Title
A Digitally Assisted Multiplexed Neural Recording System With Dynamic Electrode Offset Cancellation via an LMS Interference-Canceling Filter

Permalink
https://escholarship.org/uc/item/61z008bz

Journal
IEEE Journal of Solid-State Circuits, 57(3)

ISSN
0018-9200

Authors
Fathy, Nader Sherif Kassem
Huang, Jiannan
Mercier, Patrick P

Publication Date
2022-03-01

DOI
10.1109/jssc.2021.3116021

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at https://creativecommons.org/licenses/by-nc-nd/4.0/

Peer reviewed
A Digitally Assisted Multiplexed Neural Recording System With Dynamic Electrode Offset Cancellation via an LMS Interference-Canceling Filter

Nader Sherif Kassem Fathy, Graduate Student Member, IEEE, Jiannan Huang, Member, IEEE, and Patrick P. Mercier, Senior Member, IEEE

Abstract—This article presents a low-power (LP) area-efficient implantable neural recording system that supports high-density neural implant (HDNI) applications. The system uses a time-division multiple access method to record from 16-neural electrodes simultaneously. A least mean squares (LMS) algorithm is used to cancel the slowly varying electrode offsets from all channels simultaneously by using a single-tap digital adaptive filter (AF). The presented technique is fabricated in 65-nm CMOS technology and achieves a per-channel area of 0.00248 mm²; 68% of which is digital circuitry (and is thus scalable with technology). The overall system consumes 3.38 µW per channel while achieving 2.6 µVrms of input referred noise (IRN) in 10 kHz of bandwidth. The proposed system has a noise efficiency factor (NEF) of 1.83 and is fully integrated on-chip.

Index Terms—Digitally assisted least mean square (LMS) filter, electrocorticography (ECoG), microelectronic implants brain–machine interface (BMI), neural recording, time-division multiple access (TDMA).

I. INTRODUCTION

THE development of high-density microelectronic neural recording systems is becoming vital to study the complicated dynamics of the human brain. For example, modern research on brain–machine interfaces (BMIs) has succeeded in decoding neural signals from the brain’s cerebral cortex and has translated it into useful data for use in prosthetic applications [1]. It is feasible to restore the movement of a limb by recording from 10 000 neurons simultaneously, meanwhile 100 000 real-time neural recordings are predicted to be able to restore movement of the entire body [2]. It is immensely challenging to integrate this large number of channels on fully integrated implantable system-on-chips (SoCs).

The main challenge in designing a high-density neural implant (HDNI) system is the requirement of small channel area, since area tends to trade off with other important parameters such as noise, power, and offset blocking capabilities [3]. For example, neural local field potential (LFP) signal contents of (1–300 Hz) lie in the flicker noise band, while thermal noise affects the neural action potential (AP) signal of frequencies (0.3–10 kHz) [4]. The flicker noise issue is traditionally resolved by employing a large-area input differential pair amplifier [5] or by using chopper-stabilized instrumentation amplifier (IA), the latter of which requires a wide-bandwidth...
(BW) amplifier and, importantly, additional feedback loops to cancel up-converted electrode offsets and chopper ripples; both approaches require non-zero additional area [6], [7]. On the other hand, the thermal noise issue is solved by increasing the transconductance of the input differential pair of the neural amplifier, which requires high power consumption and/or large area utilization [5]. According to [8], the analog front-end (AFE) module is still far behind in terms of development in the pursuit of creating a true HDNI system.

Fig. 1 illustrates the overall concept of a batteryless neural recording system, powered wirelessly from outside the skull by using coils as demonstrated in [3]. The multiplexed digital signals are transmitted outside the brain by an integrated antenna and then demultiplexed on the receiver side outside the brain as implemented in [8].

Fig. 1(a) shows a typical neural recording system, where each electrode has its own dedicated analog interface. Each pixel consists of a neural amplifier and a digital to analog converter (DAC) in feedback to remove: 1) electrode offset voltage (EOV); 2) amplifier input-referred offset; and 3) other undesired signals such as motion or stimulation artifacts if present. Dynamic electrochemical reactions at the interface of the neural electrodes and tissue cause an EOV between electrodes and the analog interface that can reach magnitudes of approximately ±50 mV [4], [6], [9], [10]. In addition, if an amplifier is used in an open-loop configuration to save area by eliminating passive feedback components, an input-referred offset caused by process mismatch can reach a few millivolts. Such high offsets can easily saturate high-gain differential neural amplifiers. While ac coupling capacitors can potentially block EOVs, they do not help when choppers are placed before the coupling capacitors. On the other hand, chopping after the coupling capacitors forms a switched-capacitor (SC) parasitic resistor with the input devices parasitic capacitors. When input referred, the SC parasitic resistance noise is seen to have a $1/f^2$ shape, which requires physically large capacitors to attenuate it [11].

