The Design of Integrated Circuits to Observe Brain Activity

Severe power and size restraints on circuitry implanted in the brain present novel design challenges at both the circuit and system levels.

By Reid R. Harrison, *Member IEEE*

ABSTRACT | The ability to monitor the simultaneous electrical activity of multiple neurons in the brain enables a wide range of scientific and clinical endeavors. Recent efforts to merge miniature multielectrode neural recording arrays with integrated electronics have revealed significant circuit design challenges. Weak neural signals must be amplified and filtered using low-noise circuits placed close to the electrodes themselves, but power dissipation must strictly be limited to prevent tissue damage due to local heating. In modern recording systems with 100 or more electrodes, raw data rates of 15 Mb/s or more are easily produced. Micropower wireless telemetry circuits cannot transmit information at such high rates, so data reduction must be performed in the implanted device. In this paper, we present integrated circuits and design techniques that address the twin problems of neural signal amplification and data reduction for this severely size- and power-limited application.

KEYWORDS | Amplifiers; analog integrated circuits; biomedical signal processing; low-power circuit design; neural recording; subthreshold circuit design

I. INTRODUCTION

The development of micromachined multielectrode arrays in the 1980s and 1990s has revolutionized modern neuroscience, permitting scientists and clinicians to monitor the simultaneous activity of many neurons in localized regions of the brain [1]–[4]. One example of this technology, shown in Fig. 1 (top), is the Utah Electrode Array. This 10 × 10 array of platinum-tipped silicon

electrodes measures $4 \times 4 \times 1.5$ mm^3 [3]. Dense multielectrode arrays are now readily available from multiple commercial sources, but these devices must be connected to external instrumentation via relatively bulky wire bundles and transcutaneous connectors.

In an effort to create fully implantable neural recording devices with wireless power and data transfer, microelectromechanical system (MEMS) electrode arrays are being combined with integrated complementary metal–oxide–semiconductor (CMOS) electronics [5]–[8]. The miniaturization of these highly parallel recording systems presents several significant circuit design challenges. The weak neural signals must be amplified and digitized, and this information must be relayed out of the body using a wireless telemetry link to avoid any path for infection. Multichannel neural recording systems potentially produce large quantities of continuously streaming data that must be transmitted. Yet the power dissipation of small implanted devices must be strictly limited to prevent excessive tissue heating that can kill nearby cells [9], [10].

Fig. 1 (bottom) shows a block diagram of a generic wireless neural recording device. A bank of amplifiers must be used to boost the weak signal measured by each electrode. Differential amplifiers are used to measure the potential of each signal electrode with respect to a large, low-impedance reference electrode (such as a short length of thin platinum wire). In some recording applications, multiple reference electrodes are used, and a small number of signal electrodes (or a single signal electrode) are paired with a nearby high-impedance reference electrode.

In most neural recording applications, each signal electrode must have its own dedicated low-noise amplifier. Although it is tempting to imagine using an analog multiplexer to timeshare a single amplifier between multiple electrodes, the time constants inherent in the amplifier dynamics are typically much longer than the multiplexer switching time required to catch brief neural activity across
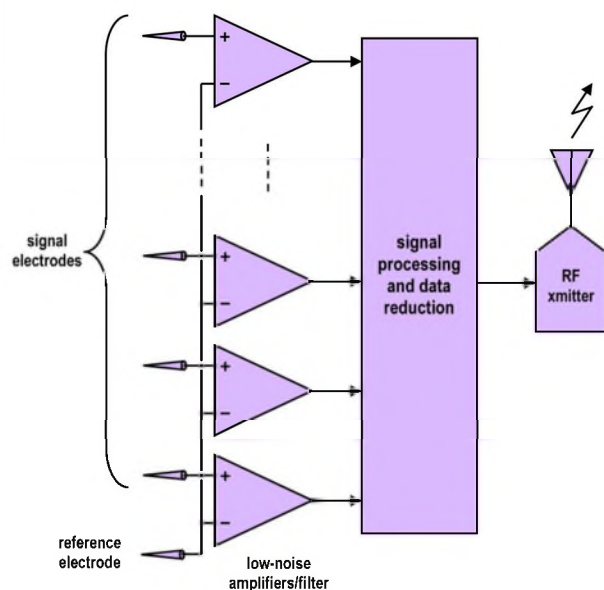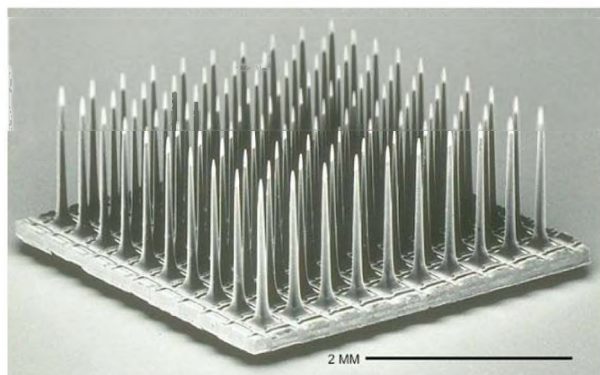
Fig. 1. *(Top) Scanning electron micrograph of silicon-based Utah Electrode Array [4]. The 100-electrode array measures 4 × 4 × 1.5 mm³. (Bottom) Block diagram of wireless neural recording device.*

an array of many electrodes. Thus, this array of amplifiers can consume relatively large amounts of power and chip area in a multichannel neural recording system.

After amplification and some continuous-time filtering (including antialias filtering, if the signal is to be digitized), the resulting vector of signals must be conditioned for wireless transmission. This requires the ordering of the parallel multielectrode signals into a single serial data stream. In many cases, this conditioning also involves digitization, as digital transmission can be made relatively robust through the proper use of source coding and carrier modulation. Finally, the serial signal is sent to a radio-frequency (RF) modulator and amplifier for transmission to a receive antenna located outside the body.

Implantable neural recording devices have great promise for advancing the understanding of brain function by allowing scientists to observe and manipulate neural

activity during normal animal behavior [11]. Clinical applications for this technology include monitoring and diagnosis of epileptic seizures and prosthetic control for the severely disabled. Recent work in the field of neuroprosthetics has demonstrated that rats [12], monkeys [13]–[15], and paralyzed humans [16], [17] can learn to control robotic arms or computer cursors by thoughts if multiple neural signals from motor regions of the cerebral cortex are used for control.

In this paper, we describe analog and mixed-signal integrated circuits designed and optimized to amplify neural signals in a power-efficient manner (Section III) and extract relevant information from these signals to aid in pretransmission data reduction (Sections IV and V). Since analog circuits are sensitive to device mismatch and other effects not fully captured in simulation, all circuits presented here have been fabricated in commercial silicon CMOS processes. To guide the design of these circuits, we begin with a description of the signals observed in the brain.

II. THE NATURE OF NEURAL SIGNALS

Electrically active cells such as neurons produce internal voltage changes on the order of 100 mV relative to the extracellular fluid [18]. While brief intracellular recordings are possible using individually guided microelectrodes, chronic recordings using multielectrode arrays make use of the smaller extracellular potentials measured several micrometers from the cell. The contact between metal electrode tip and extracellular fluid creates an electrical double layer so the electrode-tissue interface behaves primarily as a capacitance for small voltages [19], [20]. The capacitance of the interface depends on electrode area and surface roughness; values between 150 pF–1.5 nF are common in recording electrodes. A typical trace from an extracellular neural recording is shown in Fig. 2. This waveform was recorded from the motor cortex of an awake cat using a Utah Electrode Array that was implanted approximately three months prior [21].

The features present in typical extracellular neural recordings are visible in Fig. 2. The amplitude of the signal is on the order of 100 μV, and rapid neural action potentials or "spikes" from a nearby neuron are present (at $t = 100$, 140, and 180 ms). Neural spikes often appear biphasic in extracellular recordings and usually have durations of 0.3–1.0 ms. Neurons rarely fire more rapidly than 100 spikes per second (though rapid bursts of several spikes are possible), with firing rates around 10 Hz somewhat typical in cerebral cortex. Spikes are "digital" events; neurons produce spikes of nearly identical amplitude and duration, and information is encoded in the timing of spikes [18].

