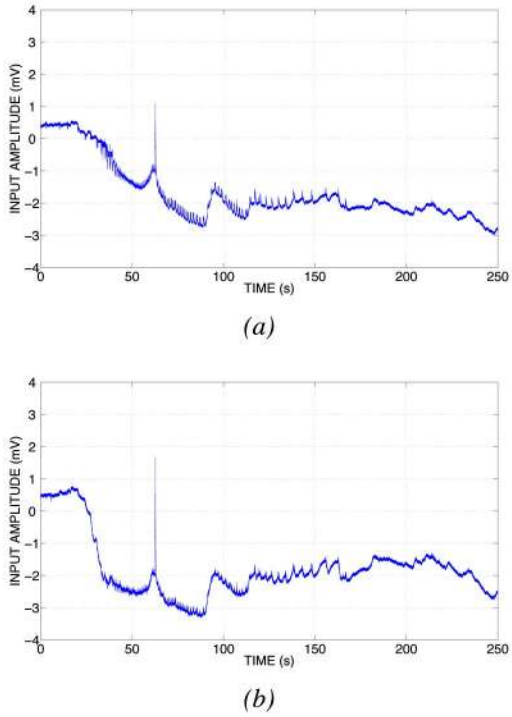


Fig. 6. Neural recordings experimentally measured with two neural amplifiers, (a) and (b). Epileptic seizure-like events were induced chemically by low $Mg^{2+}$.

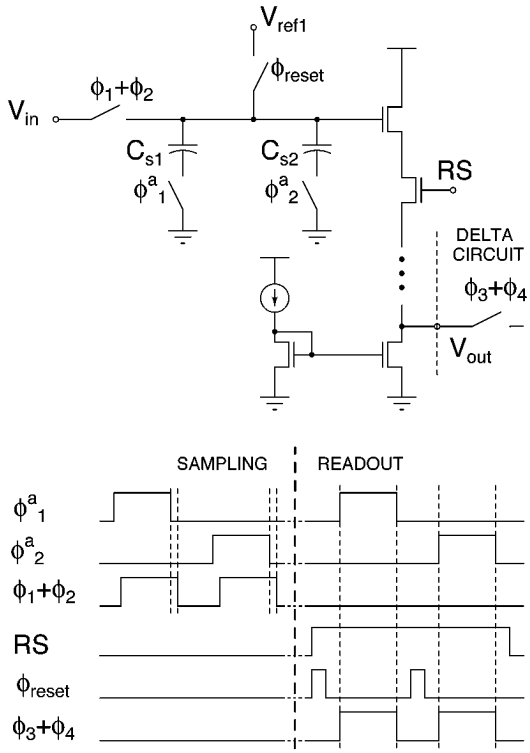


Fig. 7. Sample-and-hold circuit with double analog memory, and its timing diagram.

ratios. Using a technology with a smaller feature size has a minimal effect on the integration area as most of the transistors are not of minimum length and the capacitance per area is similar to that of the technology in [1]. Our technology, however, provided lower flicker noise which is partially due to differences in foundries.

### C. Sample-and-Hold Circuit

Accurate multi-site neural recording requires maintaining a high degree of correlation in time among all channels. Time-multiplexed recording architectures do not allow for such correlation unless the sampling frequency is very high. This necessitates a memory buffer in each channel to store the sampled signal. Two-dimensional frames of samples across the whole array are captured simultaneously and then read out sequentially.

The sample-and-hold cell includes two storage capacitors as shown in Fig. 7. The double memory array stores two subsequent neural activity frames as necessary for delta compression as described in Section III. To maintain charge injection from the sampling switch signal-independent, a two-switch sampling technique is used. The switches connected to the bottom plates of the sampling capacitors $C_{s1,2}$ have gate-to-source voltages independent of the input signal and thus always inject the same charge onto the capacitors. These switches are turned off shortly before the main sampling switch is turned off as shown in the timing diagram in Fig. 7. Charge injection by the main sampling switch is negligent as the capacitors are floating when it turns off.