To tackle the issue of EOV in IAs, a dc-servo loop with a low corner frequency (<1 Hz) is usually used [6], [7], [9], [12]. Other approaches such as [13] use a track and zoom analog to digital converter (ADC) to remove the dc-offsets, or in [14] and [15], a voltage-controlled oscillator (VCO)-based and successive approximation register (SAR) ADC with digital feedback loop are used for dc-offset removal, respectively. Unfortunately, the addition of analog or digital servo loops requires a non-trivial overhead to the system. The delta modulation presented in [16] and [17] offer a solution to remove the EOV, though as will be discussed shortly, hardware sharing to reduce per-channel area is not easily possible with this technique. As a result, the traditional in-pixel solution with EOV cancellation mechanism typically occupies large channel area (∼0.01 mm²) and consumes high power (∼3–5 μW) due to the use of a dedicated amplifier and DAC for every channel [4], [14].

On the other hand, Fig. 1(b) shows a solution which shares the analog interface among multiple electrodes. Done correctly, this hardware sharing approach can reduce the per-channel area, making it a promising candidate to solve the AFE bottleneck described in [8]. For example, a time-division multiple access (TDMA) method has been recently used in [18] and [19], resulting in per-channel areas of (0.004, 0.0023 mm²), respectively. However, multiple coarse and fine-tuning DACs, memory modules, off-chip digital filters, binary-search algorithms, and processors are used in the system to restore the distorted signal recorded by delta-encoded method, which would ultimately occupy significant area in a final implemented solution. Unfortunately, in a TDMA system, the EOV removal becomes challenging since, assuming random EOV for every channel, the input multiplexed signal to the AFE translates the slowly varying EOV into a higher frequency component that is equal to the multiplexer sampling speed. This forces all channel multiplexed EOV to pass through the AFE input with full amplitude even if an ac-capacitor or a regular dc-servo loop is used. Accordingly, the multiplexed EOVs saturate the AFE.

Importantly, it should be noted that EOV behavior is not constant: EOVs tend to have some low-frequency content from dc to 0.2 Hz [20]. As a result, static dc-cancellation approaches as in [4] and [18] are not robust in real-world conditions unless the digital cancellation calibration routine is rerun in the foreground periodically. In addition, in case of using a binary-search algorithm to remove the slowly varying EOV, the cutoff frequency between LFP and AP becomes unknown, and the offset is removed abruptly, which can lead to distortion in the acquired neural signal.

This article presents a TDMA-based least mean square (LMS)-optimized neural recording AFE which offers continuous EOV tracking and cancellation, all without requiring large transistors or capacitive feedback. The developed system is mostly digital and highly scalable with CMOS process technology, which paves a pathway toward implementing high-density channel systems. This article is organized as follows: in Section II, a study on which multiplexing technique is the best for neural recording systems is presented. In Section III, the EOV cancellation technique with adaptive filter (AF) is explained. In Section IV, the overall proposed neural recording system is illustrated. The system circuit topologies are introduced in Section IV. Finally, Sections VI and VII present the measurement results and conclusions of this article, respectively.

II. STUDY OF DIFFERENT MULTIPLE ACCESS TECHNIQUES

There are various channel access schemes in telecommunication and computer network that allows transmission medium sharing or, in this case, IA and ADC sharing. The applicable ones in neural recording systems are frequency division multiple access (FDMA), code division multiple access (CDMA), and TDMA.

A. Frequency Division Multiple Access

Fig. 2(a) shows a typical N-channels FDMA system block diagram. Each signal is up modulated to a different carrier frequency and then combined at a summing node to share the AFE. To retrieve the signals back, a dedicated tuned bandpass filter (BPF) selects each channel band individually, and then a demodulator recovers the signal.
adapted to neural recording applications.

Fig. 2. (a) FDMA, (b) CDMA, and (c) TDMA system block diagrams adapted to neural recording applications.

Such an approach has been used in distributed electroencephalogram (EEG) systems [21] to share the ADC among all recording channels, and in implanted neural recording systems [22] to share the RF module. Since on-chip EOV cancellation is required to avoid the saturation of the IA, sharing the IA and ADC by using FDMA requires on-chip demodulation to filter out the EOV from each channel and cancel it from the IA input. An FDMA on-chip demodulation typically requires analog/mixed-signal (AMS) circuitry like a frequency-locked loop (FLL), modulators, and demodulators. This overhead in terms of power and area is usually high in this approach which defeats the purpose toward HDNI applications. Cutting down the area and power of peripheral circuits to make FDMA competitive is an active area of research.

