A Single-Chip Bidirectional Neural Interface With High-Voltage Stimulation and Adaptive Artifact Cancellation in Standard CMOS

John P. Uehlin, Member, IEEE, William Anthony Smith, Member, IEEE, Venkata Rajesh Pamula, Member, IEEE, Eric P. Pepin, Steve Perlmutter, Visvesh Sathe, Member, IEEE, and Jacques Christophe Rudell, Senior Member, IEEE

Abstract—A single-chip, bidirectional brain–computer interface (BBCI) enables neuromodulation through simultaneous neural recording and stimulation. This article presents a prototype BBCI application-specified integrated circuit (ASIC) consisting of a 64-channel time-multiplexed recording front-end, an area-optimized four-channel high-voltage compliant stimulator, and electronics to support the concurrent multi-channel stimulus artifact cancellation. Stimulator power generation is integrated on a chip, providing ±11-V compliance from low-voltage supplies with a resonant charge pump. High-frequency (~3 GHz) self-resonant clocking is used to reduce the pumping capacitor area while suppressing the associated switching losses. A 32-tap least mean square (LMS)-based digital adaptive filter achieves 60-dB artifact suppression, enabling simultaneous neural stimulation and recording. The entire chip occupies 4 mm² in a 65-nm low power (LP) process and is powered by 2.5-/1.2-V supplies, dissipating 205 μW in recording and 142 μW in the stimulation and cancellation back-ends. The stimulation output drivers achieve 31% dc–dc efficiency at a maximum output power of 24 mW.

Index Terms—Artifact cancellation, brain–computer interfaces, electrical stimulation, neural recording, time-division multiplexing.

I. INTRODUCTION

REALIZATIONS of small-form factor, ultralow-power bidirectional brain–computer interfaces (BBCIs) will enable the treatment of chronic neurophysiological disorders and allow new modes to investigate brain function. Neural stimulators have been shown to effectively alleviate the symptoms of Parkinson’s disease [1], epilepsy [2], depression [3], and obsessive-compulsive disorder [4]. The development of closed-loop neural interfaces with simultaneous recording and stimulation capabilities will increase therapy effectiveness by adapting to real-time measurements of the modulated neural tissue without input from a practitioner [5]. In addition, a sufficiently complex closed-loop interface could be used to reanimate the damaged neural tissue or drive neuroprostheses with sensorimotor feedback [6]. In addition to clinical applications, simultaneous stimulation and sensing opens new research opportunities for neuroscientists, including the study of neural plasticity [7] and low-latency neural mapping [8].

In order to increase the efficacy and accessibility of potentially life-changing treatments, BBCI implants must be developed with minimal cost, size, and power use. A single-chip BBCI implementation minimizes the area and power consumption by reducing the number of interconnects. In addition, standard scaled-CMOS technology offers dense and power-efficient digital devices for complex BBCI therapy protocols. A single-chip CMOS BBCI front-end requires the integration of recording, stimulation, and power generation electronics on the same substrate along with a sophisticated digital back-end. Fig. 1 shows the components of a single-chip BBCI front-end. In addition, resilient operation in chronic implants requires the capability to continuously adapt to a slowly varying environment to maintain the desired stimulation and recording function.

Critical challenges remain before the vision of implantable single-chip BBCI neural interfaces becomes a reality; this article addresses two such challenges. First, the voltages generated at the stimulator–electrode interface (>10 V) and...
regularly exceed the acceptable gate oxide breakdown limits (<1.5 V) for reliable operation in advanced, scaled CMOS technologies [9]. Second, the electrical stimulation creates in-band artifacts that are several orders of magnitude larger than the signals targeted by neural recording front-ends (~100 mV versus ~100 µV) [10].

Recent implementations use a variety of techniques to address stimulation artifacts. One approach simply migrates to a high-voltage (HV) CMOS process, where the traditional current/voltage-driving structures can readily tolerate high stimulation voltages. The recording channels have been designed in HV processes that are naturally immune to large voltage swings due to stimulation artifacts [11]. However, the large core devices in these processes are less suited for efficient integration of digital processing algorithms as will be demanded of future generation BBCIs. Other implementations mitigate the stimulation artifacts by overdesigning the record front-end dynamic range. These approaches quantize both the large stimulus artifact and the underlying neural signals [12]–[16], relying on a DSP back-end to digitally remove the artifact. The front-end overdesign can create a recording non-linearity (voltage-controlled oscillator (VCO)-based ADCs [12]) and introduces area, complexity, or power-consumption tradeoffs. Artifacts are most efficiently mitigated using the adaptive mixed-signal feedback in the front- and back-ends [17]–[19]. The proposed multiplexed canceller is optimized to reduce the power consumption, using less than 50 nW per artifact channel during continuous operation, and less than 8 nW per channel for a standard 10-Hz stimulation rate.