Also present are low-frequency (< 200 Hz) oscillations known as local field potentials (LFPs). Local field potentials arise from the synchronous activity of many neurons in one region of the brain. These neurons are too
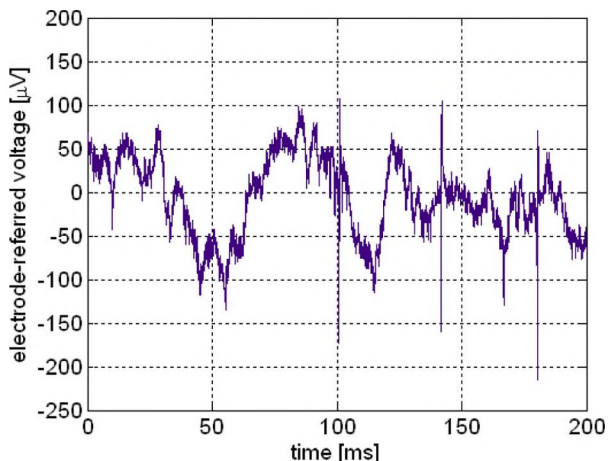
Fig. 2. *Neural recording from cat motor cortex using Utah Electrode Array and integrated CMOS amplifier. Both spikes and LFPs are visible.*
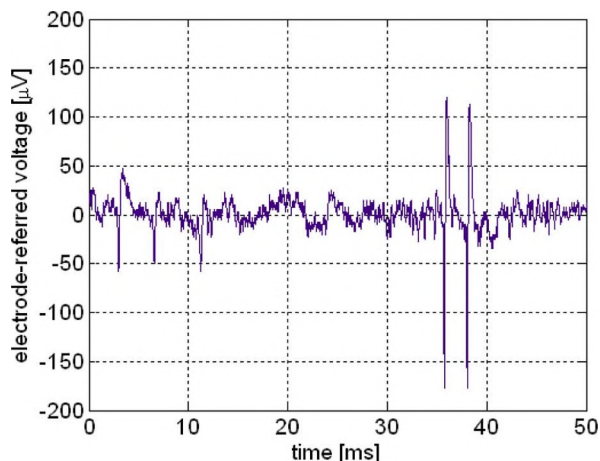

Fig. 4. *Large and small spikes observed in cat motor cortex. A one-pole high-pass filter at 300 Hz was applied to the signal to remove LFPs.*

distant from the electrode for their individual action potentials to be resolved, but the "crowd noise" of many neighboring cells creates a large signal that is easily detected [22]. The LFP is the internal correlate of the electroencephalograph signal measured on the scalp (after much attenuation and spatial blurring). The energy of LFP signals in the primate premotor and motor cortex has been shown to correlate with specific arm movement reach parameters such as direction, distance, and speed, and thus may be useful in neuroprosthetic applications [23]–[27]. Fig. 3 shows the onset of pronounced beta waves (10–15 Hz) in cat motor cortex [21]. LFPs are a robust signal. In some experiments using electrode arrays, scar tissue forms around microelectrode tips. This scar

tissue tends to attenuate spike signals from nearby neurons, but LFP signals are less affected [28].

In many applications, it is desirable to separate LFP and spike signals so they may be analyzed separately. This is easily accomplished by linear filtering since LFPs occupy frequencies from approximately 10–200 Hz, while spikes have energy concentrated in the 300 Hz–5 kHz range. Fig. 4 shows a neural signal that has been high-pass filtered at 300 Hz, isolating the spikes [21]. Here, a nearby neuron producing relatively large spikes fires twice between 35 and 40 ms, while a more distant neuron fires three spikes between 3 and 12 ms. When multielectrode arrays are placed in the brain, it is common for some electrodes to detect spikes from two to four distinct neurons, while other electrodes may see no resolvable spikes. Fig. 5 shows
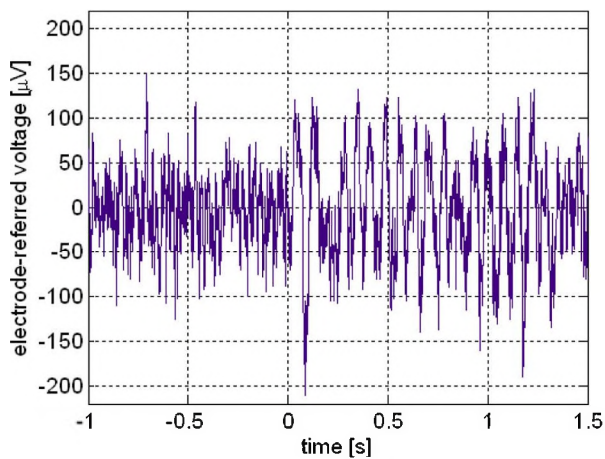

Fig. 3. *Neural recording from cat motor cortex showing spontaneous onset of 10–15 Hz beta activity in the LFP at t = 0. Note that the time scale is much longer than in Fig. 2.*
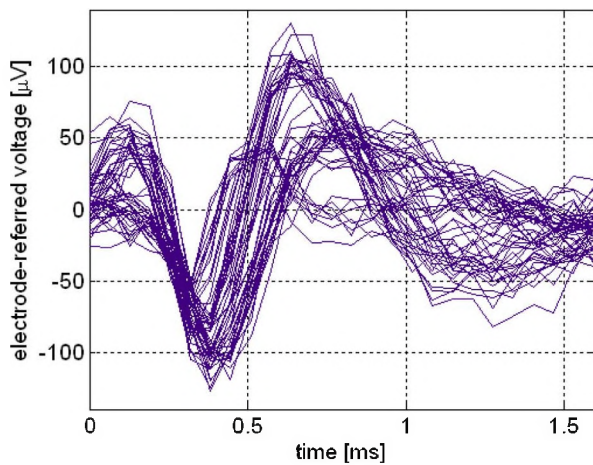

Fig. 5. *Time-aligned spikes recorded from cat auditory cortex using Utah Electrode Array, integrated amplifier, and wireless telemetry [1]. Three distinct neurons are visible.*

51 time-aligned spikes recorded from cat auditory cortex using a Utah Electrode Array. (These data were gathered by the wireless system described in [8].) Three distinct waveform shapes are visible, corresponding to three nearby neurons with different distances and/or orientations with respect to the electrode tip.

III. NEURAL SIGNAL AMPLIFICATION

A. Design Requirements

Due to the small amplitude of neural signals recorded extracellularly and the high impedance of the electrode-tissue interface, amplification must be performed before these signals can be digitized or analyzed in any way. An integrated front-end amplifier for neural signals must:

1) have sufficiently low input-referred noise to resolve spikes as small as 30 μV in amplitude;
2) have sufficient dynamic range to convey spikes or LFPs as large as ± 1–2 mV in amplitude;
3) have much higher input impedance than the electrode-tissue interface and have negligible dc input current;
4) amplify signals in the frequency bands of interest (roughly 300 Hz–5 kHz for spikes and 10–200 Hz for local field potentials);
5) block dc offsets present at the electrode-tissue interface to prevent saturation of the amplifier; and
6) consume little silicon area and use few or no off-chip components to minimize size.

In addition to these requirements, the amplifier should have a high common-mode rejection ratio to minimize interference from 50/60 Hz power line noise, and a high power-supply rejection ratio if power supply noise is significant (e.g., from ac inductive power links). Arrays of amplifiers should have low crosstalk between channels.

To reduce pickup of 50/60 Hz noise, microphonics, and other capacitively and inductively coupled interferers, the distance between electrode and amplifier should be minimized. Additionally, tethering forces introduced by wires cause problems for electrodes inserted into the soft, pliable brain tissue. Thus, the amplifiers are ideally attached directly to the electrodes very near the recording site. This proximity of the electronics to living tissue imposes strict limits on the amount of power that can be dissipated by the circuitry; if cells are exposed to elevated temperatures for extended periods of time, they will die [9], [10]. Thus, we add another requirement for neural signal amplifiers: operation at low power levels to minimize tissue heating.

The precise limits to power dissipation in implanted devices can be difficult to establish. Most devices are designed to limit the chronic heating of surrounding tissue to less than 1 °C. Thus, the size and shape of a device determine its power limits; smaller devices can dissipate less power safely. While heat conduction in the body can be simulated with some accuracy, convection from blood flow cannot be modeled accurately and thus experimental validation is required. Preliminary experiments have shown that an implanted cortical 100-electrode array with integrated electronics measuring roughly $6 \times 6 \times 2$ mm^3 can safely dissipate approximately 10 mW of power [29], [30]. This power limit poses a challenge for high-channel-count recording systems since each electrode requires a dedicated low-noise amplifier.

A rough order-of-magnitude analysis of multichannel neural recording devices presents a sobering picture for circuit designers: with modern MEMS arrays providing approximately 100 electrodes and a power dissipation limit of 10 mW, each channel must consume less than 100 μW, and this does not even include shared resources on a chip such as analog-to-digital conversion, power regulation, control, and telemetry circuits.

B. Circuit Architecture and Design Techniques

Fig. 6 shows the schematic of a neural signal amplifier that was first described in [32]. The amplifier is based around an operational transconductance amplifier (OTA) that produces a current proportional to the differential voltage applied to its inputs, where G_m is the constant of proportionality. A capacitive feedback network consisting of C_1 and C_2 capacitors sets the midband gain of the amplifier. (C_{in} models the input capacitance of the OTA, as well as any bottom-plate capacitance from C_1 and C_2.) The input is capacitively coupled through C_1, so any dc offset from the electrode-tissue interface is removed. C_1 should be made much smaller than the electrode impedance to minimize signal attenuation.