B. Code Division Multiple Access

Fig. 2(b) shows an N-channel CDMA-based neural recording system block diagram, where neural signals are coded with a set of perfect orthogonal codes like Walsh–Hadamard ($C_j$). Conveniently, the code is digital-like and can be fed to a chopper.

Unfortunately, utilizing CDMA for analog interfaces suffers from two main problems. The first problem is the inability to use the coding choppers to reduce the amplifier flicker noise. This is because an analog low-pass filter (LPF) after each decoded channel is normally essential to attenuate the up-modulated flicker noise and offset of the amplifier prior to digitization by the ADC. Accordingly, demodulation needs to be on the analog side instead of the digital side which adds undesired area per channel.

The second problem with CDMA is using a digital-like coding scheme in an AMS system. CDMA expects a constant input during the entire code length period to cancel crosstalk completely. However, when CDMA-chips code an analog waveform that is slightly changing over time during the code period, it results in slight delta errors each time the coded signal is sampled. Table I illustrates the issue mathematically with an ($N = 4$) channels example. $H_4$ Hadamard codes are used to code the analog input amplitudes {a, b, c, and d} at instance $T_0$; then, at $T_{1-3}$, all $\Delta_{Vin,j}$, where $j$ is the time index, errors start accumulating on the summing node of the CDMA-TX. An ideal CDMA system is achieved if all $\Delta_{Vin,j}$ are set to zero. On the CDMA-RX side, the received summation already includes the $\Delta_{Vin,j}$ errors; hence, the integrate and dump from $T_0$ to result N-times the signal power plus a total error sum of $\delta_i$, where $i$ is the channel index. Simulation results of a 16-channels system with neural signals of a few 100 mV at 10-kHz BW show that the input referred noise (IRN) can be as high as 30 $\mu$Vrms. To avoid this issue, a sample-and-hold at the input of the AFE is required as proposed in [23], which due to $kT/C$ noise reasons, requires an increased per-channel area.

For these reasons, CDMA encoding suffers from difficult issues and thus remains an active area of research as well.

C. Time-Division Multiple Access

It should be clear now that FDMA and CDMA require significant overhead in either the analog or digital domains to be used in HDNI systems. On the other hand, Fig. 2(c) shows a TDMA system block diagram which requires very minimal overhead: 1) AFE multiplexer can have a simple equivalent digital demultiplexer in the back end to recover the input neural signals [18]; 2) single neural amplifier of BW ($2 \times f_s \times N$), where $f_s$ is the neural signal BW, is sufficient to amplify the multiplexed neural signals; and 3) Nyquist ADC with a sampling rate similar to the neural amplifier BW. For these reasons, TDMA enables a potentially very compact overall implementation which is an excellent fit for HDNI applications. However, since each electrode can have a different (time-varying) offset voltage, careful design attention is required to keep a TDMA system from saturating the front end in a small area, even when ac coupling capacitors are employed.

III. EOV Cancellation

The removal of EOV for a multiple access system can be tricky. Since each channel has its own isolated random time-varying EOV, depending on the multiple access scheme used, the EOV at the amplifier input can change. For example, in FDMA and CDMA, the systems depend on adding the modulated channels on a single summing node, as shown in Fig. 2(a) and (b); hence, the total EOV get multiplied by $\sqrt{N}$ at the input of the neural amplifier. In contrast, in TDMA systems, each channel has its own isolated
time-varying EOV which, given a perfect back end, can be individually canceled at each measurement iteration.

### A. Single-Channel Versus Multi-Access EOV Cancellation

Fig. 3(a) shows an example of a digital servo-loop applied to a dc-coupled neural recording system block diagram. Solutions like Sharma et al. [18] use a binary-search algorithm to remove EOV; this approach does not guarantee the complete continuous removal of the slowly varying EOV, rather it considers the offset as a static voltage which is not always guaranteed.

Other solutions like Muller et al. [4] use a mix of an IIR-filter and binary-search algorithm for partial dynamic EOV removal as demonstrated in Fig. 3(b). Coarse and fine extraction of EOV and neural LFP signals are performed by using a binary-search algorithm and LPF, respectively. However, this approach is problematic when an on-chip IIR-LPF is used with a random EOV value, the TDMA system multiplexes all channel offsets together onto a single wire, which creates a high-frequency EOV artifact that can easily saturate an amplifier. Whether the neural amplifier is ac- or dc-coupled, this multiplexed EOV artifact can pass in full magnitude directly into a single neural amplifier and ADC with gain (\(G_{ADC}\)). The feedback loop consists of \(N\)-LPFs, an AF with gain (\(G_{AF}\)) that will be addressed shortly in Section III-B, and a DAC with gain (\(G_{DAC}\)).