Read-out is performed row-wise by setting the row select signal RS one row at a time as shown in the timing diagram

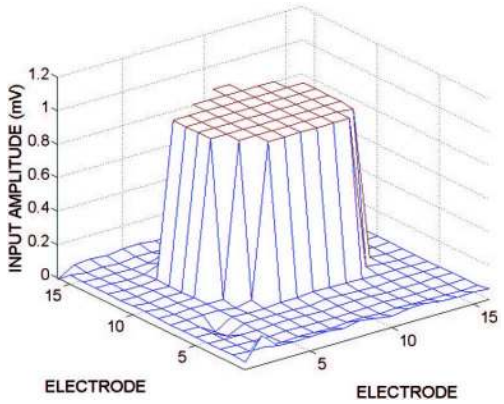


Fig. 8. A two-dimensional experimental recording of a water drop placed on a die and driven by a sinusoidal signal.

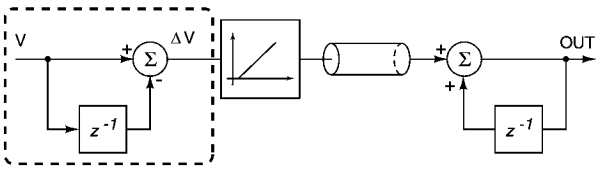


Fig. 9. Block diagram of delta compression and off-chip reconstruction of neural data.

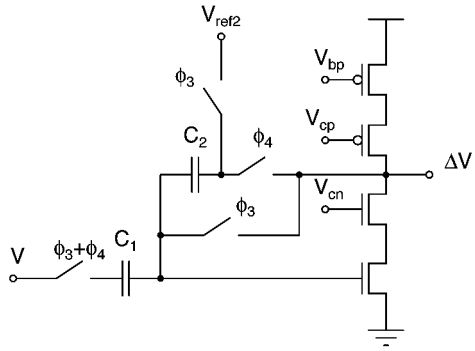


Fig. 10. Column-parallel switched-capacitor delta circuit.

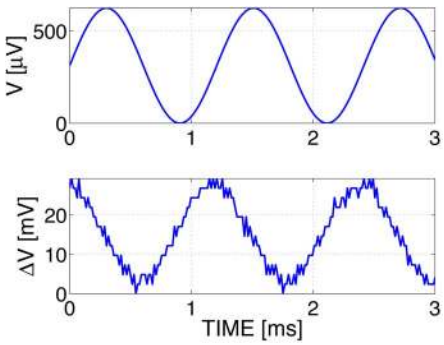


Fig. 11. *Top*, an input sinusoidal signal to the recording channel; *bottom*, experimentally measured output waveform (the temporal derivative of the input; accuracy limited by the off-chip ADC resolution).

in Fig. 7. The read-out sampling clock $\phi_3 + \phi_4$ is a sum (OR) of non-overlapping clocks $\phi_3$ and $\phi_4$ employed by the delta compression circuitry as will be described in Section III. To ensure memoryless charge sharing between the two sampling capacitors and the sampling node parasitics, the sampling node is charged to a fixed voltage $V_{\rm ref1}$ by a reset clock before each read-out.

In order to validate the two-dimensional recording functionality of the sample-and-hold circuit, the following experiment was conducted. A drop of water was placed on the surface of the $16 \times 16$ electrode array of one of the integrated prototypes and driven by a low-amplitude sinusoidal voltage. The stimulus signal was recorded at 5 kHz sampling rate and displayed on a PC in real time as an 'electronic video' stream. Fig. 8 shows a three-dimensional map of one frame recorded on a test chip corresponding to the peak value of the input sinusoid.

## III. Delta Compression

The neural recording array captures electronic images of the brain activity. Similarly to optical CMOS imagers [20], we utilize a data compression scheme known as delta compression for neural data reduction.

Delta compression is performed by computing the difference of two subsequent neural activity frames and discarding any resulting values that are below a certain threshold as depicted in Fig. 9. The neural data is compressed since only change in the activity above a certain threshold is transmitted. The original neural data is reconstructed on the receiver side by accumulating the incoming frame differences.

The $16 \times 16$-cell double analog memory in Fig. 7 captures two subsequent neural activity frames with no rolling delay. The column-parallel switched-capacitor delta circuit shown in Fig. 10 computes an approximation of a temporal derivative of the signal in each channel in one row, row by row. As mentioned above, the sampling clock $\phi_3 + \phi_4$ is a sum of two non-overlapping clocks, $\phi_3$ and $\phi_4$. As the circuit performs double sampling, it also removes any DC offsets due to channel-to-channel mismatches and reduces the effect of $1/f$ noise [21]. A single-ended cascoded common-source amplifier is employed yielding small integration area and low power consumption. The amplifier is designed and biased to ensure sufficient gain, bandwidth and slew-rate for the target 8-bit accuracy and 10 ksps frame sampling rate.