This article achieves HV stimulation compliance (±11 V) using stacked circuits in low-voltage scaled CMOS. The stacked driver architecture enables constant-current stimulation through a large range of electrode impedances from the same chip as a complex, efficient digital back-end. Charge pumps are typically used to create above-\(V_{DD}\) stimulation voltages, and the pumping capacitors dominate the system area. This article minimizes the charge pump area using integrated inductors to reclaim power while operating at a high charge pump clock frequency to reduce the area overhead. The proposed stimulator implements the first GHz-clocked charge pump to achieve aggressive area reduction in neural stimulation, while relying on resonance to effectively mitigate the adverse efficiency impact from this miniaturization.

The HV stimulators are integrated with the time-division multiplexed recording architecture from [20] and [21]. The recording system uses closed-loop adaptive feedback for stimulus artifact cancellation on multiple channels. The canceller computation hardware is also time-multiplexed to reduce power and area consumption. The resulting monolithic 65-nm CMOS chip includes all the necessary components to allow simultaneous 64-channel recording and 4-channel stimulation for BBCI-clinical applications. This article is an elaboration of the work presented in [22].

This article is organized as follows. Section II discusses the overall system architecture. Section III elaborates on the resonant H-bridge stimulator and analysis of power generation and efficiency. Section IV briefly describes the integrated recording channels. Section V describes the design and optimization of the adaptive artifact canceller. Section VI provides the system and circuit implementation details, as well as benchtop and in vivo measurements. Concluding remarks are presented in Section VII.

II. System Overview

A block diagram of the full system implemented is shown in Fig. 2. Four integrated H-bridge stimulators are independently tunable with digitally programmable current waveforms. The resonant charge pumps generate up to 11 V to drive up to 2 mA of sink-regulated current through the electrode–tissue load. The shared IDAC is switched between the two sides of the H-bridge.

A time-multiplexed, delta-encoded recording front-end [20], [21] records 64 channels using a single recording chain. A multiplexer routes all channels to a shared input capacitive DAC (CDAC), front-end amplifier, and successive approximation register (SAR) ADC. A digital integration loop increments the CDAC at the recording amplifier input to subtract low-frequency signal content. Such a delta-encoded loop creates a frequency-shaped dynamic range well suited for the colored spectral shape of neural signals [23]. The system targets 64-channel electrocorticography (ECoG) recordings at 2 kS/s.

The digital back-end implements an adaptive artifact cancellation filter. A simplified least mean square (LMS) algorithm processes the recording output and stimulator control signals into a time-domain reconstruction of unwanted stimulus artifacts. The artifact voltage is subtracted using the CDAC at the recording amplifier input. The canceller computation hardware associated with adapting the filter is time-multiplexed between channels, and the filter coefficients are stored using an on-chip dedicated SRAM.

III. Resonant Charge Pump H-Bridge Stimulator

A. HV Stimulation

Constant-current stimulation pulses create large voltages at the load interface due to high electrode impedances [24]. This article uses an H-bridge stimulator topology to double the voltage compliance range for a given maximum output voltage. The stimulator output is ground-referenced, and each differential side has a dedicated charge pump to supply positive voltage as needed to each side of the load. The charge pumps are dynamically enabled to generate only as much voltage as necessary to drive the current through the tissue–electrode load. Specifics of the constant-current voltage control loop can be found in [25]. Compared with a fixed-supply stimulator [11], [26]–[28], dynamic voltage supplies save energy when stimulating below the maximum output voltage [29].

B. Resonant Charge Pump Voltage Supplies

This article uses a modified version of the cross-coupled switched capacitor charge pump presented in [30], which is well suited for use in triple-well, deep sub-micrometer integrated CMOS processes. The charge pump draws current from an input voltage, \(V_N\), and adds the voltage proportional to the number of cascaded stages, \(n\), and the clock voltage.
swing, $V_{CLK}$, applied to the flying pump capacitors, $C_{FLY}$ (Fig. 3). The output voltage, $V_O$, droops for given output current, $I_O$, in proportion to the clock-switching frequency, $f_{CLK}$, and the flying capacitor size at each stage. This gives the following output voltage characteristic:

$$V_O = V_{IN} + n(V_{CLK} - I_O/f_{CLK}C_{FLY}).$$

The switched capacitor charge pumps are typically implemented with large switches and capacitors at low frequencies, mitigating frequency-dependent reactive losses. Milliampere-level stimulation current with clock frequencies in the tens of megahertz requires large capacitors ($\sim 100$ pF) in the charge pumps, thus, dominating the overall stimulator silicon area [31].

Existing stimulators often use large ($\sim 1$ $\mu$F) off-chip flying capacitors switched at low frequencies ($\sim 100$ kHz) [32], taking up more area than the core stimulator IC. This article reduces the silicon footprint dedicated to charge-pump flying capacitors by increasing $f_{CLK}$ to maintain the output power levels with smaller capacitor sizes. Efficiencies afforded by resonance ensure that $f_{CLK}$ increase does not adversely impact switching losses. A significant reduction in charge-pump capacitance allows for single-chip implementations that include multiple stimulation sites.

Resonant charge pumps have been previously implemented for neural stimulators, but without an increase in switching frequency and corresponding area savings. A large, single-coil inductor created around the chip boundary resonates with a single charge pump, with the goal of using the 24-nH coil as a wireless energy-harvesting mechanism [29]. The charge pump and inductor occupy 3 mm $\times$ 3 mm of the silicon area while only providing enough power for 145 $\mu$A of stimulation at $\pm 3.3$ V.

