A 4.78mm² Fully-Integrated Neuromodulation SoC Combining 64 Acquisition Channels with Digital Compression and Simultaneous Dual Stimulation

Department of Electrical Engineering and Computer Science, University of California Berkeley, Berkeley, CA 94704, USA

Abstract
A 65nm CMOS 4.78mm² integrated neuromodulation SoC consumes 417μW from a 1.2V supply while operating 64 acquisition channels with epoch compression at an average firing rate of 50Hz and engaging two stimulators with a pulse width of 250μs/phase, differential current of 150μA, and a pulse frequency of 100Hz. Compared to the state of the art, this represents the lowest area and power for the highest integration complexity achieved to date.

Introduction
Brain Machine Interfaces (BMI) have the potential to revolutionize our understanding of the complex workings of the brain and to immeasurably enhance the quality of life of patients with debilitating nerve damage. Increasing the lifetime of the implant [1] requires next-generation interfaces to offer a large number of channels for simultaneous neural data acquisition and stimulation while being untethered, microscopic, and bandwidth and energy limited. In this paper, we present an integrated neuromodulation system (Fig. 1) that combines 64 channels of real-time neural recording with on-chip compression and dual stimulation on 8 selectable channels without any off-chip components, hence offering substantially enhanced functionality over current state of the art [2-4]. The ultra-low area and power of this IC pave the road for fully-implanted, wirelessly-powered, high-density, neural interfaces.

Neural Signal Acquisition
Recording channels must provide adequate noise performance while consuming minimal power and area. This work achieves the lowest area per channel (0.0258mm² with biasing and digitization) for an AC-coupled design by 3x [5], allowing for integration with compression and stimulation in one IC. Performance is summarized in Figs. 2, 6: gain, bandwidth and noise/power performance are adjustable per channel, enabling power savings on high SNR electrodes.

Several methods are utilized to minimize overall power consumption. Each VGA drives a separate sampling capacitor to allow maximum settling time. A time-multiplexed SC amplifier (BUF) drives the 260fF ADC input capacitance. To minimize power consumption of this buffer, the ADC sampling window utilizes 22/32 of each conversion period, while successive approximation consumes the remaining 10/32 cycles. The BUF requires 1.2μA to settle at 8*20kHz.

The ADC utilizes bottom plate sampling. To avoid charge pumps and swings beyond the rails or a power-hungry \( V_{DD}/2 \) reference, half of the top plates are charged to \( V_{DD} \) and the other half are charged to GND during the sampling phase, effectively creating a \( V_{DD}/2 \) reference. The split array has 6 binary and 4 thermometer bits to reduce DNL to < 1bit. Custom 260aF MOM unit capacitors allow for a compact low power ADC.

Digital Compression
High-density implantable recording systems necessitate data reduction, specifically when combined with a data-rate constrained trans-cranial link. Prior state-of-the-art multichannel neural signal compression implementations [2-4, 6] have not addressed the level of integration and area-efficiency necessary for a long-term implantable system. A diagram of the digital back-end (Fig. 3) shows the entire interface to the block, with direct connections to the ADCs and a 4-wire interface. A nonlinear energy operator (NEO, [6]) based spike detector extracts spike events, enabling data reduction by only sending the 2.1ms spike window around an event (“epochs”), and/or spike counts in a 2.4-50ms programmable window. Sending epochs, spike rates, and uncompressed data (“streams”) can be enabled on a per channel basis. All packets are put into a clock domain crossing FIFO, which allows the system clock to operate at a different frequency than the output data rate, resulting in further power savings. Fig. 3 presents annotated power consumption, data rates, and compression ratios with an average firing rate of 50 Hz on each channel. Firing rates, which are sufficient for BMI control [7], provide the highest compression ratio of 700x. The 64 channel digital back-end occupies a total area of 0.675mm², a 2.95X reduction compared to previous state of the art [6] normalized to a 65nm technology, and is lower power than [2] which implements a simpler algorithm.

Stimulation
Two independent, differential, bi-phasic current stimulators (Fig. 4) can each be multiplexed onto four electrode pairs for a total of 8 unique stimulation sites. Each stimulator utilizes a single current source for both positive and negative stimulation phases, which mitigates the effects of current mismatch between phases and eliminates the need for calibration. The stimulators have configurable pulse length and current amplitude of >500μA (differential) with a 7μA LSB. The INL and DNL were measured to be 0.082 and 0.039 LSB respectively. The electrode pairs can also be used to drive LEDs for optogenetic stimulation.

The stimulator output common mode is set at mid-rail from a fully single-integrated switched-capacitor DC-DC converter. The DC-DC is implemented using a Dickson ladder topology with a maximum tunable input voltage of ~1.3V and a conversion ratio of 1:7. A maximum output voltage of 8.7V was measured and is limited by the NWELL/PSUB breakdown voltage of the process.

The stimulator implements an adiabatic, charge-recycling architecture without off-chip components, enabling a fully integrated system as opposed to state of the art [8]. A supply sensor dynamically tracks the charging electrode voltage and selects one of 7 supply voltages from the DC-DC in order to minimize power consumption over a stimulation cycle. A typical stimulation pattern of 300μA differential current with ~150μs per phase was performed on bench-top with a 1kOhm/10nF RC electrode model; the measured electrode
voltages and stimulator supplies are shown in Fig. 4. The measured input referred current supplied by the DC-DC over one cycle of operation is also shown in Fig. 4.

In Vivo Measurement Results
Extracellular recordings were performed using a 16-channel micro-wire array implanted in the visual cortex of an adult Long-Evans rat. A typical subset of the recorded raw data is shown in Fig. 5 as well as the time-aligned epochs, which were simultaneously recorded from that channel. In vivo stimulation was performed with a 210μA differential current for 125μs per phase. Stimulation artifacts are shown across multiple channels in Fig. 5, and the relative amplitude correlates with proximity to the stimulation site. The amplifier worst-case stimulation recovery time is <1ms.

Conclusion
The system was fabricated in TSMC 65nm LP CMOS and occupies 4.78mm² including pads; a die photo is shown in Fig. 5. The key metrics of the design are summarized in Fig. 6 and compared with the state of the art. The high integration level provides the next step in enabling fully-implanted wirelessly-powered high-density neural interfaces.

Acknowledgements
The authors thank TSMC University Shuttle Program for chip fabrication, BWRC sponsors, NSF, and BDA for support.

References

Figure 1. 64 recording channels, 2 stimulators, digital compression and power conditioning are all integrated on a single 4.78mm² IC.

Figure 2. (Bottom) Block diagram of the signal acquisition chain. (Top) Performance summary and noise spectral density.

Figure 3. (Left) The digital block diagram. (Right) Power, data rate and compression ratios under the 3 different modes of operation.

Figure 4. (Left) The schematics for the differential stimulator and on-chip SC DC-DC. (Right) Typical stimulation waveform recorded on benchtop depicting dynamic supply switching and the measured stimulator current consumption over one cycle.

Figure 5. (Left) A typical set of raw data recorded *In Vivo* and the set of epochs captured simultaneously over the full recording period. (Right) Die photo and *In Vivo* stim. artifacts on adjacent channels.

Figure 6. System summary and comparison table.