To experimentally characterize the delta circuit, a sinusoidal signal is applied to its input. As the delta circuit performs discrete differentiation, the output is another sinusoid with a 90° phase shift and an amplitude proportional to the ratio of the input frequency and the clocking frequency. Fig. 11 shows the experimentally measured input signal and output signal of the double sampling circuit. The shown output waveform accuracy is limited by the resolution of an off-chip analog-to-digital converter.

To demonstrate the utility of the delta data compression algorithm in neurophysiological applications, we simulate it on experimentally recorded epileptic seizure data shown in Fig. 6. A short segment of the recording in Fig. 6(a) is shown in Fig. 12(a). It is delta-compressed using thresholds of 50, 100 and 200 $\mu$V. The resulting reconstructed waveforms are shown in Figs. 12(b), (c) and (d), respectively. It can be observed that the seizure waveform is well preserved at these threshold levels. Other existing on-chip data compression methods such as adaptive thresholding in [4] are generally not well-suited for compressing such abnormal neural activity.

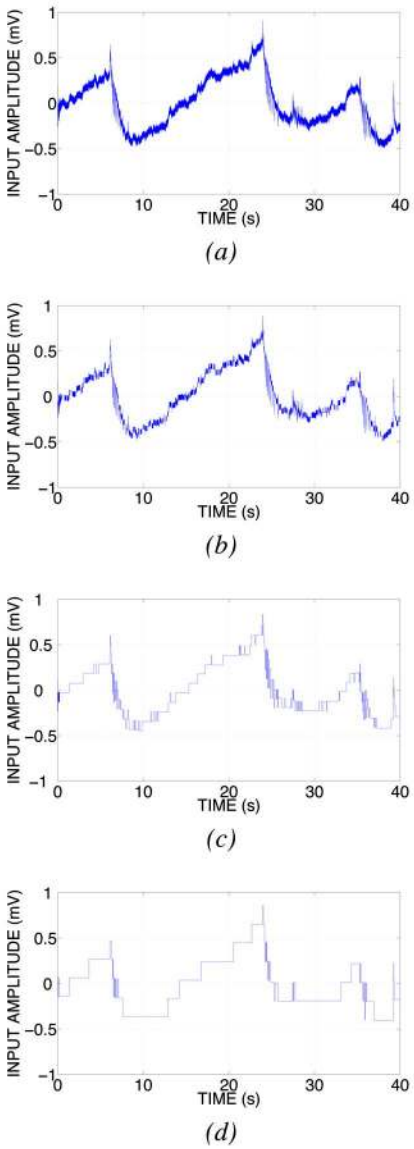


Fig. 12. (a), A segment of the recording in Fig. 6(a), and its simulated reconstructed versions for thresholds of: (b), 50 $\mu$V, (c), 100 $\mu$V, and (d), 200 $\mu$V.

By increasing the sensitivity threshold level, the compression ratio can be increased at the expense of reducing the accuracy of reconstructed data. To quantify this trade-off, the peak signal-to-noise ratio is calculated. It compares the quality of the reconstructed data with the original data:

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{3}$$

where MAX is the maximum recording channel output value for all samples, and MSE is the mean squared error for all samples. Fig. 13 illustrates the resulting simulated peak signal-to-noise ratio versus the compression ratio for the data shown in Fig. 12(a). The plot reflects a typical quality-size trade-off in data compression. The threshold value is set off-chip as needed for a specific neurophysiological application. Based on the data in Fig. 12(c), the threshold of 100 $\mu$V is deemed appropriate for neural data containing epileptic seizures corresponding to a compression ratio of over 800.

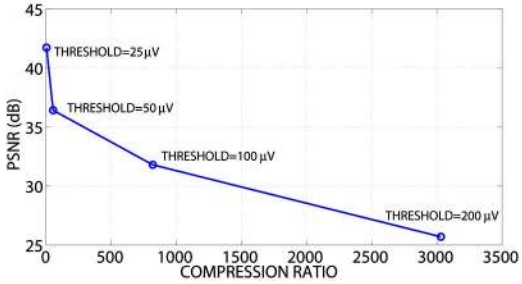


Fig. 13. Simulated peak signal-to-noise ratio (PSNR) of the reconstructed neural activity data.