This article integrates eight resonant charge pumps dedicated in pairs to realize four fully integrated stimulators on a single chip. Each stimulator runs from a 500-mV $V_{DD}$ and is capable of independently delivering up to 2 mA of current, while ramping to a maximum driver voltage of 11 V. During normal operation, the $LC$ tank forms a free-running oscillator, obviating the need for a phased-locked loop.

The capacitance associated with a 12-stage cascaded charge pump is placed in parallel with a single differential inductor to create a resonant oscillator, as shown in Fig. 3. Large, cross-coupled NMOS devices generate negative impedance, compensating for resistive losses in the $LC$ tank to ensure oscillation. An inductor center tap set to $V_{DDRES}$ results in a $2-V_{DDRES}$ voltage swing at the charge pump clock input. This differential oscillator topology was originally proposed for microprocessor clock generation in [33]. Here, a similar clocking concept is used to drive a capacitive charge pump, as opposed to a clock distribution network.

Optimizing the charge pump in the context of an $LC$ tank requires a steady-state model to estimate the tank quality factor and the resulting charge pump efficiency. Calculating the charge pump efficiency begins by finding the dc current necessary to generate enough negative impedance in the cross-coupled NMOS pair to sustain a free-running oscillator. Resonating parasitic capacitances with an inductor mitigates $CV^2f$ losses. As such, the dc bias current of the oscillator drivers and the resistive losses in the series charge pump stages dominate the stimulator power loss.

The steady-state model in Fig. 3(b) assumes that each intermediate node between charge pump stages is an ac ground. Two complimentary switches are “ON” at any given time in each stage, and this is represented as $R_{SW}$. $C_{FLY}$ combines in series with the much smaller switch parasitics, so $C_{FLY}$ is not included in this model. The remaining parasitic capacitances are represented by $C_{PAR}$, which contain the average of state-dependent device parasitics (depletion region $C_{P,N,off,on}$), and state-independent capacitances ($C_{gb}$, $C_{ds}$, and capacitor...
bottom plate parasitics $C_{\text{plate}}$. The total capacitance to ac
ground, averaged between all clock states, is given by
\[ C_{\text{PAR}} = 2C_{\text{plate}} + \frac{C_{g_s} + C_{g_d} + C_{g_s} + C_{g_d}}{2} + C_{gd} + C_{ds} + C_{ds}. \quad (2) \]

The capacitance between differential clock terminals, $C_{SW}$, is
given by the gate–drain capacitance of the switches
\[ C_{SW} = 2C_{gd} + 2C_{gd}. \quad (3) \]

The cross-coupled NMOS devices also introduce their
own parasitic resistance and capacitance, represented by
$r_{ds}$, $C_{gs}$, and $C_{gd}$.

Transforming the series impedances in Fig. 3(b) into a set of
parallel impedances yields a simplified expression to calculate
the driver $g_m$. The inductor quality factor used to calculate
an equivalent series resistance and transformed into a parallel
resistance at the resonant frequency. With all the impedances
in parallel form, the
\[ C_{SW0} = \frac{1}{g_m} = \frac{R_{LS} + \frac{1}{C_{Total}L_S} \frac{12}{R_{SW}}}{R_{LS}}. \]

This allows the estimation of oscillator $g_m$ in terms of operating
frequency and inductor size. This expression includes the
frequency-dependent quality factors from foundry-provided
integrated inductors. In previous analysis, $Q(f, L)$ was
simplified as a series resistance, $R_{LS}$. The total parasitic capacitance,
$C_{Total}$, is shortened to $C$ for clarity. The switch parameters
scale linearly with total capacitance, using simulation-verified
values as the starting points $(g_{ds0}, R_{SW0}, C_0)$
\[ g_m(f, L) = g_{ds0} + \frac{R_{LS}C_{Total}}{L_S} + \frac{12}{R_{SW}}. \]

The $g_m$ necessary to sustain oscillation can be translated into
dc current drawn through the $g_m/I_D$ ratio for a given operating
regime and process technology. The oscillator startup
begins in the subthreshold region, with $g_m/I_D = 25$ in a
65-nm CMOS. Efficiency is calculated in terms of the ideal
output power and the $g_m$ given in (6).
\[ \eta = \frac{P_{OUT}}{25 g_m(f, L) \ast 500 \text{ mV}} = \frac{2 \text{ mA}(12 \text{ V} - 12 \text{ mA} \ast R_{SW}(f))}{25 g_m(f, L) \ast 500 \text{ mV}}. \quad (7) \]

Fig. 4 graphically shows the resonant charge pump efficiency as a function of inductor size and self-oscillation.

---

**Fig. 3.** Transformation of (a) the resonant charge pump schematic into (b) an equivalent steady-state model and (c) equivalent half-circuit model for efficiency calculations.
frequency. The parallel combination of the charge pump and integrated inductor gives an efficiency peak in the low GHz.