The R_2 elements shown in the feedback loop represent lossy elements that set the low-frequency amplifier cutoff;
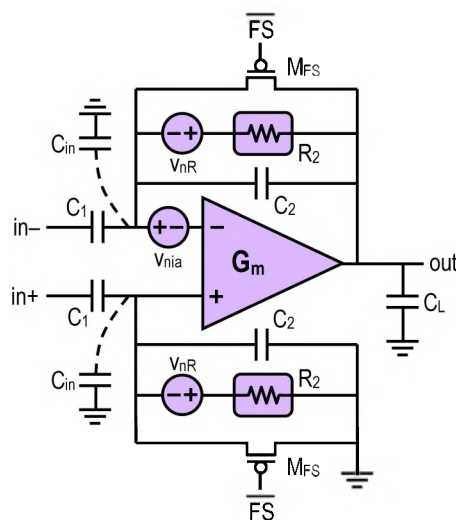

Fig. 6. Schematic of OTA-based neural signal amplifier with capacitive feedback.
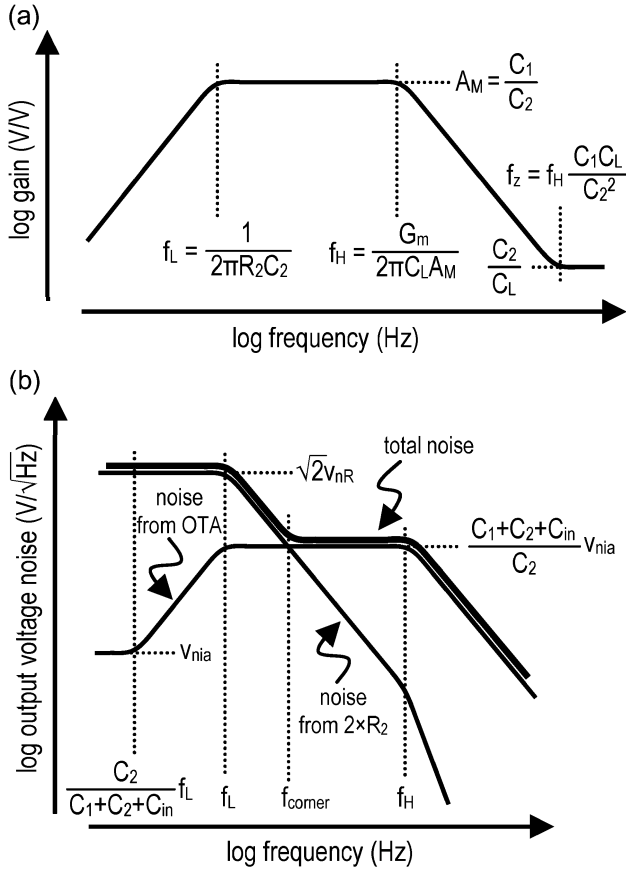
(a)

(b)

Fig. 7. *Log–log plot of (a) gain versus frequency for the neural amplifier shown in Fig. 6 and (b) neural amplifier output noise versus frequency.*

they may be implemented using real resistors, but the MOS-bipolar element used in [32] provides an area-efficient means of creating a small-signal resistance of $> 10^{12}$ Ω for low-frequency operation (i.e., LFPs). The long time constant associated with this pole can cause the amplifier to recover slowly from large transients, so the M_{FS} transistors can act as switches to implement a "fast settle" function.

Fig. 7(a) shows a gain versus frequency plot for the neural amplifier in Fig. 6. The approximate transfer function is given by

$$\frac{v_{out}}{v_{in+} - v_{in-}} = \frac{C_1}{C_2} \cdot \frac{1 - sC_2/G_m}{\left(\frac{1}{sR_2C_2} + 1\right)\left(s\frac{C_LC_1}{G_mC_2} + 1\right)}$$

$$= A_M \frac{1 - s/(2\pi f_z)}{\left(\frac{2\pi f_L}{s} + 1\right)\left(\frac{s}{2\pi f_H} + 1\right)}. \quad (1)$$

The midband gain A_M is set by the capacitance ratio C_1/C_2, and the gain is flat between the lower and upper cutoff

frequencies f_L and f_H. The lower cutoff frequency is determined by the product of R_2 and C_2, while the upper cutoff is determined by the load capacitance C_L, the OTA transconductance G_m, and the midband gain. Capacitive feedthrough introduces a right-half-plane zero at f_z. This zero can be pushed to very high frequencies (higher than secondary poles due to parasitic capacitances in the OTA) by setting

$$C_2 \ll \sqrt{C_1C_L} \quad (2)$$

so that it has little practical effect on amplifier operation.

The thermal noise sources in the neural amplifier are shown in Fig. 6 as voltage sources v_{nia} and v_{nR}. The source v_{nia} models the input-referred voltage noise of the OTA. (If MOSFETs are used as input devices, then the current noise is negligible at low frequencies.) The two v_{nR} sources model the thermal noise (or Johnson noise) contributed by the resistive R_2 elements in the feedback loop. Fig. 7(b) shows the contributions to the total amplifier output noise when both v_{nia} and v_{nR} are taken to by white (i.e., ignoring $1/f$ noise). The OTA contributes noise primarily between f_L and f_H. Below a particular frequency, the noise contribution from v_{nR} will dominate; we denote this frequency f_{corner}. If R_2 is implemented as a real resistor so that its noise spectral density is

$$v_{nR}^2(f) = 4kTR_2 \quad (3)$$

and $C_1 \gg C_2, C_{in}$, then f_{corner} is approximately

$$f_{corner} \approx \sqrt{\frac{3C_L}{2C_1}f_Lf_H}. \quad (4)$$

(A similar result is obtained for the MOS-bipolar element used as R_2 in [32].) To minimize the noise contribution from the R_2 elements, we should ensure that $f_{corner} \ll f_H$. For resistive R_2 elements, this can be accomplished by designing the amplifier so that

$$\frac{C_L}{C_1} \ll \frac{2}{3}\frac{f_H}{f_L}. \quad (5)$$

In practical circuits, the $1/f$ noise from the OTA may dominate the noise contributed by the R_2 elements. However, if multitransistor, amplifier-based circuits are used as R_2 feedback elements, the increased thermal noise from these circuits may masquerade as increased $1/f$ noise as shown in Fig. 7(b).

If the noise contribution from R_2 is negligible (i.e., $f_{corner} \ll f_H$) and $C_1 \gg C_2, C_{in}$, then the output root mean square (rms) noise voltage of the neural amplifier in Fig. 6 is dominated by the noise from the OTA. Thus, the design of the OTA is crucial to minimize the overall noise of the neural amplifier. We use a cascoded current-mirror OTA as shown in Fig. 8, but other topologies such as a folded cascode amplifier would work as well. The input-referred thermal noise spectral density of this OTA is given by

$$v_{nia}^2(f) = \frac{16kT}{3g_{m1}}\left(1 + 2\frac{g_{m3}}{g_{m1}} + \frac{g_{m7}}{g_{m1}}\right) \qquad (6)$$

where g_{m1} is the transconductance of the input devices M_1 and M_2, g_{m3} represents the transconductance of the nMOS current mirror devices $M_3 - M_6$, and g_{m7} represents the transconductance of the pMOS current mirror devices M_7 and M_8. The biasing transistors (M_{M1} and M_{M2}) and the cascode transistors (M_{C1} and M_{C2}) contribute negligible noise.

As described in [32], the input-referred noise of this OTA can be minimized by ensuring that $g_{m1} \gg g_{m3}, g_{m7}$.

This is accomplished by sizing the transistors so that M_1 and M_2 operate in weak inversion where the ratio of device transconductance to drain current (g_m/I_D) is maximum and $M_3 - M_8$ operate deep in strong inversion where g_m/I_D is greatly reduced [33]–[36].

Perhaps the most critical tradeoff in neural amplifier design is that between power dissipation and input-referred noise. A dimensionless figure of merit that captures the essence of this tradeoff clearly is the noise efficiency factor (NEF), first proposed in [31]

$$\text{NEF} \equiv V_{ni,\text{rms}}\sqrt{\frac{2I_{tot}}{\pi \cdot U_T \cdot 4kT \cdot \text{BW}}} \qquad (7)$$

where I_{tot} is the total amplifier supply current, U_T is the thermal voltage kT/q, BW is the amplifier bandwidth, and $V_{ni,\text{rms}}$ is the amplifier's input-referred rms voltage noise. An amplifier with noise contributed only by the thermal noise of a single ideal bipolar transistor has an NEF = 1; all physical circuits have NEF > 1. In [32], we demonstrated that the NEF of CMOS neural amplifiers is minimized by selectively operating transistors in weak or strong inversion as described above.
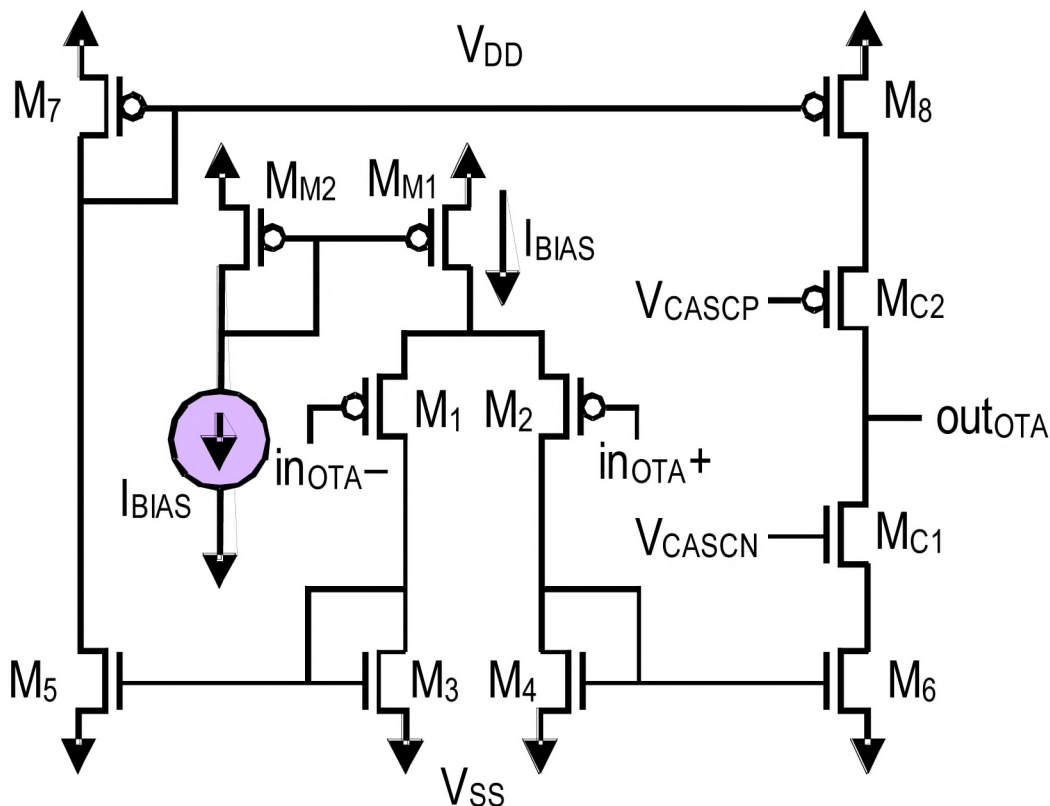