## IV. Microsystem Integration

As the amplitude of recording signals is small (microvolts to millivolts) and the impedance of recording electrodes is high (kilo-ohms to mega-ohms), ambient electro-magnetic radiation interferes with the neural signals being recorded. Fabricating recording electrodes directly on the surface of the neural recording die minimizes such interference. The geometry and pitch of on-chip recording electrodes depends on requirements of a specific neurophysiological application. We previously reported a on-chip fabrication process of short (100 $\mu$m) Au electrodes for neural recording from acute brain slices of mice [11]. The neural recording prototype presented here can also be used for recording from acute slices at 256 sites utilizing the same electrodes as in [11]. In this paper we focus on applications where recording from a deeper volume of brain tissue is required.

For higher penetration depth applications, a Utah electrode array shown in Fig. 1 was bonded onto the die. The array is comprised of $10 \times 10$ silicon electrodes with shanks insulated with Parylene-C. Tips of the electrodes are exposed for 40–60 $\mu$m and coated with platinum for precisely localized extracellular recording. Bases of the electrodes are insulated with glass. Each electrode is 1.5 mm long. As the 400 $\mu$m Utah electrode array electrode pitch is twice the recording channel cell pitch, a set of $8 \times 8$ electrodes were bonded for a total of 64 recording sites. An electrode array with 200 $\mu$m electrode pitch can be bonded following the same procedure yielding a total of 256 recording sites.

Fig. 14 shows a cross-section of the assembled microsystem. A technology similar to a conventional flip-chip bonding technology is utilized. It is a reliable and low-cost method of integrating the two heterogeneous pitch-matched components together. The bonding process involves several steps.

The first step is gold stud bumping of the CMOS die. The gold stud bumping process is a modification of a regular wire-bonding procedure. In the wire bonding process, a bonding machine creates an electric connection between an aluminum pad on a die and the corresponding package lead by a golden wire. In stud bumping, the golden wire is broken off above the pad at a controlled height. The resulting stud bump provides a reliable electrical connection through aluminum oxide of the pad to the underlying metal layer. All stud bumps are subsequently coined in order to planarize their surface. Each stud bump is individually pressed by a wire pressing tool. Next, electrically-conductive epoxy is spotted onto the bases of the electrodes on the back

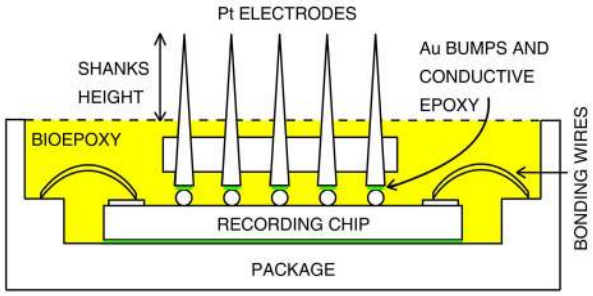


Fig. 14. A cross-section of the golden stud flip-chip bonded neural recording microsystem.

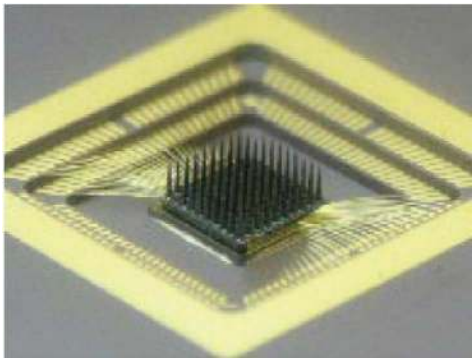

Fig. 15. A three-dimensional view of the assembled neural recording microsystem. The Utah electrode array is bonded on the CMOS die and packaged. The bio-compatible insulating epoxy is not shown.

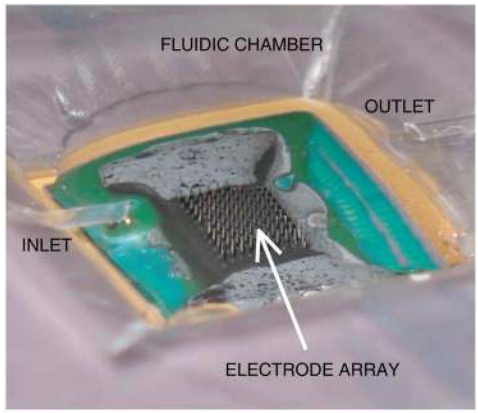


Fig. 16. The neural recording microsystem placed in a fluidic chamber.