Since every channel is connected to an electrode with a known cutoff frequency (usually \(\sim 300\) Hz), and the gain of the filter passband region must be exactly equal to \(1/(AV \times G_{ADC} \times G_{DAC})\) which is difficult due to process variation that affects the gain of each stage causing incomplete offset cancellation. On the other hand, if the LPF is implemented as an integrator, the loop gain might get very large and cause instability to the overall system; therefore, an attenuation factor should be added to the loop to maintain stability. If the attenuation factor is too small, the loop might not track the slowly varying EOV and hence, cause the amplifier to saturate. This factor can be manually adjusted after fabrication since the loop gain is prone to change due to process variations. Accordingly, Muller et al. [4] designed the system to have an off-chip IIR-LPF and a binary-search algorithm to manually control the filter coefficient to reach optimum performance in terms of EOV cancellation and cutoff frequency accuracy. The transfer function of the LPF is very tricky to design for a single-channel system, as explained in [4], because the gain coefficient of the LPF controls two coupled parameters: 1) loop stability and 2) loop filter poles location. The stability becomes increasingly difficult if the ADC and DAC introduced delays higher than one unit delay [4].

Fig. 3(c) shows the proposed work approach, which is an extension of the work in [4], but with a multi-access technique and simultaneous EOV cancellation across all channels. The system uses an analog multiplexer to multiplex all channels into a single neural amplifier and ADC with gain (\(G_{ADC}\)). The feedback loop consists of \(N\)-LPFs, an AF with gain (\(G_{AF}\)) that will be addressed shortly in Section III-B, and a DAC with gain (\(G_{DAC}\)).

Since every channel is connected to an electrode with a known cutoff frequency (usually \(\sim 300\) Hz), and the gain of the filter passband region must be exactly equal to \(1/(AV \times G_{ADC} \times G_{DAC})\) which is difficult due to process variation that affects the gain of each stage causing incomplete offset cancellation. On the other hand, if the LPF is implemented as an integrator, the loop gain might get very large and cause instability to the overall system; therefore, an attenuation factor should be added to the loop to maintain stability. If the attenuation factor is too small, the loop might not track the slowly varying EOV and hence, cause the amplifier to saturate. This factor can be manually adjusted after fabrication since the loop gain is prone to change due to process variations. Accordingly, Muller et al. [4] designed the system to have an off-chip IIR-LPF and a binary-search algorithm to manually control the filter coefficient to reach optimum performance in terms of EOV cancellation and cutoff frequency accuracy. The transfer function of the LPF is very tricky to design for a single-channel system, as explained in [4], because the gain coefficient of the LPF controls two coupled parameters: 1) loop stability and 2) loop filter poles location. The stability becomes increasingly difficult if the ADC and DAC introduced delays higher than one unit delay [4].
value that guarantees cancellation of the EOV. This technique ensures loop stability and high conversion speed.

Fig. 3(d) shows an example of the transient signals of the proposed system with \( N = 4 \). Neural signals, \( V_{IN} \), for each channel are multiplexed in the time domain at \( V_{MUX} \), where the aggregated signal is then amplified and digitized, and then demultiplexed for low-pass filtering at the \( V_{LFP} \) node, which results in a sum of LFP and EOV signals. Finally, the low-frequency signals are multiplexed one more time and converted into the analog domain for subtraction of EOV at the node \( V_{MLFP} \). The transient signals at \( V_{MLFP} \) confirm that the sum of EOV and LFP signals is almost static after \( N \)-delays.

B. Adaptive Filter Architecture

An LMS algorithm is often used to cancel interference in various systems. For example, the system in [24] uses two LMS AFs to remove stimulation artifacts from the recorded neural signal. Fig. 4(a) shows the equivalent block diagram of the first AMS loop of the system presented in [24]; the neural signal with dc offset is amplified, then digitized, and finally applied to an LPF to extract dc offset. A standard LMS adaptive algorithm calculates an output \( y[n] \) based on the input signal \( u[n] \) and the error signal \( e[n] \). The LMS algorithm is set to an interference-canceling mode by feeding \( e[n] \), being the difference between a desired signal \( d[n] \), resulting from the ADC, and the output signal \( y[n] \). Accordingly, the interference-canceling algorithm follows

\[
w[n + 1] = w[n] + \mu u[n] \cdot (d[n] - y[n])
\]