Fig. 8. *Schematic of OTA used in the neural amplifier shown in Fig. 6.*

Note that NEF does not directly account for power dissipation since supply voltage is not present in the expression. However, in modern IC amplifiers, supply voltage only varies by a factor of perhaps five (e.g., 1–5 V), while supply current can vary by many orders of magnitude (e.g., 1 nA–1 A). So NEF is closely correlated with power consumption. In weak inversion, transconductance is linearly proportional to bias current [33]–[36], so from (6) we can see that noise spectral density is inversely proportional to bias current.

C. Noise vs. Layout Area

The rms input-referred noise of the neural amplifier in Fig. 6 is given as

$$\overline{v_{ni}^2} = \left(\frac{C_1 + C_2 + C_{in}}{C_1}\right)^2 \frac{16kT}{3g_{m1}}\left(1 + 2\frac{g_{m3}}{g_{m1}} + \frac{g_{m7}}{g_{m1}}\right)\Delta f. \quad (8)$$

If the amplifier is designed so that $C_1 \gg C_2$, C_{in} and $g_{m1} \gg g_{m3}$, g_{m7}, the input-referred noise is minimized to

$$\overline{v_{ni}^2} \approx \frac{16kT}{3g_{m1}}\Delta f. \quad (9)$$

The equivalent brick-wall bandwidth (Δf) of the amplifier (which has one dominant pole) is given by $(\pi/2)f_H$ if $f_H \gg f_L$ [37]

$$\Delta f = f_H \cdot \frac{\pi}{2} = \frac{1}{2\pi} \cdot \frac{G_m C_2}{C_L C_1} \cdot \frac{\pi}{2} = \frac{G_m}{4C_L A_M}. \quad (10)$$

Therefore, the total rms noise referred to the input of the amplifier is found to be

$$V_{ni,\text{rms}} \approx \sqrt{\frac{4kT}{3C_L A_M}}. \quad (11)$$

From this expression, it is clear that the total amplifier noise integrated across its bandwidth is only a function of temperature, load capacitance C_L, and closed-loop gain A_M. Temperature will be nearly constant in implanted applications, and A_M must be limited to be significantly less than the (poorly controlled) open-loop gain of the amplifier if we want the gain to be well controlled by the ratio of C_1 to C_2. Thus, (11) demonstrates a clear tradeoff between input-referred noise and silicon area, which is often dominated by capacitors.

Since C_2 is usually made as small as possible while staying above parasitic capacitances, A_M is set by sizing C_1. If the area of the amplifier is dominated by capacitors, then the layout area will be proportional to $C_L + 2C_1 + 2C_2$. We can write an expression for amplifier area as

$$\text{Area} = \frac{C_L + 2C_1 + 2C_2}{C'} = \frac{C_L + 2(A_M + 1)C_2}{C'} \quad (12)$$

where C' is the capacitance per unit area of linear capacitors (typically around 1 fF/μm^2 in modern CMOS processes). Solving (11) for C_L, and substituting into (12), we derive an expression for amplifier area as a function of input-referred noise, midband gain, and C_2

$$\text{Area} = \frac{4}{3}\frac{kT}{V_{ni,\text{rms}}^2 A_M C'} + 2(A_M + 1)\frac{C_2}{C'}. \quad (13)$$

For varying midband gain, (13) has a minimum at

$$A_{\text{Mopt}} = \frac{1}{V_{ni,\text{rms}}}\sqrt{\frac{2}{3} \cdot \frac{kT}{C_2}}. \quad (14)$$

Taking $C_2 = 200$ fF as a practical lower limit (to stay above stray parasitic capacitances) and $V_{ni,\text{rms}} = 2$ μVrms, an optimum gain of 60 is found. If the input-referred noise specification is raised to 5 μVrms, the optimum gain to minimize layout area is 24.

Thankfully, these values of A_M fall in a practical range. A gain of perhaps ten or more is desirable to boost the signal above the input-referred noise of successive circuits. A gain of greater than 100 or so is difficult to achieve with high accuracy unless the open-loop gain of the amplifier is extremely high and a large C_1/C_2 ratio is used.

The data from Figs. 2–4 were obtained using a commercial RHA1016 integrated neural amplifier (Intan Technologies, LLC, Salt Lake City, UT) developed using the techniques described here. This amplifier uses a fully differential design throughout to improve common-mode noise rejection, and has an input-referred noise of 2 μVrms [21].

The data from Fig. 5 was obtained using a 100-channel neural recording system with an integrated analog-to-digital converter (ADC) and wireless RF telemetry [8]. Fig. 9 shows a photograph of this chip, which measures 4.7 mm $\times$ 5.9 mm^2 after fabrication in a 0.5-μm 2-poly, 3-metal CMOS process. Each amplifier fits into a layout area of 400 $\times$ 400 μm^2 so that it may be flip-chip bonded to the back of a Utah Electrode Array for complete integration. The amplifiers on this chip were designed for an input-referred noise of 5 μVrms to reduce the required layout area. Since the layout area of neural amplifiers is typically dominated by capacitance and C' for linear capacitors does not scale dramatically in deep submicrometer
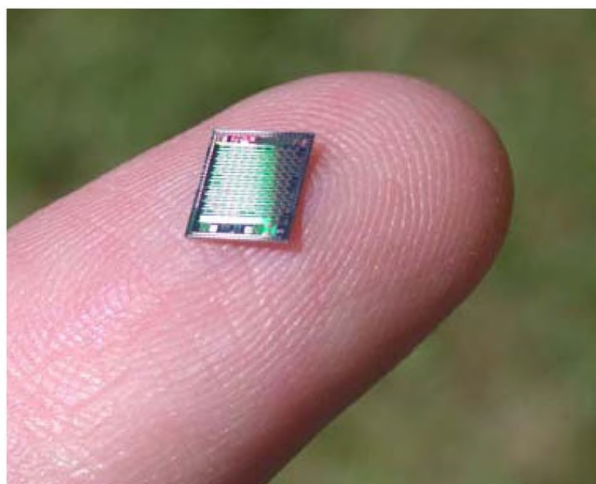
Fig. 9. *Photograph of 100-channel neural recording integrated circuit. The chip measures 4.7 × 5.9 mm² and includes an ADC, spike detectors, and a wireless RF telemetry system [8].*

processes, moving to smaller processes results in modest area savings.

D. Signal Digitization

To permit the robust transmission of neural data across a wireless channel, the amplified neural signals must be converted into a digital representation. Fig. 10 shows a variety of techniques for performing this digitization. In all cases, a preamplifier must be used first to boost the microvolt-level electrode signal and dramatically lower the driving impedance.
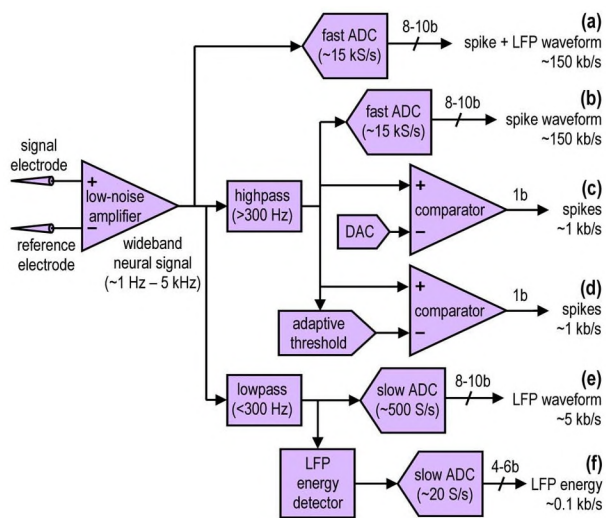

Fig. 10. *Block diagram showing six different techniques for digitizing various aspects of a neural signal. Estimated single-electrode data rates for all six techniques are shown.*

The most straightforward technique for digitizing the neural signal is to pass the wide-band amplified signal through an ADC, as shown in Fig. 10(a). If LFP information is not needed, it can be eliminated with a high-pass filter prior to digitization, as shown in Fig. 10(b). Most commercial neural recording equipment (mounted in large rack-mount cases and supplied by ac wall power) operates in one of these modes using sampling rates of approximately 30 kS/s with resolutions of 12–16 bits (e.g., [38]). These systems thus produce data rates of 36–48 Mb/s from a 100-electrode array. A reduced sampling rate of 15 kS/s and resolution of 10 bits [as shown in Fig. 10(a) and (b)] is sufficient for most scientific and clinical applications, but this still yields a data stream of 15 Mb/s for 100 electrodes.