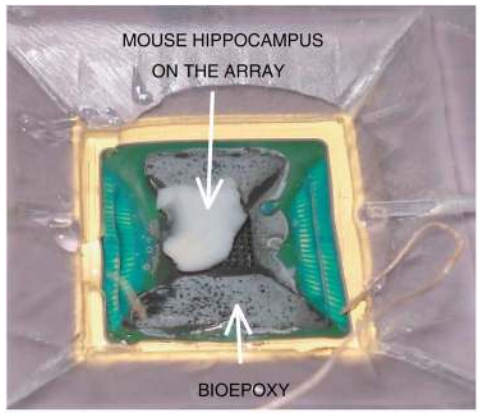


Fig. 17. An intact hippocampus of a mouse inserted onto the array for *in vitro* recordings.

side of the Utah electrode array. The stud bumps on the die are then optically aligned with the bases of the electrodes of the Utah electrode array. They are attached together, and the conductive epoxy is cured at 150C to make an electrical connection. After the bonding, a non-conductive under-fill adhesive is applied to completely fill the space between the bumps by the capillary action. The under-fill adds mechanical strength to the assembly. Heat-curing the under-fill adhesive at 150C completes the assembly process. Lastly, the package cavity is filled with a bio-compatible epoxy in order to electrically insulate and mechanically protect bonding wires, as well as to prevent any chemical damage to the neural tissue. It is cured at 125C. Fig. 15 shows a photograph of the fabricated neural recording microsystem prototype before bio-compatible epoxy is poured into the package cavity.

Gold stud bump flip-chip bonding offers several advantages. The bumping equipment, a wire bonder or a dedicated stud bumper, is widely available and well characterized. Since stud bumps are formed by wire bonders, they can be placed anywhere a wire bond can be placed. Thus, this technique is suitable for tight pitch electrode arrays. Since stud bumping can be done on a wire bonder, it does not require wafer-scale processing or under-bump metalization. A conventional CMOS die can be bumped and flip-chip bonded without any pre-processing. This makes stud bump flip-chip bonding low-cost, fast, and flexible for product development, prototyping and low-volume production, while easy to scale up to high-volume wafer-based production with automated equipment. Because stud bumping is a serial process, the bumping time increases with the number of bumps. Another limitation of stud bumping flip-chip bonding lies is somewhat low precision of die placement equipment and lower tolerance to placement errors than that of self-aligning solder assemblies [22].

For *in vitro* validation, the microsystem is placed into a custom-manufactured fluidic chamber as shown in Fig. 16. Fluidic perfusion of neural tissue is necessary to maintain its vitality. Additionally, certain neurophysiological behaviors can be induced by chemical modifications to the perfusion fluid. The chamber is manufactured from a bio-compatible plexiglass. It has one inlet and one outlet, and is sealed to the chip package with a liquid gasket.

Fig. 17 shows an intact hippocampus of a mouse (5–15 days old) inserted onto the electrodes of the microsystem for *in vitro* recording. Heated artificial cerebrospinal fluid is circulated through the inlet and outlet. The hippocampus rests on the tapered shanks of the electrodes and remains suspended above the bio-epoxy floor as necessary to keep it perfused from all sides. The composition of the circulating artificial cerebrospinal fluid is modified with low $Mg^{+2}$ in order to induce epileptic seizure-like events, which are then recorded through on-chip Utah electrode array electrodes, quantized by an off-chip ADC and transferred to a desktop computer for visualization through a custom-developed user interface. Fig. 18 depicts an example of neural activity recorded by the neural amplifier (the first

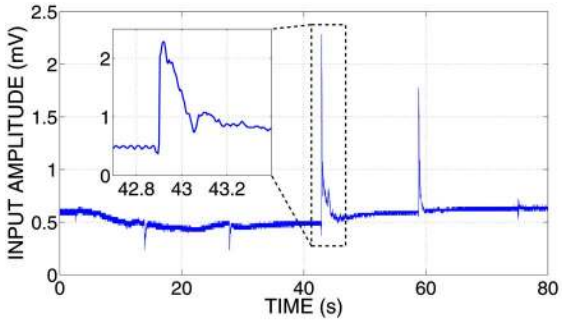


Fig. 18. An experimental recording from an intact hippocampus of a mouse through an on-chip Utah electrode array electrode. Seizure-like events were induced chemically by low $Mg^{2+}$.