Fig. 4 also includes an estimate of the overall area occupied by the resonant charge pump. This LC tank model assumes that the flying capacitors and integrated inductors dominate the overall area, and that the switches are sized to ensure the total parasitic capacitance is less than 10% of the flying capacitance. Maintaining this sizing limits the clock amplitude reduction arising from the series capacitive division.

The implemented resonant charge pump was designed at the minimum area point along the efficiency peak for various inductor sizes, as shown in Fig. 4. This optimum lies at a 3-GHz free-running frequency using a 180-pH integrated differential inductor. The 0.042-mm$^2$-charge pump area estimate matches the silicon area of the actual charge pump (including the inductor) in the fabricated device. The measured dc/dc converter efficiency of 31% matches the power efficiency predicted by the LC tank resonant charge pump model.

Compared to prior work with charge pumps clocked at 100 MHz [25], this architecture achieves similar losses in a 6× small-form factor with a 3-GHz charge-pump clock frequency (31% versus 40% efficiency). Using larger, non-integrated inductors with a higher quality factor would enable higher efficiency with this topology, at the expense of a larger board-form factor.

C. H-Bridge Stimulator

The resonant charge pumps fit into the H-bridge stimulator architecture as shown in Figs. 5(b) and 6(a). The charge pump supplies power to one side of the H-bridge while the current is sink-regulated through the opposite side of the load. During the return phase, the opposite charge pump is enabled, and the current sinks through the previously active side. The H-bridge-switching structure is formed with diodes and distributed current buffers to overcome the low 1.2-V voltage tolerance of individual CMOS devices.

To avoid wasted energy, the charge pumps for a given supply side are only enabled when the shared supply comparator detects that the IDAC is going out of saturation. This regulates the stimulator output voltage to be no larger than necessary for the load current into the electrode impedance. The same supply comparator is reused for the two sides of the H-bridge. After stimulation on a given side, the track comparator on that side detects a forward/reverse bias in the diode through a capacitive divider. This comparator drives a discharge path in the charge pump until the diode is “OFF.”

While one side of the H-bridge supplies voltage across the electrode–tissue interface, the current flows through a high-voltage adapter (HVA) into the current-controlling IDAC. A schematic of the HVA is shown in Fig. 5(c). The HVA is a multi-stage cascode operating as a current buffer, protecting the 1.2-V IDAC from large voltages as seen at the stimulator–electrode interface. When an H-bridge side is sourcing current, the charge pump output voltage is capacitively distributed across the cascode gates to ensure no transistor sees more than 1.2 V across its terminals. After stimulation, the floating nodes of the capacitive divider are discharged with accessory charge pumps (not shown). When sinking the current, the HVA gates are pre-charged to place the devices in triode and act as closed switches.

The sinking IDAC shown in Fig. 5(a) has 8 bits of resolution with a tunable least significant bit, nominally 10 µA. An active cascode at the IDAC output provides a high output impedance of 600 MΩ (simulated).

The ground-referenced H-bridge requires discharge phases after each stimulation pulse. In addition, the stimulator operates on the condition that there are no other paths to ground connected to the tissue. Because the monopolar voltage supplies are used to double the headroom, a low-impedance ground path would cause the supply side of the H-bridge to source extra, non-regulated current. The implemented H-bridge provides a ground reference when not stimulated, removing the need for an additional reference.

IV. DIFFERENTIAL ARTIFACT CANCELLATION

A. Artifact Cancellation in the System

When delivering current into tissue, which is also connected to the recording electronics, the stimulation voltage propagates to the recording inputs and is measured as a so-called stimulus artifact. Artifacts at the recording channel input can be several orders of magnitude larger than the signals of interest in ECoG/neural applications and can be decomposed into differential-mode and common-mode components. We have observed differential artifacts as large as ±100-mV in vivo during HV stimulation with microwire recording and stimulation electrodes 1-mm apart. As shown in Fig. 7, the artifact canceller uses the recording system output to create a negative copy of differential artifacts. The canceller output sums with the input signal, eliminating the input artifact and revealing the underlying neural signal. Common-mode artifacts are mitigated with a switched-capacitor offset cancellation network, as described in [21].
The time-multiplexed recording system uses a 10-bit CDAC at the input of the transconductance amplifier to delta-encode low-frequency signal content which relaxes the required dynamic range of a single recording channel ADC [20], [21]. Adaptive differential digital artifact cancellation is performed through a separate feedback loop which utilizes the same CDAC at the recording channel input (Fig. 7). The recording architecture in this article is an enhancement of [20] and [21], with increased CDAC resolution and expanded digital tunability. This article focuses on a new implementation of
an on-chip differential cancellation technique, realized using an ensemble of adaptively trained impulse response filters for various stim/record channel pairs.