Transmitting data at these rates over a wireless transcutaneous link is difficult or impossible to achieve in small, implanted systems that are severely power constrained. RF links are handicapped by the fact that the tissue absorption of electromagnetic radiation follows an f^2 trend. Thus the FCC-approved Medical Implant Communication System band was allocated in the relatively low-frequency band of 402–405 MHz and consists of ten channels each with only 300 kHz of bandwidth. Infrared light penetrates bone and tissue with little attenuation, but optical links require a fair amount of power. Recently, transcutaneous data transfer at 40 Mb/s was demonstrated, but the power consumption of the transmitter was 120 mW [39]. Clearly, implantable high-channel-count neural recording devices will likely require circuitry for on-chip data compression.

IV. ADAPTIVE NEURAL SPIKE DETECTION

When one considers the nature of typical neural signals, it is clear that far too much information is being transmitted in Fig. 10(a) and (b). For many scientific and neuroprosthetic applications, the only relevant information is the presence and timing of action potentials to an accuracy of approximately 1 ms. Detecting the presence or absence of a spike every 1 ms produces a 100 kb/s data stream for a 100-electrode system. This data rate could be reduced even further by the use of an asynchronous protocol that transmits data only when spikes appear (e.g., [40]). Cortical neurons exhibit firing rates around 10 Hz, and in a 100-channel system, the "address" of each spike can be encoded in a 7-bit number representing its electrode of origin. If we transmit an address only when a spike occurs, our data rate can be reduced to an average of 7 kb/s. A system described in [7] sends the address of spikes and uses a 5-bit ADC to transmit the amplitude of the spike as well.

The remaining problem is how to perform this data reduction from noisy analog waveform to identified spikes in a small, low-power device. The amplitude of spikes recorded extracellularly can vary widely from one electrode to the next depending on the relative position

and orientation of the recording site and the cell carrying the impulse. Additionally, background noise caused by distant neural activity, electrode noise, and electronic noise in the preamplifier can vary with time, temperature, and electrode position.

A straightforward technique shown in Fig. 10(c) is to set a spike-detection threshold manually using a digital-to-analog converter (DAC). This technique has been implemented in a 100-channel wireless neural recording system that transmits one user-selectable channel using the technique shown in Fig. 10(b), while spike data from all 100 channels are transmitted using manual spike thresholding [8]. The ADC allows the user to observe the waveform from each electrode in turn and set an appropriate spike-detection threshold using local DACs.

In the future, it would be advantageous for the implanted device to autonomously set spike detection thresholds for each channel. In pursuit of this goal, we developed a small mixed-signal circuit to adaptively set spike detection thresholds above a background noise level.

A. Adaptive Spike Detection Algorithm

The goal of our spike-detection algorithm (first described in [41]) is to adaptively set a detection threshold that is low enough to capture action potentials but high enough to reject occasional peaks in the background noise. We assume Gaussian background noise having a mean of zero. (Measured background noise from actual neural recordings has a roughly Gaussian distribution, though the tails are slightly wider [42].) Therefore the noise is entirely described by its rms value, which is equivalent to its standard deviation σ.

If we can measure the rms level σ of the background noise, we can set a threshold to some multiple of σ and reject all but a vanishingly small fraction of the background noise. For example, with a threshold of 5σ, the probability of Gaussian noise triggering the spike detector is approximately 3×10^{-7}.

To develop a simple method for measuring σ, we observe that if a threshold is set at σ, the probability of Gaussian noise exceeding this threshold is 0.159. Fig. 11(a) shows a noise waveform and a threshold level of 1σ. After comparing the noise with this threshold, we get a digital waveform having a duty cycle (i.e., the fraction of time the waveform is high) of 0.159 [see Fig. 11(b)]. The duty cycle is proportional to the dc level of this digital waveform, and we can use this signal as feedback to servo a reference voltage to the 1σ level of the waveform.

Fig. 12 shows a block diagram of the proposed adaptive spike detection algorithm. Comparator A is used in a feedback loop (with a gain of K) that servos the duty cycle of its output to 0.159, thus setting $V_{1\sigma}$ to the rms level of the input waveform. This voltage is then amplified by a constant N typically having a value of five or greater. The resulting voltage $V_{N\sigma}$ is used as the threshold level for Comparator B. Thus, the circuit performs spike detection using a specified multiple of the background noise rms value.
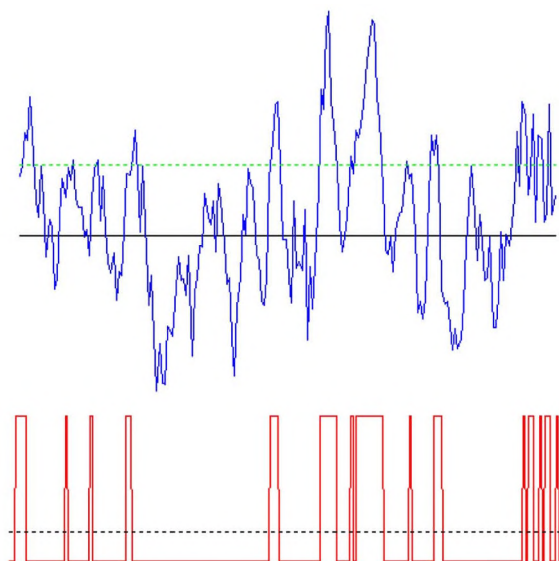

Fig. 11. *If Gaussian noise (top) is passed through a comparator having a threshold set to the rms value of the noise (dotted line), the resulting digital signal (bottom) made up of zeros and ones has a dc level of 0.159 (dotted line).*

The presence of spikes in the waveform will lead to errors in our estimate of the noise rms level since the $V_{1\sigma}$ feedback loop does not distinguish between spikes and background noise. However, if the spikes are approximately ac balanced (as most biphasic spike waveforms are) and occur relatively infrequently, they should have little effect on the rms noise estimate. With cortical firing rates around 10 Hz and action potentials approximately 1 ms in duration, spikes are present only about 1% of the time.

B. Circuit Design and Implementation

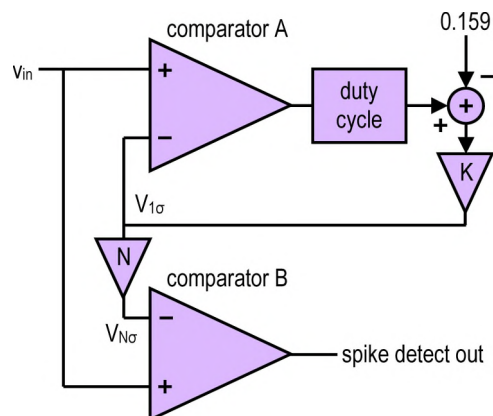We implemented the adaptive spike detection algorithm in a CMOS integrated circuit with the goal of

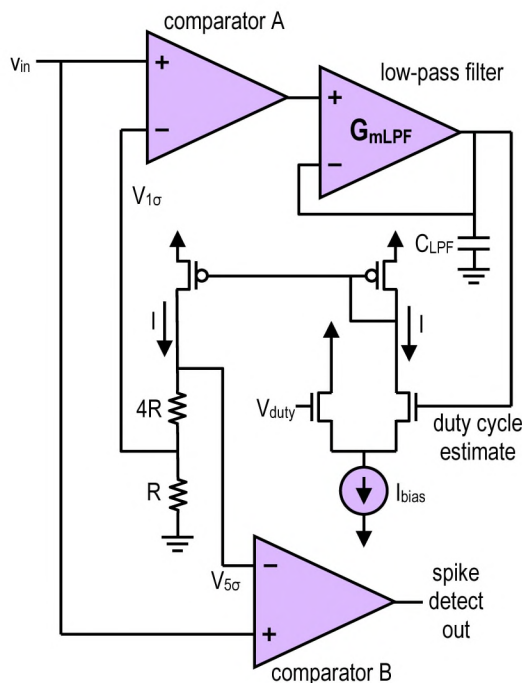**Fig. 12.** *Block diagram of the adaptive spike detection algorithm.*

Fig. 13. *Schematic of the adaptive spike detection circuit.*

minimizing power consumption and chip area. The circuit was completely integrated in a 1.5-μm 2-metal 2-poly CMOS process, using no off-chip components.

A schematic of the adaptive spike detection circuit is shown in Fig. 13. Comparators A and B are implemented using standard regenerative latch-and-hold topologies [37]. The duty cycle of Comparator A is calculated using an OTA to realize a $G_m - C$ low-pass filter. By biasing this OTA in the subthreshold region, cutoff frequencies below 1 Hz may be achieved [34]. The high-frequency oscillations of the digital waveform are attenuated leaving only the dc level, which is proportional to the duty cycle of the waveform. By taking a "running average" of the duty cycle using this leaky integrator, the circuit is able to adapt to time-varying levels of background noise. The time constant of this filter sets the adaptation time constant.