where \( w[n] \) is the estimated AF weight. The undesired EOV signal \( u[n] \) is generated by applying an LPF to the neural amplifier output, while \( \mu \) is a constant set by the designer which controls speed and accuracy of the AF.
The system in [24] added an extra digital LMS loop to the configuration of Fig. 4(a) for complete artifact signal cancellation. The addition of the extra loop is necessary due to a fundamental issue with this configuration that arises when a differential amplifier is used in feedback with an artifact canceling LMS loop. In other words, the LMS interference-canceling algorithm in Fig. 4(a) is changed due to the added feedback from node $y[n]$ to node $d[n]$ and, hence, the algorithm can no longer perform its intended role. This change introduces two problems: the first problem is that node $d[n]$ gets converted into an error node $e[n]$ when the algorithm runs for some time. This occurs because the error node is defined as $e[n] = d[n] - y[n]$ which is natively forced in topology of Fig. 4(a) as the amplifier performs continuous subtraction. The second problem is that the offset at the output of the amplifier will disappear over time since it is subtracted from the input of the neural amplifier, so the LPF will produce zero output at some point in time; this will force node $u[n]$ to be zero, which disturbs the stability of the filter and causes an incorrect interference cancellation mechanism.

Fig. 4(b) shows the proposed solutions for the problems of topology of Fig. 4(a). Since the output signal of the AF, $y[n]$, carries the EOV value, it can be added to both nodes $u[n]$ and $d[n]$ to cancel the EOV subtraction effect of the neural amplifier. Fig. 4(c) compares the LMS error across multiple algorithm iterations for the method in Fig. 4(a) and the proposed method in Fig. 4(b). Here, it can be seen that the LMS algorithm as applied in Fig. 4(b) can reach its full accuracy and conversion speed capability when deployed with a differential amplifier by using the proposed feedback.

The overall proposed LMS AF block diagram is shown in Fig. 4(d). An additional unit-delay block is inserted after the node $y[n]$ for signal timing flow consistency of the LMS loop in a TDMA system. Finally, saturation blocks are inserted at the input of the nodes $u[n]$ and $d[n]$ to avoid algorithm overflow at system startup which helps with LMS conversion and locking to the correct $w[n]$ value.

IV. PROPOSED NEURAL RECORDING SYSTEM

Fig. 5 shows a block-level diagram of the proposed neural recording system. The system is composed of a high-density neural pixel (HDNP), a highspeed digital MUX, and an RF module with integrated antenna. This work focuses on the design of the HDNP only, and the rest will be considered as future work. Each HDNP block consists of 16 fully differential input recording channels: a 16:1 analog multiplexer combines the neural signals onto an ac-coupled fully differential neural amplifier which is biased by pseudo resistors, $R_p$. A 10-bit SAR ADC digitizes each code and passes it to the on-chip DSP module for EOV estimation. Then, the DSP passes the continuously updated estimates to a 10-bit capacitive DAC (CDAC) which performs electrode EOV cancellation in the current domain. The neural amplifier is ac-coupled after the multiplexer for two main reasons: 1) to properly bias the neural amplifier without recourse to EOVs and 2) to convert the input voltage into the current domain for EOV subtraction. If the amplifier were dc-coupled, then $R_p$ must be much less than the electrode resistive impedance to correctly bias the amplifier. However, the lower $R_p$ gets, the higher the signal losses become due to the potential division between the electrodes and the amplifier input impedance. On the other hand, if the capacitors are increased, the input signal will be less attenuated, but this trades off with the analog module area which will increase. In addition, it attenuates the CDAC canceling signal which sees a potential divider between the DAC capacitors, the amplifier input device parasitic capacitance, and the ac-coupling capacitance. This limits the magnitude of the offset cancellation signal to $\pm 65$ mV if the CDAC is supplied by 1.2 V.

Since the DAC needs to cover the full dynamic range of both the EOV and the small neural signal, the effective number of bits (ENOB) requirement is relatively high at 15-bit. Therefore, the CDAC is implemented as a second-order $\Delta \Sigma$-Modulator, designed as an error-feedback noise-shaping loop [25], to reduce the DAC’s actual bit-width from 15 to 10 bit. Although the $\Delta \Sigma$-Modulator requires additional power, the reduced bit-width offers more savings in terms of DAC area and eases the matching requirement of unit DAC elements. The second-order $\Delta \Sigma$-Modulator has an oversampling ratio of 32; this makes the highest digital clock required to operate the system to be 10.24 MHz.

Finally, the open-loop amplifier saves area by eliminating the need for passive component feedback and instead running open loop; however, the offset caused by the mismatch of the amplifier input differential pairs can easily cause saturation. Fortunately, the proposed digital algorithm can natively recover from the amplifier input-referred offset.

The drawback of this open-loop configuration is the inability of using choppers to further decrease noise. If choppers were used, the amplifier offset would be up modulated, causing the amplifier to saturate, and the digital algorithm would not be able to recover the neural signal in that case. Instead, the only ways to reduce flicker noise in the proposed approach are: 1) use a separate feedback LMS loop that deals with chopped offset signals and 2) increase the devices sizes, but this solution increases the input parasitic capacitance, hence, lowering DAC cancellation capabilities. The latter solution is used in the proposed system as it requires less on-chip area and is still amortized across all of the channels.