### B. FIR-Based Adaptive Filters

Fig. 8 shows an adaptive canceller algorithm where the finite impulse response (FIR) filter coefficients are adapted using a LMS algorithm. For an FIR-based canceller implementation, the filter input signal, \( x(n) \), is a time-domain representation of the stimulation current pulse. This input is convolved with a set of coefficients, \( \{c_0, c_N\} \) to create the filter output, \( y(n) \), to create the FIR filter transfer function

\[
y(n) = \sum_{i=0}^{N} c_i x(n - i). \tag{8}
\]

The coefficients are trained based on the recording system output, \( e(n) \), which is the residual artifact and neural signal remaining after subtracting the FIR filter’s output from the input signal, \( y(n) \). The FIR coefficients can be trained with an LMS algorithm, where the update coefficient, \( \mu \), tunes the step size at each update interval. Coefficients adapt the algorithm as

\[
c_N(n+1) = c_N(n) + \mu e(n) x(n - N). \tag{9}
\]

Implementing the full LMS/FIR adaptive filter shown in Fig. 8 requires power and area-intensive hardware. Each tap of the FIR filter requires two multipliers, one adder, and two delay cells. This cost is multiplied by the number of implemented filter channels. This article proposes a simplification of this adaptive filter that dramatically reduces the necessary hardware for similar functionality and performance across multiple channels.

### C. Impulse-Based Adaptive Filter

The end goal of the adaptive filter is a time-domain representation of the stimulation voltage artifact. The FIR coefficients in the previously described approach are an intermediate step, quantifying the channel which changes the stimulator output current into the voltage artifact at the recording channel input. This intermediate step can be removed by changing the filter input into a discrete delta function. The FIR coefficients can be trained to a time-domain representation of the artifact voltage waveform, with several added benefits: most multiplicands become unity, the updated hardware can be multiplexed, and the tapped delay line can be removed.

With a discrete delta input signal propagating through the FIR filter, only one tap of the filter is active at a time. This removes the need for a sum at the filter output. In addition, the multiplication of each tap coefficient with the delayed filter input, \( x(n) \), is replaced with multiplication by zero or unity as the discrete impulse input propagates through the filter. For \( N \) samples after stimulation onset at \( n_0 \), the filter simplifies to

\[
y(n) = \sum_{i=0}^{N} c_i \delta(n - n_0 - i) = c_{n-n_0}, \quad 0 \leq n - n_0 \leq N. \tag{10}
\]

In this implementation, the same LMS update algorithm in (9) updates one coefficient at a time. Given the high-speed logic available in a scaled CMOS and the low sampling frequencies necessary for adequately representing neural signals, it is possible to time-domain multiplex one set of LMS update hardware between all the filter taps. Furthermore, the time-domain multiplexed recording architecture lends itself well to the multiplexing of filter hardware between multiple stimulation–recording channel pairs. A representation of this multiplexing is shown in Fig. 9. While the filter coefficients are adapting, the scaling of the error signal used in the LMS adaptation is simplified to a division via a bitwise shift. The update operation first rotates through the recording channels before progressing down the time-domain taps of the impulse response filter, as indicated by the helical spiral in Fig. 9. Somewhat noteworthy is the fact that although the adaptation hardware is being reused between different channels, the filter response for each stim-s’ sense combination is adapted uniquely. Each artifact produced at one stimulation site propagates to each recording site and has a unique artifact response and a dedicated set of coefficients. The distribution of the unique stored artifact values is shown in Fig. 10(b).

The filter can be further simplified to remove the tapped delay line. Only one tap is active at a time, so the only functionality required is the storage and recall of filter values. This article uses a writable lookup table to replace the FIR filter, removing the power-consuming delay line. An integrated SRAM stores the filter values. The artifact samples, \( y_{R,S}(n) \) in Fig. 10, are equivalent to the coefficients in (10), \( c_{n-n_0} \). The artifact samples are indexed by the filter tap number, \( n \), recording channel, \( R \), and stimulation channel, \( S \). The previously stored coefficient for a given index is read out of

---

![Fig. 8. Behavioral block diagram of a full LMS-based artifact canceller.](image-url)
Fig. 9. Demonstration of the artifact canceller multiplexing in the context of a full FIR-based adaptive filter.

Fig. 10. (a) Implemented SRAM lookup-based artifact canceller architecture. (b) Translation of SRAM values to multiple channels and time samples.

Fig. 11. System verification and measurement setup. Emulated electrode loads couple test signals to the on-chip recording inputs along with artifacts generated by the on-chip stimulator. The electrode model is altered to dc bias the recording input to ground; the 5-GΩ resistors are shunted to the ground rather than being parallel with the model capacitance.

The four stimulators operate independently, so that multiple artifacts can overlap on the same recording channel. In order to accommodate the overlapping artifacts, the canceller hardware is physically multiplexed for the four stimulators, while it is time-domain multiplexed across recording channels. This article implements four separate canceller back-ends, one for each stimulator, with each time-domain multiplexed between four recording channels. The CDAC sums all the four canceller outputs before subtracting them from the input. This operation assumes that overlapping artifacts linearly superimpose.