An nMOS differential pair is used to compare the output of the low-pass filter to the reference voltage $V_{\mathrm{duty}} = 0.159 V_{\mathrm{DD}}$, which corresponds to a low-pass filter output indicating Comparator A is operating at the 1σ threshold level. Current from one leg of the differential pair is mirrored using a pMOS current mirror and driven into two resistors in series. These resistors convert the current into two voltages: $V_{1\sigma} = IR$ and $V_{5\sigma} = 5IR$. To save chip area, these resistors were implemented as nMOS transistors operating in the deep triode (linear) region. By sizing the transistors appropriately, a drain-to-source resistance R of approximately 10 kΩ was obtained.

C. Circuit Testing

We tested the adaptive spike detector using a synthetic waveform programmed into an arbitrary waveform generator (Agilent 33120A). The test waveform consisted of three typical extracellular action potentials embedded in a background of Gaussian noise and represented the output from a preamplifier in a neural recording system. The first 10 ms of the test waveform is shown as the input waveform in Fig. 14. The rest of the waveform consisted only of noise. The waveform was 80 ms in length and was played in a loop so the burst of three spikes appeared periodically at a rate of 12.5 Hz.

We applied this waveform to the input of the adaptive spike detector. The amplitude of the waveform was set so that the largest spike had an amplitude of 70 mV and the background noise had an rms value of 5.5 mV. (Assuming a preamplifier with a gain of 60 dB, this corresponds to a spike amplitude of 70 μV and a noise rms value of 5.5 μV at the electrode.) Fig. 14 shows the input waveform along with the value of $V_{5\sigma}$ and the output of Comparator B. The adaptive spike detector successfully sets the threshold to an appropriate level to detect spikes but reject noise.

The amplitude of the input waveform (largest spike) was varied from 23 to 116 mV (and the rms noise level varied from 1.8 to 9.2 mV). The circuit functioned correctly as the amplitude of the background noise changed by a factor of five. Fig. 15 shows the response of the circuit to a waveform containing only noise and no spikes. The algorithm succeeds in rejecting the noise completely despite occasional peaks in the Gaussian waveform. (In Figs. 14 and 15, the 0–5 V digital output voltage is scaled down for clarity.)

The circuit consumed 0.094 mm^2 of chip area in a 1.5-μm process, and its power consumption was 57 μW when run from a 5 V power supply. The two comparators
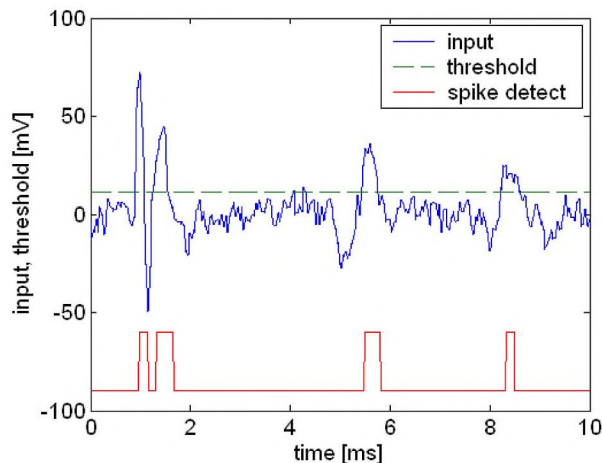

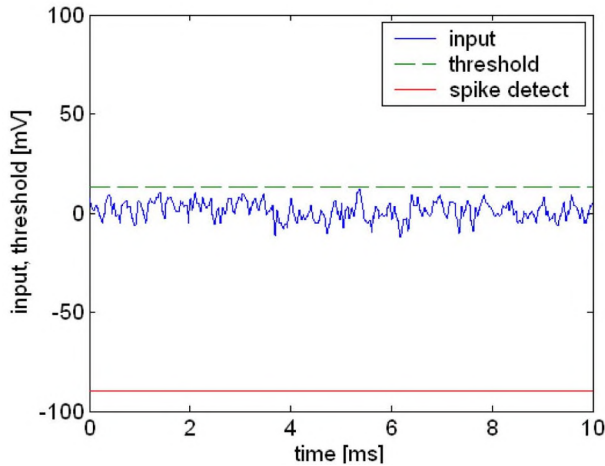**Fig. 14.** *Measured output of adaptive spike detection chip for input amplitude of 70 mV.*

Fig. 15. *Output of adaptive spike detection chip with input of bandlimited Gaussian noise only.*

Fig. 17. *Schematic of stagger-tuned fourth-order bandpass filter tuned for a one-octave passband of 20–40 Hz.*

consumed 91% of this power, so future work will focus on reducing their power dissipation.

V. LOCAL FIELD POTENTIAL ENERGY DETECTION

A. LFP Energy Detection Algorithm

As discussed above, local field potential signals also contain useful information. Since LFPs occur below 200 Hz, they may be digitized using a slow ADC, as shown in Fig. 10(e). Most analysis of LFP information involves tracking the energy of LFP signals in a certain narrow range of frequencies (e.g., 20–40 Hz) [28]. If this analysis is performed on chip before transmission, a slower ADC with lower resolution may be used to capture the changes in energy in particular frequency bands of interest, as shown in Fig. 10(f).

Fig. 16 presents a diagram of our proposed algorithm for on-chip LFP energy detection [43]. To measure the energy in an LFP signal, we first use a bandpass filter to isolate the frequencies of interest. The filtered signal is then squared, and the squared signal is passed to a leaky integrator that effectively calculates a running average of signal energy within the passband of the system. The time constant of the leaky integrator determines the "persistence of memory" in this running average.
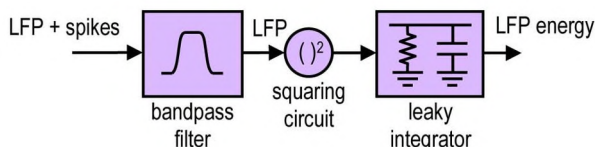

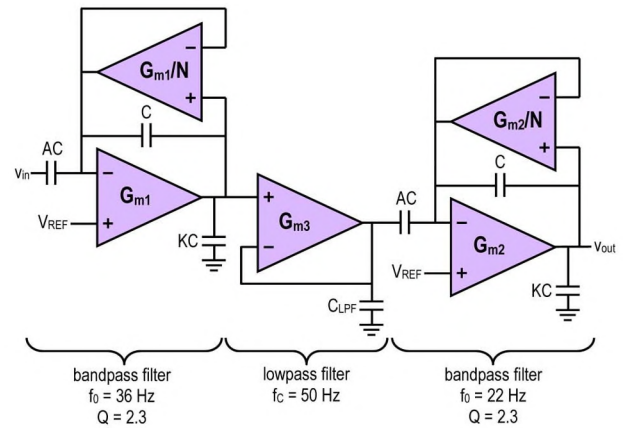**Fig. 16.** *Block diagram of local field potential energy detector.*

B. Circuit Design and Implementation

We designed the LFP energy detector for implementation in a CMOS integrated circuit with the goal of minimizing power consumption and chip area. All transistors were operated in the subthreshold region to conserve power and achieve large time constants with integrated capacitors [34].

A schematic of the bandpass filter is shown in Fig. 17. The circuit uses five OTAs and seven capacitors to realize a fourth-order bandpass filter. We use the standard technique of stagger tuning where two second-order bandpass filters are cascaded and tuned to slightly different frequencies. The combination of their transfer functions results in a wider, flatter passband than in either single filter.

Each second-order bandpass filter consists of two OTAs and three capacitors. The OTAs use diode-connected transistors for source degeneration to extend their linear range. The transfer function of each filter is given by

$$\frac{v_{\text{OUT}}(s)}{v_{\text{in}}(s)} = -\frac{AN}{\beta} \cdot \frac{\left(1 - \frac{s}{\omega_0 \beta}\right)}{\frac{s}{\omega_0} + \frac{1}{Q} + \frac{\omega_0}{s}} \qquad (15)$$

where

$$\omega_0 = 2\pi f_0 = \frac{G_m}{\beta C} \qquad (16)$$

$$\beta = \sqrt{N(A+1)(K+1) - N}$$
$$\approx \sqrt{NAK} \quad \text{if } A, K \gg 1 \qquad (17)$$

$$Q = \frac{\beta}{(K+N)}. \qquad (18)$$

The output signal v_{OUT} is centered around V_{REF}, a dc voltage that was set to 1.2 V. We sized the capacitors in our

circuit to give $A = 20$ and $K = 5$ (with $C = 0.1$ pF). The transconductance of the lower OTA was set by adjusting its bias current I_B

$$G_m = \frac{\kappa}{(\kappa + 1)} \cdot \frac{I_B}{2U_T} \qquad (19)$$

where κ is the weak inversion slope (approximately 0.7) and U_T is the thermal voltage kT/q (approximately 27 mV at body temperature). We set the bias current in the upper OTA five times smaller to achieve $N = 5$.