The input impedance of the HDNI AFE system should be designed to be much higher than the electrode impedance to avoid signal attenuation, due to potential division, at the system input interface. Since the implantable electrode sizes are expected to be very small as the number of recording channels increase, the electrode impedance is expected to increase as well. This poses a design challenge on the overall AFE. Fig. 6 compares input impedance simulation results of a chopper-based AFE versus a different number
Fig. 5. Overall block diagram of the proposed system where a 16-channel TDMA signal acquisition and a DSP block for EOV removal are implemented.

Fig. 6. (a) Chopped-based input AFE. (b) TDMA-based input AFE. (c) Lowest average input impedance magnitude for a chopper-based AFE versus TDMA-based AFE with $C_{IN} = 2 \text{ pF}$ and $f_s = 20 \text{ kHz}$ at different number of multiplexed channels.

system in a single period of the clock $(f_s)$ is half the value of the chopper-based AFE. Accordingly, the lowest differential input impedance for a TDMA-based system at relatively low frequencies can be expressed as $Z_{IN} = 1/f_s C_{IN}$. Since the change in voltage between any two consecutive channels in TDMA-based AFE is random, the input impedance can be higher than the given expression. Recent development of implantable micro-electrocorticography ($\mu$ECoG) in [26] shows that the impedance of a 100-$\mu$m-diameter coated carbon nanotube (CNT) is 100 k$\Omega$. This gives the TDMA-based systems input impedance a promising compatibility of high-density implantable electrodes.

B. HDNP DSP Module

The DSP module consists of a digital controller which regulates the channel selection of the analog MUX and digital MUX–DEMUX. In addition, it samples the LMS filter weight and switches off the digital multiplier M1, shown in Fig. 4(d), to save power in case the filter weight remains within a programmable range of $\pm 2$ digits for 10 s.

For EOV cancellation, the digitized neural signal gets demultiplexed first and passes through a digital LPF for each channel separately. The LPF architecture used in this work is a delay-free digital integrator. Both the EOV and LFP get extracted and then the signals get multiplexed again to form a reference signal for the LMS algorithm. The interference-canceling LMS-AF senses the raw neural signals from the ADC and EOV/LFP reference signal and computes the exact EOV/LFP voltage level to be removed from the neural amplifier input summing node.

The LMS algorithm in the digital module has $\mu$ value set to $2^{-34}$; this value is considered a good balance in terms of speed and offset removal accuracy for the overall system.

V. CIRCUIT IMPLEMENTATION

To achieve the lowest noise and area from the AFE, the employed circuit topology should be carefully considered. The analog multiplexer consists of thick-oxide NMOS
The overall approximate mid-band gain is given by 

\[ A_{\text{od}} \approx (g\text{m}_{Mn1} + g\text{m}_{MP1}) (r\text{o}_{Mn1} + r\text{o}_{MP1}) \times (g\text{m}_{Mn2} + g\text{m}_{MP2}) (r\text{o}_{Mn2} + r\text{o}_{MP2}) R_i \]

whereas the approximate BW is \(1/2\pi R_i C_i\).

The PMOS current tails are divided into bleeders to avoid latching at startup. The common mode feedback (CMFB) current sources are fed by pseudo-resistors directly without a CMFB amplifier. This topology reduces the loop gain of the CMFB which increases stability in the open-loop configuration in addition to saving area and power. The second stage has an optional bank of Miller-capacitors and resistors to linearly adjust the GBW of the amplifier using 3-bit digital controllers each if needed. The total capacitance is 280 fF on each differential output, resulting in BW variation from 210 to 830 kHz when operating with the 40-dB gain setting. This range was chosen to cover any process variations after fabrication such that the exact desired BW would be set to 320 kHz. The resistive bank varies the amplifier gain linearly from 35 to 52 dB with a 3-bit controller. However, for a 10-bit accuracy requirement, the dynamic settling error \(\epsilon\) must be \(\leq 0.1\%\); this requires the amplifier BW to be set to a one-step higher BW setting (397 kHz) to ensure accuracy. The practical amplifier BW is given by 

\[ f_{BW} \geq -f_i \ln(\epsilon)/2\pi \approx 1.1 f_i \] [28].