TABLE III
EXPERIMENTAL CHARACTERISTICS

| | |
|---|---|
| Die Size | 3.5mm×4.5mm |
| Technology | $0.35\mu m$ CMOS |
| Number of Channels | 16×16 |
| Channel Dimensions | $200\mu m \times 200\mu m$ |
| Electrodes | $1500\mu m$, Pt tips |
| Supply Voltage | 3V |
| Programmable Gain | 48dB - 68dB |
| Total Input-Referred Noise | $7\mu V_{rms}$ |
| Noise Bandwidth | 1Hz - 5kHz |
| Noise Efficiency Factor | 4.6 |
| THD@4.4m$V_{pp}$ (first stage) | 0.7% |
| LPF Cut-off Frequency | 500Hz - 5kHz |
| HPF Cut-off Frequency | 0.01Hz - 70Hz |
| Max Sampling Rate | 10ksps |
| Power Dissipation | |
| Channel | $15\mu W$ |
| $\Delta$ Modulator | $45\mu W$ |
| Readout Circuits | 0.48mW |
| Total Power Dissipation | 5.04mW |

stage) without compression through one of the Utah electrode array electrodes bonded onto the chip. The recording represents a characteristic episode of a weak epileptic activity.

The experimentally measured characteristics and physical properties of the fabricated brain-silicon interface prototype are summarized in Table III. The immediate applications of the microsystem are in *in vitro* animal studies and high-throughput drug screening. The small form factor of the microsystem allows for future on-chip integration of analog-to-digital converter and digital circuits as needed for wireline *in vivo* animal studies. The power dissipation budget is suitable for through-skin wireless power transmission and harvesting [23]. Recorded data can also be transmitted wirelessly as the on-chip data compression circuitry relaxes the requirements on the data transmitter power dissipation [24]. The long-term applications of the microsystem include neural prostheses for neurological disorders therapy such as for epileptic seizure automated prediction and prevention [25] and central and peripheral nervous system rehabilitation [7].

## V. CONCLUSION

We have presented an architecture and VLSI implementation of a neural interface microsystem for high-resolution neural activity recording and analog delta compression. A low-power compact telescopic neural amplifier design is presented. The delta compression scheme allows for a flexible trade-off between accuracy and output data rate. An array of three-dimensional recording microelectrodes is integrated with the die by means of a low-cost on-chip bonding process. The microsystem prototype is validated in extracellular *in vitro* recording from intact hippocampus of mice.

## ACKNOWLEDGMENT

The authors thank Canadian Microelectronics Corporation for providing the fabrication services and facilitating the electrode bonding process.

## REFERENCES

[1] R. R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE J. Solid-State Circuits*, vol. 38, no. 6, pp. 958–965, Jun. 2003.

[2] P. Mohseni and K. Najafi, "A fully integrated neural recording amplifier with DC input stabilization," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 5, pp. 832–837, May 2004.

[3] F. Heer, S. Hafizovic, W. Franks, A. Blau, C. Ziegler, and A. Hierlemann, "CMOS microelectrode array for bidirectional interaction with neuronal networks," *IEEE J. Solid-State Circuits*, vol. 41, no. 7, pp. 1620–1629, Jul. 2006.

[4] R. H. Olsson and K. D. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2796–2804, Dec. 2005.

[5] R. Harrison, P. Watkins, R. Kier, R. Lovejoy, D. Black, R. Normann, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," in *IEEE Int. Conf. Solid-State Circuits Dig. Tech. Papers*, Feb. 6–9, 2006, pp. 2258–2267.

[6] K. Wise, "Silicon microsystems for neuroscience and neural prostheses," *IEEE Eng. Med. Biol. Mag.*, vol. 24, no. 5, pp. 22–29, Sep.–Oct. 2005.

[7] W. Patterson, Y. Song, C. Bull, I. Ozden, A. Deangellis, C. Lay, J. McKay, A. Nurmikko, J. Donoghue, and B. Connors, "A microelectrode/microelectronic hybrid device for brain implantable neuroprosthesis applications," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 10, pp. 1845–1853, Oct. 2004.

[8] B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes, "A 128 × 128 CMOS biosensor array for extracellular recording of neural activity," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2306–2317, Dec. 2003.

[9] J. N. Y. Aziz, R. Genov, B. Bardakjian, M. Derchansky, and P. L. Carlen, "256-channel integrated neural interface and spatio-temporal signal processor," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 21–24, 2006, pp. 5075–5078.