As we see from (15), this circuit acts as an ac-coupled second-order bandpass filter centered at f_0 with a quality factor Q set to 2.5 by capacitor and bias current ratios. The circuit also has a zero at βf_0, but since $\beta = 25$ in our circuit, the zero takes effect outside the passband and thus has little practical effect on the filter. We added a single-pole low-pass filter to the circuit (center OTA in Fig. 17) to ensure that the gain decreases at high frequencies despite the presence of this zero.

We used bias currents of $I_B = 43$ and 24 pA in the lower OTAs and 8.8 and 4.8 pA in the upper OTAs to center the passbands of the first and second bandpass filters near 36 and 22 Hz, respectively. The combined system has a passband from 20 to 40 Hz with a two-pole (40 dB/dec) rolloff at each end.

Fig. 18 shows a schematic of the circuit that squares the signals from the bandpass filter and produces a running average of signal energy. We use a subthreshold CMOS Gilbert multiplier to square the input voltage v_{LFP} provided by the bandpass filter. The output of the Gilbert multiplier is a bidirectional current (due to the nMOS current mirror) that is forced into an OTA configured as a $G_m - C$ filter (G_{m-int} and C_{int} in Fig. 18). The time constant of this filter was set to 100 ms, a reasonable value for the integration of LFP signal energy in the 20–40 Hz range.
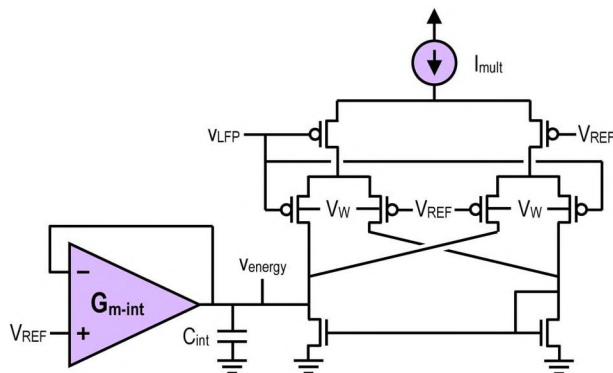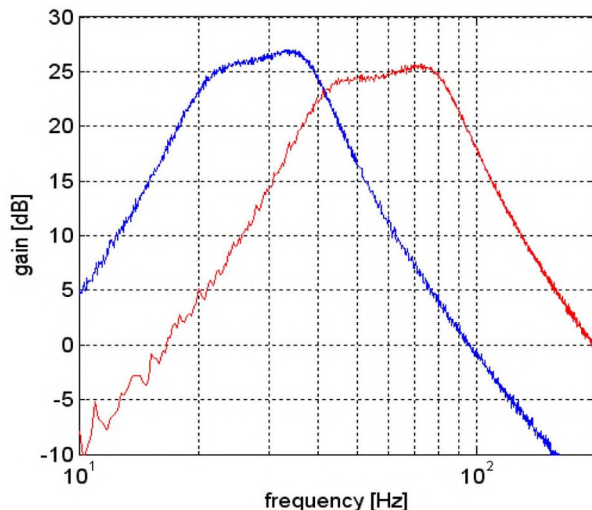

Fig. 19. *Measured filter passbands for two different biasing conditions. Left curve: 21–40 Hz. Right curve: 39–88 Hz.*

C. Circuit Testing

The LFP energy detection circuit was fabricated in a 1.5-μm 2-metal 2-poly CMOS process. The circuit was completely integrated, using no off-chip components. The circuit contained eight capacitors with a total capacitance of 8.2 pF. The LFP detector layout consumed a chip area of 586×79 μm^2 (0.046 mm^2). Capacitors consumed approximately 50% of the layout area.

The bandpass filter was biased to have a pass band of approximately 20–40 Hz. A network analyzer was used to measure the filter transfer function (see Fig. 19). As expected, the filter exhibits a 40 dB/dec rolloff outside of the passband of 21–40 Hz. The filter has a midband gain of approximately 26 dB. To demonstrate the tunable nature of the circuit, the transfer function was also measured with the bias currents doubled, yielding a passband of 39–88 Hz (right curve in Fig. 19). The passband was set back to 21–40 Hz for all later experiments.

As a test of overall circuit performance, we used prerecorded broadband neural data from a Utah Electrode Array chronically implanted in monkey premotor and motor cortex (0.3 Hz–7.5 kHz; 30 kS/s, 12-bit samples).[1] We programmed a two-second segment of neural data from one of the 100 electrodes into a function generator [see Fig. 20(a)] and played it to the chip, amplified by a factor of 60 dB. (In an actual neural recording system, the LFP energy detection circuit would be connected to the recording electrode via a low-noise preamplifier such as the one described in Section III.)

The chip's response was compared to a Thomson multitaper spectral analysis performed on a PC using MATLAB as shown at the bottom of Fig. 20. An arbitrary

Fig. 18. *Schematic of circuit to square v_{LFP} signal and perform a leaky integration to produce a running average of LFP energy (v_{energy}).*

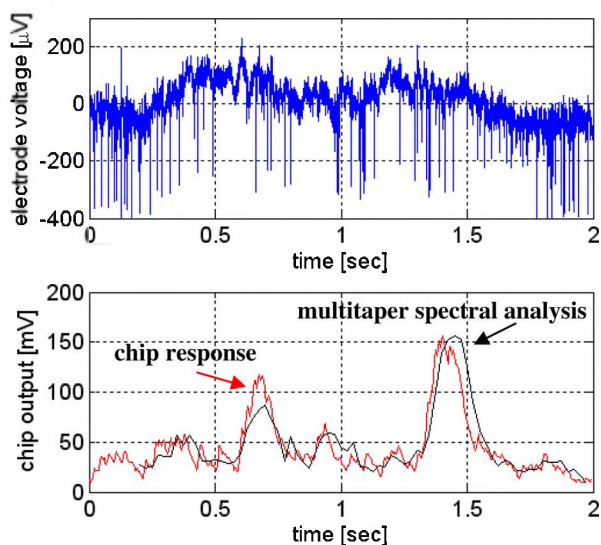[1]Data courtesy of K. Shenoy, Stanford University.

Fig. 20. *(Top) Neural waveform containing spikes and local field potentials. (Bottom) Response of the LFP energy detection circuit (jagged line) compared to the response of a Thomson multitaper spectral analysis run offline on a PC (smooth, dark line) [41]. Both techniques measured LFP energy in the 20–40 Hz band.*

scaling factor was used to align the two curves' amplitudes. The circuit output closely follows the multitaper analysis results, which is widely considered to be the optimal spectral analysis method for LFP analyses [28]. The circuit consumes only 5 nW of power when run from a 5 V power supply. The circuit accurately detects LFP energy levels within a specified pass band. The small size and low power operation of this circuit make it compatible with fully implanted multielectrode applications.

VI. CONCLUSION

In this paper, we have presented integrated circuits and design techniques for amplifying weak potentials produced in the brain and distilling relevant information from these signals. When designing implantable devices, power and size are paramount, so a firm understanding of circuit design tradeoffs is an essential tool. Analog-to-digital converters consume large amounts of silicon area and, if operated

at high sampling rates, can consume significant power; they should be used sparingly. Analog computation such as adaptive spike thresholding and LFP energy calculation can provide practical alternatives to the traditional approach of A/D conversion and digital signal processing.

The spike detection algorithms presented here have not addressed the problem of "spike sorting" if multiple neurons are observed from a single electrode (e.g., Fig. 5). Other neural recording systems reported in the literature digitize and transmit spike peak amplitudes [7] or peak and trough amplitudes [44]. This would enable a simple clustering algorithm to discriminate between spikes with significant size differences.

Many software-based spike sorting techniques developed during the past few decades are quite computationally intensive (e.g., [45]), but some purely digital sorting algorithms are being adapted for power-efficient integrated implementation [46]. Any on-chip spike sorting algorithm must deal with the fact that spike amplitudes and shapes change with time as the electrode array moves slightly in the tissue [47].

Although it is tempting to imagine increasingly elaborate circuits for on-chip data processing, size and power limitations dictate that we perform only the bare minimum operations in implanted neural recording devices; all other processing can be implemented outside the body where heat, volume, and mass are less of a concern. This implies that an implanted system should only amplify neural signals, perform basic filtering, and find the most efficient way to transmit relevant data through the "transcutaneous bottleneck." Thankfully, today's advances in circuit miniaturization enable increasingly sophisticated methods of extracting relevant signals from the neural cacophony and threading just the right information through the bottle. ∎

REFERENCES

[1] J. Pine, "Recording action potentials from cultured neurons with extracellular microcircuit electrodes," *J. Neurosci. Meth.*, vol. 2, pp. 19–31, 1980.

[2] A. C. Hoogerwerf and K. D. Wise, "A three-dimensional microelectrode array for chronic neural recording," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 1136–1146, Dec. 1994.

[3] A. L. Owens, T. J. Denison, H. Versnel, M. Rebbert, M. Peckerar, and S. A. Shamma, "Multi-electrode array for measuring evoked potentials from the surface of ferret

primary auditory cortex," *J. Neurosci. Meth.*, vol. 58, pp. 209–220, 1995.

[4] C. T. Nordhausen, E. M. Maynard, and R. A. Normann, "Single unit recording capabilities of a 100 microelectrode array," *Brain Res.*, vol. 726, pp. 129–140, 1996.