To save more power and energy, the segmented charge-redistribution circuit topology shown in Fig. 7(b) is used in both the DAC and the ADC with slight differences between the two. It is built out of metal parasitic capacitance as demonstrated in [15]. The DAC unit capacitors \(C_a + C_f\) equal 8.2 fF in all combinations. \(C_{1a}\) and \(C_{1f}\) are 4 and 4.2 fF, while \(C_{2a}\) and \(C_{2f}\) are 3.9 and 4.3 fF, respectively. The first five LSB differential capacitor blocks \((B0 \rightarrow B4)\) are binary weighted. The remaining 5-MSBs \((B5 \rightarrow B9)\) are thermometer weighted, with \(C_{5a}\) and \(C_{5f}\) set to 0.9 and 7.3 fF, respectively. Each capacitor block is controlled by a differential binary input BN and BB on its bottom plate; hence, the effective unit capacitance is 0.2 fF.

Fig. 7(c) shows a conventional 10-bit asynchronous SAR ADC algorithm used in the system [29]. Unlike the DAC capacitor-bank topology, the SAR ADC uses nine capacitance blocks only instead of ten blocks.

With the DAC having 8.2 fF \(\times 36\) total equivalent capacitance \((C_{DAC})\), CMOS input parasitic capacitance \((C_{INP})\) of 0.25 pF, and 2-pF ac coupling capacitors \((C_{ac})\), the neural and DAC signals get attenuated by

\[ \Gamma_{\text{Neural Signal}} = \frac{C_{ac}}{C_{ac} + C_{INP} + C_{DAC}} \] (2)

\[ \Gamma_{DAC} = \frac{C_{DAC}}{C_{ac} + C_{INP} + C_{DAC}} \] (3)

which result in \(\Gamma_{\text{Neural Signal}} = 0.78\) and \(\Gamma_{DAC} = 0.11\). Assuming a \(V_{DD}\) of 1.2 V, the maximum DAC cancellation capability reaches \(\pm 69.5\) mV. Monte Carlo simulation results show that the input-referred offset can be as high as 4.5 mV, this gives an EOV cancellation margin of \(\pm 65\) mV which is sufficient to prevent the neural amplifier saturation from common electrode types.

VI. MEASUREMENT RESULTS

The proposed neural recording system IC was fabricated in 65-nm 1p9m low-power (LP) CMOS technology. Fig. 8 shows a microphotograph of the fabricated IC; the total area of the 16 channels is 0.0397 mm², including all analog and digital blocks. The analog module consumes 13.8 \(\mu A/16\) Ch from a 1.2-V supply, while the digital module uses 37.5 \(\mu A/16\) Ch from a 1-V supply. The entire IC power consumption is
Fig. 8. Proposed system IC chip microphotograph.

Fig. 9. IC power and area breakdown for 16 channels.

54.06 $\mu$W for all 16 channels; Fig. 9 summarizes the power and area breakdown for each module.

The measured input impedance $Z_{IN}$ at 100 Hz is 28 MΩ, which is very close to the expected worst case input impedance. The lowest measured common mode rejection ratio (CMRR) and power supply rejection ratio (PSRR) are 66 and 79 dB, respectively. This is measured while injecting ±50 mV slowly varying offsets to all channels and observing CH6 and CH14 for the ripple amplitude. Fig. 10 shows the power spectral density (PSD) of the ADC sampled at 320 kHz with a direct injected signal of 400 mVpp at 47.969 kHz by a Stanford Research System (SRS) DS360 function generator. For the standalone ADC, the measured SNR is 50.79 dB and signal-to-noise and distortion ratio (SNDR) is 50.16 dB resulting in ENOB of 8.04 bits. The ADC third harmonic distortion (THD) is 0.146%. Fig. 11 shows the measured integral non-linearity (INL) and differential non-linearity (DNL); some codes cross ±0.5 LSB due to mismatches of the small-area nature of the overall parasitic capacitors DAC of the ADC.
This could be improved in a future version of the chip. The overall system of measured SNR with ±50 mV injected EOV is 46.96 dB, and SNDR is 46.29 dB, resulting in ENOB of 7.4 bits.

The measured closed-loop transfer function of AP and LFP for all 16 channels is shown in Fig. 12. Measurements of five different dies show that the neural amplifier mid-band gain mean value is 40 dB with a standard deviation of 4.12 dB and BW mean of 397 kHz with 194.1 Hz standard deviation, respectively. The neural amplifier statistical analysis is measured from a debugging buffer, shown in Fig. 8, connected to the output of the neural amplifier to avoid additional pads loading effect.

Fig. 13 shows the PSD of all 16-channels simultaneously; channels 1 → 8 are injected with an ACCES USB-DA12-8A-PR arbitrary waveform generator. Each channel is injected with a different amplitude, frequency, and slowly varying EOV. Channels 1 → 7 are wired off-chip to channels 9 → 15 due to limitation of laboratory devices. Channel 16 is connected to the SRS-DS360 function generator. Channels 6 and 14 are injected with slowly varying EOV and zero neural-band signals to measure IRN; both channels had a maximum IRN of 2.1 μV rms. In another trial with all channels injected with EOV only, the maximum measured IRN is 2.6 μV rms. The worst case measured EOV rejection is 53 dB. Fig. 14 shows the system test bench setup and a simplified schematic view. The neural signals pass through an emulated electrode impedance of 100 kΩ. With 28-MΩ AFE input impedance, the signals are not affected much.