When processing recording data at 2 kS/s for 16 stim-sense pairs with 40 stimulation pulses per second, the canceller dissipates a total of 780 nW, or \( \sim 50 \, \text{nW per channel} \). On the contrary, the full LMS method in [18], with multipliers for each filter tap, dissipates 910 nW per channel.
The adaptive artifact cancellation hardware readily scales with the number of recording channels by simply expanding the amount of SRAM integrated on a chip. This implementation includes reconfigurable on-chip memory for 16 possible stim-sense artifact combinations, each with 32 10-bit taps. The 5120-bit low-voltage custom SRAM occupies an area of 250 μm × 300 μm. The canceller can also interface with the off-chip memory through source-synchronous serialized communication.

The artifact cancellation depth is limited by the digital-to-analog interface. In this article, a 10-bit CDAC gives a possible 60 dB of cancellation with a ±125-mV full scale. Additional off-chip post-processing can achieve an increased cancellation depth above the 60 dB provided by this chip.

V. Measurement Results

The proposed system was fabricated in a TSMC 65-nm low power (LP) 1P9M CMOS and evaluated on both the bench and in vivo. These in vivo measurements were carried out on ketamine-sedated macaque monkeys with Utah (Blackrock Microsystems) or microwire electrode arrays chronically implanted in the motor cortex. Benchtop measurements involved emulating the electrode–tissue load using discrete resistors and capacitors. Resistive dividers translate the stimulator output voltage to a differential artifact voltage, coupled into the recording inputs with various test tones. The bench test configuration is shown in Fig. 11. The impedances were selected to replicate a microwire electrode array.

The resonant stimulation driver topology had a measured voltage compliance of ±11-V while driving up to 2 mA of output current. The stimulator charge pumps have a measured power efficiency of 31% while delivering maximum output current (2 mA, 11 V). A bench demonstration of the multi-channel stimulation and the ability of this chip to deliver programmable current shapes are shown in Fig. 12. Details of a rectangular stimulation pulse into a resistive load are shown in Fig. 13. The 2-mA biphasic pulses are driven through a 5-kΩ load impedance with pulse widths of 10 μs, separated by 5 μs. The ripple from the comparator-based voltage regulation loop can be seen during the constant-current pulse.

Fig. 14 shows an example of the artifact filter convergence and cancellation of full-scale artifacts with bench recordings at 2 kS/s. A ±125-mV artifact was generated at 40 pulses/s using the on-chip stimulator driving an emulated electrode–tissue load (as shown in Fig. 11). This artifact was combined with a 10-μV, 50-Hz test tone. Fig. 14(a) shows the full convergence progression. Fig. 14(b) shows the saturated artifact and the beginning of filter training during the first 200 ms after reset. Fig. 14(c) shows the recording and filter output after convergence. The power spectral density (PSD) plot in Fig. 14(d) shows signal integrity after artifact cancellation through the

Authorized licensed use limited to: University of Washington Libraries. Downloaded on October 13,2020 at 00:15:12 UTC from IEEE Xplore. Restrictions apply.
Fig. 16. Measured *in vivo* LFP recordings (a) without stimulation-evoked potentials, (b) with stimulation-evoked potentials and artifacts, and (c) with stimulation-evoked potentials and cancelled artifacts.

preservation of the 10–μV, 50-Hz test tone in the presence of full-scale artifacts.

Fig. 15 shows the convergence with artifacts from two independent stimulators on the same recording channel. Artifacts identical to those in Fig. 14 are applied at \( t = 0 \), and a second stimulator is activated at \( t = 1 \text{s} \). The second stimulator applies pulses to create ±75-mV artifacts at 40 pulses/s. This is the same rate as the first set of artifacts, and a 4-ms delay ensures that the artifacts overlap for 50% of their duration. The filter converges to an optimum solution after 5 s, as shown in Fig. 15(a). Details of the artifacts are shown in Fig. 15(b). This demonstrates the ability to simultaneously cancel artifacts from multiple stimulators with the on-chip cancellers.

The measurement shown in Fig. 16 illustrates artifact suppression *in vivo*. The recordings are shown of local field potentials (LFPs) in the motor cortex of a sedated non-human primate with a bandwidth of 10 Hz–1 kHz. The recordings are shown without stimulation, during stimulation without artifact suppression, and during stimulation with artifact suppression. Biphasic, differential stimulation of ±150 μA at five pulses per second was applied to differential electrodes 2 mm away from the differential recording electrodes on the same array. The ±50-mV differential artifacts were observed at the recording channel input via the oscilloscope, which the canceller suppresses to within ~100 μV at the recording amplifier’s input. Note that the artifacts in Fig. 16(b) are subject to the recording front-end saturation, reducing the post-recording amplitude to ~±2 mV.

In an *in vivo* recording environment, the canceller distinguishes between neural signals and artifacts by only locking on to signals correlated with the periodicity of stimulation and limiting the length of its adaptive response. Uncorrelated neural activity appears as random noise to the LMS algorithm over many samples. Neural responses to stimulation are correlated, and cancellation of these signals is avoided by limiting the length of the adaptive filter in time to cover only the stimulus artifact.