[5] K. Najafi and K. D. Wise, "An implantable multielectrode array with on-chip signal processing," *IEEE J. Solid-State Circuits*, vol. SC-21, pp. 1035–1044, Dec. 1986.

[6] K. D. Wise, D. J. Anderson, J. F. Hetke, D. R. Kipke, and K. Najafi, "Wireless implantable microsystems: High-density electronic interfaces to the nervous system," *Proc. IEEE*, vol. 92, pp. 76–97, Jan. 2004.

[7] R. H. Olsson, III, and K. D. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, vol. 40, pp. 2796–2804, Dec. 2005.

[8] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuits*, vol. 42, pp. 123–133, Jan. 2007.

[9] T. M. Seese, H. Harasaki, G. M. Saidel, and C. R. Davies, "Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of muscle tissue to

chronic heating," *Lab. Invest.*, vol. 78, no. 12, pp. 1553–1562, 1998.

[10] J. C. LaManna, K. A. McCracken, M. Patil, and O. J. Prohaska, "Stimulus-activated changes in brain tissue temperature in the anesthetized rat," *Metabolic Brain Disease*, vol. 4, pp. 225–237, 1989.

[11] A. Jackson, J. Mavoori, and E. E. Fetz, "Long-term motor cortex plasticity induced by an electronic neural implant," *Nature*, vol. 444, pp. 56–60, 2006.

[12] J. K. Chapin, K. A. Moxon, R. S. Markowitz, and M. A. L. Nicolelis, "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex," *Nature Neurosci.*, vol. 2, pp. 664–670, 1999.

[13] J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K. Chapin, J. Kim, S. J. Biggs, M. A. Srinivasan, and M. A. L. Nicolelis, "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates," *Nature*, vol. 408, pp. 361–365, 2000.

[14] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, and J. P. Donoghue, "Instant neural control of a movement signal," *Nature*, vol. 416, pp. 141–142, 2002.

[15] D. M. Taylor, S. I. H. Tillery, and A. B. Schwartz, "Direct cortical control of 3-D neuroprosthetic devices," *Science*, vol. 296, pp. 1829–1832, 2002.

[16] P. R. Kennedy, R. A. E. Bakay, M. M. Moore, K. Adams, and J. Goldthwaite, "Direct control of a computer from the human central nervous system," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 198–202, Jun. 2000.

[17] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, pp. 164–171, 2006.

[18] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, 4th ed. Boston, MA: McGraw-Hill, 2000.

[19] R. C. Gesteland, B. Howland, J. Y. Lettvin, W. H. Pitts, "Comments on microelectrodes," in *Proc. IRE*, 1959, vol. 47, pp. 1856–1862.

[20] C. D. Ferris, *Introduction to Bioinstrumentation*. Clifton, NJ: Humana, 1978.

[21] R. R. Harrison, "A versatile integrated circuit for the acquisition of biopotentials," in *IEEE 2007 Custom Integr. Circuits Conf. Dig. Tech. Papers (CICC 2007)*, San Jose, CA, 2007, pp. 115–122.

[22] A. C. Metting van Rijn, A. Peper, and C. A. Grimbergen, "High-quality recording of bioelectric events," *Med. Biol. Eng. Comput.*, vol. 29, pp. 1035–1044, 1986.

[23] V. N. Murthy and E. E. Fetz, "Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys," in *Proc. Nat. Acad. Sci. USA*, 1992, vol. 89, pp. 5670–5674.

[24] J. P. Donoghue, J. N. Sanes, N. G. Hatsopoulos, and G. Gaal, "Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements," *J. Neurophysiol.*, vol. 79, pp. 159–173, 1998.

[25] K. V. Shenoy, M. M. Churchland, G. Santhanam, B. M. Yu, and S. I. Ryu, "Influence of movement speed on plan activity in monkey pre-motor cortex and

implications for high-performance neural prosthetic system design," in *Proc. 2003 Int. Conf. IEEE Eng. Med. Biol. Soc.*, Cancún, Mexico, 2003, pp. 1897–1900.

[26] C. Mehring, J. Rickert, E. Vaadia, S. Cardoso de Oliveira, A. Aertsen, and S. Rotter, "Inference of hand movements from local field potentials in monkey motor cortex," *Nature Neurosci.*, vol. 6, pp. 1253–1254, 2003.

[27] H. Scherberger, M. R. Jarvis, and R. A. Andersen, "Cortical local field potential encodes movement intentions in the posterior parietal cortex," *Neuron*, vol. 46, pp. 347–354, 2005.

[28] B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, and R. A. Andersen, "Temporal structure in neuronal activity during working memory in macaque parietal cortex," *Nature Neurosci.*, vol. 5, pp. 805–811, 2002.

[29] S. Kim, R. A. Normann, R. Harrison, and F. Solzbacher, "Preliminary study of thermal impacts of a microelectrode array implanted in the brain," in *Proc. 2006 Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, 2006, pp. 2986–2989.

[30] S. Kim, P. Tathireddy, R. Normann, and F. Solzbacher, "Vitro and in vivo study of temperature increases in the brain due to a neural implant," in *Proc. 3rd Int. IEEE EMBS Conf. Neural Eng.*, Kohala Coast, HI, 2007.

[31] 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. SC-22, pp. 1163–1168, Dec. 1987.

[32] R. R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE J. Solid-State Circuits*, vol. 38, pp. 958–965, Jun. 2003.

[33] E. A. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation," *IEEE J. Solid-State Circuits*, vol. 12, pp. 224–231, Jun. 1977.

[34] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA: Addison-Wesley, 1989.

[35] C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integr. Circuits Signal Process.*, vol. 8, pp. 83–114, 1995.

[36] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. Boston, MA: McGraw-Hill, 1998.

[37] D. A. Johns and K. Martin, *Analog Integrated Circuit Design*. New York: Wiley, 1997.

[38] K. S. Guillory and R. A. Normann, "A 100-channel system for real time detection and storage of extracellular spike waveforms," *J. Neurosci. Meth.*, vol. 91, pp. 21–29, 1999.

[39] K. S. Guillory, A. K. Misener, and A. Pungor, "Hybrid RF/IR transcutaneous telemetry for power and high-bandwidth data," in *Proc. 2004 Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC 2004)*, San Francisco, CA, 2004, pp. 4338–4340.

[40] K. A. Boahen, "Point-to-point connectivity between neuromorphic chips using address-events," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 416–434, May 2000.

[41] R. R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," in *Proc. 2003 Int. Conf. IEEE Eng. Med. Biol. Soc.*, Cancún, Mexico, 2003, pp. 3325–3328.

[42] P. T. Watkins, G. Santhanam, K. V. Shenoy, and R. R. Harrison, "Validation of adaptive threshold spike detector for neural recording," in *Proc. 2004 Int. Conf. IEEE Eng. Med. Biol. Soc.*, San Francisco, CA, 2004, pp. 4079–4082.

[43] R. R. Harrison, G. Santhanam, and K. V. Shenoy, "Local field potential measurement with low-power analog integrated circuit," in *Proc. 2004 Int. Conf. IEEE Eng. Med. Biol. Soc.*, San Francisco, CA, 2004, pp. 4067–4070.

[44] T. K. Horiuchi, T. Swindell, D. Sander, and P. Abshire, "A low-power CMOS neural amplifier with amplitude measurements for spike sorting," in *Proc. 2004 IEEE Int. Symp. Circuits Syst.*, Vancouver, BC, Canada, 2004, vol. 4, pp. 29–32.

[45] Z. S. Zumsteg, C. Kemere, S. O'Driscoll, G. Santhanam, R. E. Ahmed, K. V. Shenoy, and T. H. Meng, "Power feasability of implantable digital spike-sorting circuits for neural prosthetic systems," *IEEE Trans. Neural Syst. Rehab.*, vol. 13, pp. 272–279, Sep. 2005.

[46] A. F. Atiya, "Recognition of multiunit neural signals," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 723–729, Jul. 1992.

[47] G. Santhanam, M. D. Linderman, V. Gilja, A. Afshar, S. I. Ryu, T. H. Meng, and K. V. Shenoy, "HermesB: A continuous neural recording system for freely behaving primates," *IEEE Trans. Biomed. Eng.*, pp. 2037–2050, Nov. 2007.

ABOUT THE AUTHOR

Reid R. Harrison (Member, IEEE) was born in DeFuniak Springs, Florida. He received the B.S. degree in electrical engineering from the University of Florida, Gainesville, in 1994 and the Ph.D. degree from the California Institute of Technology, Pasadena, in 2000.

He joined the University of Utah, Salt Lake City, in 2000, where he is now an Associate Professor of electrical and computer engineering and an Adjunct Associate Professor of bioengineering. He has more than 50 refereed publications since 1999 in the fields of low-power analog and mixed-signal CMOS circuit design, integrated electronics for neural interfaces and other biomedical devices, and hardware for biologically inspired computational systems.

Dr. Harrison organized the 2001 IEEE SSCTC Workshop on Low-Power Circuits in Arlington, VA. He received the National Science Foundation CAREER Award in 2002 and the Jack Raper Award for Outstanding Technology Directions Paper from the International Solid-State Circuits Conference in 2006. He is a member of the Technical Program Committee of ISSCC and the IEEE International Symposium on Circuits and Systems.