[10] S. Kim, R. Normann, R. Harrison, and F. Solzbacher, "Preliminary study of the thermal impact of a microelectrode array implanted in the brain," in *Proc. IEEE Engineering in Medicine and Biology Conf.*, Aug. 30–Sept. 3 2006, pp. 2986–2989.

[11] J. N. Y. Aziz, R. Genov, B. Bardakjian, M. Derchansky, and P. L. Carlen, "Brain-silicon interface for high-resolution in-vitro neural recording," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 1, Apr. 2007.
[12] J. Aziz, R. Genov, M. Derchansky, B. Bardakjian, and P. Carlen, "256-channel neural recording microsystem with on-chip 3D electrodes," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 11–15, 2007, pp. 160–161.
[13] T. Delbruck and C. A. Mead, "Adaptive photoreceptor with wide dynamic range," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1994, pp. 339–342.
[14] J. Aziz, R. Karakiewicz, R. Genov, B. L. Bardakjian, M. Derchansky, and P. L. Carlen, "Real-time seizure monitoring and spectral analysis microsystem," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2006.
[15] W. Wattanapanitch, M. Fee, and R. Sarpeshkar, "An energy-efficient micropower neural recording amplifier," *IEEE Trans. Biomed. Circuits Syst.*, vol. 1, no. 2, pp. 136–147, Jun. 2007.
[16] R. R. Harrison, "A versatile integrated circuit for the acquisition of biopotentials," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 16–19, 2007, pp. 115–122.
[17] C. D. Motchenbacher and J. A. Connelly, *Low-Noise Electronic System Design*, 1st ed. New York: Wiley, Jun. 1993.
[18] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. Oxford, U.K.: Oxford Univ. Press, Jun. 2003.
[19] M. S. J. Steyaert, W. M. C. Sansen, and C. Zhongyuan, "A micropower low-noise monolithic instrumentation amplifier for medical purposes," *IEEE J. Solid-State Circuits*, vol. 22, pp. 1163–1168, Dec. 1987.
[20] U. Mallik, M. Clapp, E. Choi, G. Cauwenberghs, and R. Etienne-Cummings, "Temporal change threshold detection imager," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2005, pp. 362–603.
[21] C. C. Enz and G. C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: Autozeroing, correlated double sampling, and chopper stabilization," *Proc. IEEE*, vol. 84, no. 11, pp. 1584–1614, Nov. 1996.
[22] G. Riley, "Bump, dip, flip: Single chip," in *Surface Mount Int.*, Sep. 1987, pp. 535–541.
[23] N. Chaimanonart, K. R. Olszens, M. D. Zimmerman, W. H. Ko, and D. J. Young, "Implantable RF power converter for small animal in vivo biological monitoring," in *Proc. IEEE Engineering in Medicine and Biology Conf.*, Sep. 2005, pp. 5194–5197.
[24] S. Atluri and M. Ghovanloo, "A wideband power-efficient inductive wireless link for implantable microelectronic devices using multiple carriers," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 21–24, 2006, pp. 1131–1134.
[25] J. N. Y. Aziz, R. Karakiewicz, R. Genov, A. W. L. Chiu, B. L. Bardakjian, M. Derchansky, and P. L. Carlen, "In vitro epileptic seizure prediction microsystem," in *Proc. IEEE Int. Symp. Circuits and Systems*, May 2007.

**Joseph Aziz** (S'05–M'06) received the B.Sc. degree in electronics and electrical communications engineering from Cairo University, Giza, Egypt, in 2003, and the M.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 2007.

He is currently with Broadcom Corporation, Irvine, CA. His current research interests focus on the design of mixed-signal integrated circuits.

**Karim Abdelhalim** received the B.Eng. and M.A.Sc. degrees in electrical engineering from Carleton University, Ottawa, ON, Canada, in 2005 and 2007, respectively. His M.A.Sc. thesis focused on the implementation of an ultra-low-power ADC for biomedical applications. He is currently working toward the Ph.D. degree in electrical and computer engineering at the University of Toronto, Toronto, ON, Canada. His research interests are in the area of low-power implantable integrated circuits for neural interfaces.

Mr. Abdelhalim is a recipient of the Alexander Graham Bell Canada Graduate Scholarship awarded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

**Ruslana Shulyzki** received the B.A.Sc. degree from the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, in 2006. Currently, she is working toward the M.A.Sc. degree in the same department. Her research interests are in the area of low-power implantable neural microsystems.