Table II summarizes the achieved results and compares them with the state-of-the-art neural recording systems. The proposed technique has the advantage of scalability with node advancement compared with other techniques which could easily bring the channel area to below 0.001 mm² in a more scaled CMOS process.

VII. Conclusion

A multi-channel TDMA neural recording system with an LMS filter that continuously detects and removes EOV is proposed in this article. The proposed LMS-topology, with modified feedback, enables the use of an LMS algorithm with
a differential amplifier in the AMS loop while maintaining its further cancellations capabilities in terms of speed and accuracy. Fabricated in 65-nm technology, this technique enables a reduction in the per-channel area compared with other recent approaches while maintaining LP and low-noise capabilities. The proposed system power consumption is 3.38 μW/Ch, and the IRN is 2.6 μVRMS, which gives an noise efficiency factor (NEF) of 1.83. The overall per-channel area is 0.00248 mm², of which 68% is digital area which is highly scalable with node technology.

**REFERENCES**


[23] R. Ranjandish and A. Schmid, “Walsh–Hadamard-based orthogonal circuits, and brain–machine interfaces. Research interests are analog integrated circuits and systems design, biomedical design with the University of California San Diego, San Diego, CA. His graduate research assistant working on biomedical circuits and systems during the summer of 2020 and 2021 with the Analog Silicon Engineering.


Nader Sherif Kassem Fathy (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electronics and telecommunication engineering from Ain Shams University, Cairo, Egypt, in 2013 and 2018, respectively. From 2014 to 2016, he worked as a Teaching Assistant with the Department of Electronics and Communications, Ain Shams University. He was a former Senior Software Development Engineer with the Caliber Research and Development Department, Mentor Graphics—Siemens, Cairo, from 2013 to 2018. He has held two internships with Apple Inc., Cupertino, CA, USA, during the summer of 2020 and 2021 with the Analog Silicon Engineering Group focusing on phase-locked loop (PLL) designs. He is currently a Graduate Research Assistant working on biomedical circuits and systems design with the University of California San Diego, San Diego, CA. His research interests are analog integrated circuits and systems design, biomedical circuits, and brain–machine interfaces.

[29] Jiannan Huang (Member, IEEE) received the B.Sc. degree in electrical and computer engineering from the University of Alberta, Edmonton, AB, Canada, in 2006, and the S.M. and Ph.D. degrees in electrical engineering and computer science from Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2008 and 2012, respectively. He is currently an Associate Professor of electrical and computer engineering with the University of California San Diego (UCSD), San Diego, CA, USA, where he is also the Co-Director of the Center for Wearable Sensors and the Site Director of the Power Management Integration Center. His research interest includes the design of energy-efficient microsystems, focusing on the design of RF circuits, power converters, and sensor interfaces for miniaturized systems and biomedical applications.

Prof. Mercier is a member of the International Solid-State Circuits Conference (ISSCC) International Technical Program Committee, the CICC Technical Program Committee, and the Very Large Scale Integration (VLSI) Symposium Technical Program Committee. He was a recipient of the Natural Sciences and Engineering Council of Canada (NSERC) Julie Payette Fellowship in 2006, NSERC Postgraduate Scholarships in 2007 and 2009, an Intel Ph.D. Fellowship in 2009, the 2009 IEEE ISSCC Jack Kilby Award for Outstanding Student Paper at ISSCC 2010, a Graduate Teaching Award in Electrical and Computer Engineering at UCSD in 2013, the Hellmann Fellowship Award in 2014, the Beckman Young Investigator Award in 2015, the DARPA Young Faculty Award in 2015, the UC San Diego Academic Senate Distinguished Teaching Award in 2016, the Biocom Catalyst Award in 2017, the NSF CAREER Award in 2018, a National Academy of Engineering Frontiers of Engineering Lecture in 2019, and the San Diego County Engineering Council Outstanding Engineer Award in 2020. He served as an Associate Editor of the IEEE Transactions on Very Large Scale Integration (TVLSI) from 2015 to 2017. Since 2013, he has served as an Associate Editor of the IEEE Transactions on Biomedical Circuits and Systems (TBioCAS). He was the Coeditor of “Ultralow-Power Short Range Radios” (Springer, 2015), “Power Management Integrated Circuits” (CRC Press, 2016), and “High-Density Electrocortical Neural Interfaces” (Academic Press, 2019). He is an Associate Editor of the IEEE Solid-State Circuits Letters.