Fig. 17 shows the bench testing of artifact cancellation while recording at 16 kS/s. Synthesized neural “spikes” are coupled into the recording input with voltage artifacts generated from an on-chip stimulator. The control signal is a series of 50–μV neural spikes at 100 spikes/saeec from a Coulburn Bio-Signal Calibrator. The stimulation artifacts are ±125 mV in amplitude at 77 pulses/s, avoiding co-periodicity. Fig. 17(a) shows the recording measurements under the following conditions: 1) transient signal recording without stimulation; 2) with stimulation and artifact cancellation disabled; and 3) artifact cancellation hardware enabled and post-processing for further cancellation. The residual signal after post-processing is non-correlated stimulator noise. Fig. 17(b) shows the corresponding PSD profiles for 16-kS/s neural spike recordings in (a).

Fig. 18. Measured stimulator power efficiency for a range of dc output currents.
TABLE I
COMPARISON WITH THE STATE–OF–THE–ART SOLUTIONS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology / Voltage</td>
<td>HV180nm</td>
<td>HV180nm</td>
<td>130nm 3.3V</td>
<td>65nm 1V</td>
<td>65nm 0.8V</td>
<td>130nm</td>
<td>65nm 1.2/2.5V</td>
</tr>
<tr>
<td>Artifact Suppression</td>
<td>None</td>
<td>Fast Recovery</td>
<td>None</td>
<td>77dB</td>
<td>92dB Fast Recovery</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>System Power (mW)</td>
<td>18</td>
<td>0.7</td>
<td>1.07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.62</td>
</tr>
<tr>
<td>Chip Area (mm²)</td>
<td>25</td>
<td>11.52</td>
<td>4.98²</td>
<td>5.14</td>
<td>1</td>
<td>11²</td>
<td>4</td>
</tr>
<tr>
<td># of Ch.</td>
<td>16 SENSE 40 STIM</td>
<td>64 SENSE 4 STIM</td>
<td>64 SENSE 2 STIM</td>
<td>16 SENSE</td>
<td>64 SENSE 4 STIM</td>
<td>64 SENSE</td>
<td>4 STIM</td>
</tr>
<tr>
<td>Stim Compliance</td>
<td>±12V</td>
<td>±12V</td>
<td>3.1V</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>±11V</td>
</tr>
<tr>
<td>Efficiency</td>
<td>From Supply</td>
<td>From Supply</td>
<td>From Supply</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>31%</td>
</tr>
<tr>
<td>Area/Ch (Stim, mm²)</td>
<td>-</td>
<td>-</td>
<td>w/ Sense</td>
<td>-</td>
<td>-</td>
<td>w/ Sense</td>
<td>0.36mm²</td>
</tr>
<tr>
<td>Area/Ch (Rec, mm²)</td>
<td>-</td>
<td>-</td>
<td>0.013</td>
<td>0.18</td>
<td>0.024</td>
<td>0.018</td>
<td>0.0025</td>
</tr>
<tr>
<td>Ch. Power (Rec, μW)</td>
<td>5.4</td>
<td>8</td>
<td>0.63</td>
<td>2.7</td>
<td>0.8</td>
<td>0.79</td>
<td>3.21</td>
</tr>
<tr>
<td>IRN (μWرم)</td>
<td>7.68</td>
<td>1.6</td>
<td>1.13</td>
<td>8.2</td>
<td>0.73</td>
<td>2.1</td>
<td>2.9²</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>7kHz</td>
<td>500Hz</td>
<td>500Hz</td>
<td>8.3kHz</td>
<td>5kHz</td>
<td>500Hz</td>
<td>Tunable &lt;32kHz</td>
</tr>
<tr>
<td>Max DM Range</td>
<td>-</td>
<td>100mVpp</td>
<td>13mVpp</td>
<td>200mVpp</td>
<td>260mVpp</td>
<td>Rail-Rail</td>
<td>110mVpp</td>
</tr>
<tr>
<td>ADC Type</td>
<td>10b Pipeline</td>
<td>ΔΣ OSR 1024</td>
<td>ΔΣ OSR 100 ENOB 11.7</td>
<td>10b Nyquist</td>
<td>ΔΣ OSR 32 ENOB 10.7</td>
<td>ΔΣ OSR 10k ENOB 10</td>
<td>Nyquist Δ- Encode ENOB 14</td>
</tr>
</tbody>
</table>

▼Includes stimulation power ▲Reported area includes Wireless Power/TX or DSP †AFE dynamic range □Calculated over 1-1kHz BW

VI. CONCLUSION

A complete BBCI system with simultaneous recording and stimulation capabilities has been optimized for area efficiency and scalability in advanced CMOS technology (65-nm LP). The demonstrated chip integrates the HV stimulation on the same substrate as a LP recording front-end and digital computation. This scalable chip architecture enables uncorrupted neural recording on many channels during the full-amplitude stimulation.

A comparison with the state-of-the-art solutions is shown in Table I. This article condenses the stimulation power generation into a minimal silicon area, when compared to other works which require external HV supplies or charge pumps. The HV stimulation power train is integrated on the same standard CMOS substrate as a high-efficiency digital back-end utilized for closed-loop artifact cancellation, opening the way for multi-application single-chip BBCIs.