**Roman Genov** (S'96-M'02) received the B.S. degree (first rank) in electrical engineering from Rochester Institute of Technology, Rochester, NY, in 1996, and the M.S. and Ph.D. degrees in electrical and computer engineering from Johns Hopkins University, Baltimore, MD, in 1998 and 2002, respectively.

He held engineering positions at Atmel Corporation, Columbia, MD, in 1995 and Xerox Corporation, Rochester, NY, in 1996. He was a visiting researcher in the Laboratory of Intelligent Systems at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1998 and in the Center for Biological and Computational Learning at the Massachusetts Institute of Technology, Cambridge, MA, in 1999. He is presently an Associate Professor in the Department of Electrical and Computer Engineering at the University of Toronto, Canada. His research interests include analog and digital VLSI circuits, systems and algorithms for energy-efficient signal processing with applications to electrical, chemical and photonic sensory information acquisition, biosensor arrays, brain-silicon interfaces, parallel signal processing, adaptive computing for pattern recognition, and implantable and wearable biomedical electronics.

Dr. Genov received the Canadian Institutes of Health Research (CIHR) Next Generation Award in 2005, the DALSA Corporation Componentware/CAD Award in 2006, and the Brian L. Barge Award for Excellence in Microsystems Integration in 2008. He is an Associate Editor of IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS and IEEE SIGNAL PROCESSING LETTERS.

**Berj L. Bardakjian** received the Ph.D. degree in electrical engineering (Biomedical Engineering Group) from McMaster University, Hamilton, Ontario, Canada.

His previous positions included being a Medical Research Council (MRC) postdoctoral fellow in the Department of Physiology, then MRC Scholar in the Institute of Biomaterials and Biomedical Engineering, at the University of Toronto, and an investigator in the Playfair Neuroscience Unit at the Toronto Western Hospital. He is currently a Professor of biomedical and electrical engineering at the University of Toronto, Canada. He is also an associate editor for Annals of Biomedical Engineering. His research interests include biological and artificial neural networks, generation and coupling mechanisms of neural bioelectricity, prediction and control of epileptic seizures, modeling of nonlinear physiological systems, biological clocks, spatiotemporal processing of nonstationary electrical signals from the brain.

**Miron Derchansky** received the Ph.D. degree in physiology and neuroscience from the University of Toronto, Toronto, ON, Canada.

Utilizing extracted brain tissues, his primary interests include the electrophysiological mechanisms and spread of seizure activity. To address this issue, he employs single-cell and multisite field recordings, coupled with voltage sensitive dyes. He has also received an M.B.A. from the University of Massachusetts and is currently the cofounder of a startup biotechnology company focused on developing electrophysiological lab equipment, adaptive signal processing algorithms, and medical devices for the prediction and control of dynamic brain states.

**Demitre Serletis** received the B.Sc. degree in biological sciences and the M.D. degree from the University of Calgary, Alberta, Canada.

He is currently training as a Neurosurgery Resident Physician at the University of Toronto, Canada, where he is also pursuing a collaborative Ph.D. degree in the Department of Physiology and the Institute of Biomaterials and Biomedical Engineering with support through the Surgeon Scientist and Clinician Investigator Programs. His main research interests are the neurodynamical and physiological complexity of rhythms and noise-like activity in the brain, and the detection and prediction of epileptic seizures.

**Peter L. Carlen** is a clinical neurologist, specializing in epilepsy and neurodegenerative diseases at the Toronto Western Hospital of the University Health Network. He was trained in medicine and neurology at the University of Toronto, Canada. He studied cellular electrophysiology for three years at the Neurobiology Department of the Hebrew University of Jerusalem. He then returned to Toronto where he was a staff neurologist and researcher at the Toronto Western Hospital and the Addiction Research Foundation starting in 1975. In 1989, he was appointed Director of the Playfair Neuroscience Unit and Neuroscience Research at the University Health Network for a 10 year term, where he is now a senior scientist and Head of the Division of Fundamental Neuroscience. He is also a Professor in the Departments of Medicine (Neurology), Physiology and the Institute of Biomaterials and Biomedical Engineering of the University of Toronto. He has over 200 peer-reviewed biomedical publications and holds 6 patents. His main research interests are mechanisms of epilepsy and neurodegeneration.