REFERENCES


Venkata Rajesh Pamula (Member, IEEE) received the B.Tech. degree in electrical engineering from IIT (BHU) Varanasi, Varanasi, India, in 2007, and the M.Sc. degree in electrical and electronics engineering from Imperial College London, London, U.K., in 2010.

From 2013 to 2017, he was a Research Assistant with MICASIESAT, KU Leuven, Leuven, Belgium, in collaboration with IMEC, Leuven. He is currently a Visiting Research Scientist with the Processing Systems Laboratory (PsyLab), University of Washington, Seattle, WA, USA. His research interests include biomedical circuits, low-power sensor circuit design, and hardware security circuits.

Mr. Pamula received four gold medals by IIT (BHU) Varanasi in 2007. He was a recipient of the Analog Devices Outstanding Student Designer Award in 2016.

Eric P. Pepin received the B.S. and M.S. degrees in electrical engineering from the University of Washington, Seattle, WA, USA, in 2012 and 2015, respectively.

He is currently an RFIC Design Engineer with Space Exploration Technologies (SpaceX), Redmond, WA. In addition to neural interfaces, his interests include RF/millimeter-wave beamforming integrated circuits and phased-array antennas.

Mr. Pepin received the Outstanding Student Designer Award from Analog Devices in 2014.

Steve Perlmutter received the Sc.B. degree in biomedical engineering from Brown University, Providence, RI, USA, in 1979, the M.S. degree in biomedical engineering from the University of California, Los Angeles (UCLA), Los Angeles, CA, USA, in 1982, and the Ph.D. degree in physiology and neuroscience from Northwestern University, Evanston, IL, USA, in 1991.

He is currently a Research Professor with the Department of Physiology and Biophysics, University of Washington, Seattle, WA, USA, where he is also a Research Affiliate with the Washington National Primate Research Center and a member of the Center for Neurotechnology and the University of Washington Institute for Neuroengineering. His research interests include spinal control of voluntary movements, neural plasticity, and neuroprosthetics. His lab is developing therapies for spinal cord injury and stroke that use activity-dependent, targeted, electrical, and optical stimulation of the nervous systems.

Vivesh Sathe (Member, IEEE) received the B.Tech. degree in electrical engineering from IIT Bombay, Mumbai, India, in 2001, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of Michigan, Ann Arbor, MI, USA, in 2004 and 2007, respectively.

From 2007 to 2013, he served as a Technical Staff Member with the Low-Power Advanced Development Group, AMD, where his research focused on inventing, developing, and productizing new technologies into next-generation microprocessors, including low-power circuit design, high-speed circuits, adaptive clocking for supply noise mitigation, and resonant clocking. In 2010, he led the research and development effort which resulted in the first resonant-clocked commercial microprocessor. He joined the University of Washington, Seattle, WA, USA, in 2013, where he currently serves as an Assistant Professor. His current research interests lie in areas of digital, mixed-signal, and power-management circuits and architectures.

Dr. Sathe was a recipient of the NSF CAREER Award. He has served as a Guest Editor for the Journal of Solid-State Circuits and as a member of the Technical Program Committee for the Custom Integrated Circuits Conference.

Jacques Christophe Rudell (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1994, and the M.S. and Ph.D. degrees in electrical engineering from University of California, Berkeley (UC Berkeley), Berkeley, CA, USA, in 1997 and 2000, respectively.

After completing his PhD, he worked as an RFIC Designer with Berkana Wireless (now Qualcomm), San Jose, CA, USA, and Intel Corporation, Santa Clara, CA, USA, for several years. In January 2009, he joined the University of Washington, Seattle, WA, USA, as a Faculty Member, where he is currently an Associate Professor of electrical engineering.

He is an Active Member with the Center for Sensorimotor Neural Engineering (CSNE), an NSF Engineering Research Center (ERC) based at the University of Washington, and is also the Co-Director of the Center for Design of Analog-Digital Integrated Circuits (CDADIC). His research interests include topics in RF and millimeter-wave integrated circuits design for communication systems, in addition to biomedical electronics for imaging and neural interface applications.

Dr. Rudell received the Demetri Angelakos Memorial Achievement Award, a citation given to one student per year by the EECS Department, while a Ph.D. Student at UC Berkeley. He has twice been a co-recipient of the best paper awards at the International Solid-State Circuits Conference, the first of which was the 1998 Jack Kilby Award followed by the 2001 Lewis Winner Award. He received the 2008 ISSCC Best Evening Session Award and was a co-recipient of the 2011 and 2014 RFIC Symposium Best Student Paper Awards. He was a recipient of the 2015 NSF CAREER Award.

He served on the International Solid-State Circuits Conference (ISSCC) Technical Program Committee from 2003 to 2010 and the Radio Frequency Integrated Circuits (RFIC) Symposium Steering Committee from 2002 to 2013, where he also served as the 2013 General Chair. He currently serves on the technical program committees of the European Solid-State Circuits Conference (ESSCIRC) and the Custom Integrated Circuits Conference (CICC). He also serves as the Founding Seattle Chapter Chair for the Solid-State Circuits Society. He was an Associate Editor for the Journal of Solid-State Circuits from 2009 to 2015.