An Event-Driven SoC for Neural Recording

Song Luan†⋆, Yan Liu†⋆, Ian Williams†⋆, and Timothy G. Constandinou†⋆
†Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, UK
⋆Centre for Bio-Inspired Technology, Institute of Biomedical Eng., Imperial College London, SW7 2AZ, UK
Email: {s.luan, yan.liu06, i.williams10, t.constandinou}@imperial.ac.uk

***Abstract*—This paper presents a novel 64-channel ultra-low power/low noise neural recording System-on-Chip (SoC) featuring a highly reconfigurable Analogue Front-End (AFE) and block-selectable data-driven output. This allows a tunable bandwidth/sampling rate for extracting Local Field Potentials (LFPs) and/or Extracellular Action Potentials (EAPs). Realtime spike detection utilises a dual polarity simple threshold to enable an event driven output for neural spikes (16-sample window). The 64-channels are organised into 16 sets of 4-channel recording blocks, with each block having a dedicated 10-bit SAR ADC that is time division multiplexed among the 4 channels. Each channel can be individually powered down and configured for bandwidth, gain and detection threshold. The output can thus combine continuous-streaming and event-driven data packets with the system configured as SPI slave. The SoC is implemented in a commercially-available 0.35 μm CMOS technology occupying a silicon area of 19.1 mm^2 (0.3 mm^2 gross per channel) and requiring 32 μW/channel power consumption (AFE only).**

I. Introduction

Understanding how the trillions of action potentials (spikes) of the brain's billions of neurons produce our thoughts, perceptions, and actions is one of the greatest challenges of 21st century science. In recent years we have thus witnessed an increased appetite for new innovative neuro-technologies, in particular leveraging on modern microtechnology. The semiconductor economies of scale have enabled high density multichannel silicon probes and integrated electronics to extend the number of recording channels from 10s to 100s and soon 1,000s [1].

There are however several, often conflicting requirements that make implementing resource-optimised ASICs for neural recording particularly challenging:

1) Conventional neural recording systems stream digitised data to storage or a backend processor for neural signal processing (e.g. spike detection and sorting) [2]. As the number of recording channels increases, this method becomes inefficient in terms of power particularly when recording EAPs (≈3 kHz signal bandwidth). Here, power can be saved if only the useful spike data is being transmitted [3], [4].
2) There is increasing interest in recording LFPs for scientific purposes, but also for future brain machine interfaces (BMIs) [5]. As LFP signals are a relatively low frequency signal (≈200 Hz signal bandwidth), power can be saved potentially through reducing the sampling rate.
3) However, it is of great interest to observe the AP and LFP at the same time such that relations between them can be studied [6].

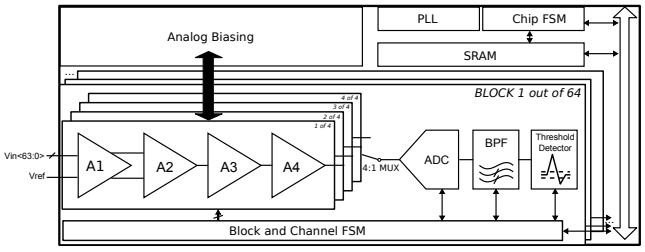

Fig. 1. System Architecture showing the 64-channels grouped into 16× 4-channel AFE blocks.

In the system presented, each channel can be individually configured to one of four different modes: (1) EAP streaming – sending out raw spike recordings; (2) EAP spike event driven – utilising spike detection to send out only the spike 'packets'; (3) LFP streaming – decimated to lower the data rate; (4) Combined LFP and EAP streaming.

The remainder of this paper is organised as follows: Section II describes the system architecture and outlines the top level function of the system components; Section III describes the function and design of the system in more detail; Section IV presents simulation results with emulated neural signals; and finally Section V summarises the specifications and discusses areas of future work.

II. System Architecture

A simplified block diagram of the system is shown in Fig. 1, including digital controllers, PLL, biasing, mixed-signal front-end array and SRAM. The digital controller communicates with an external controller via an SPI link, and generates an asynchronous handshake when data is waiting to be sent so as to request that the SPI link is activated.

The mixed-signal block contains an array of 64 Analogue Front-Ends (AFEs), organised in blocks of 4, with each block sharing a digital filter, threshold detector and 10-bit ADC. The AFE contains 4-stages of amplification organised as follows: (stage-1) provides low-noise high-gain; (stage-2) and (stage-3) fixed gain with programmable bandpass (for selecting between EAP and LFP); and (stage-4) programmable gain, such as to utilize the full dynamic range of the ADC. Within each block, the 4-channels are Time Division Multiplexed (TDM) via the Sample and Hold (S&H) circuit and 10-bit SAR-ADC that are controlled by the FSM – allowing for a tunable sampling rate. A 1^{st}-order IIR High-Pass Filter (HPF) removes the signal offset in addition to other common-mode interference, followed by spike detection. The in-block FSM handles the

TABLE I
DIFFERENT SUPPORTED RECORDING MODES

	Channel Enable	Detection Enable	IIR Enable	Record LFP	LPF Enable
Disable	0	X	X	X	X
LFP Stream	1	0	0	1	1
LFP + AP Stream	1	0	0	1	0
AP Stream	1	0	X	0	0
AP Spike Events	1	1	1	0	0

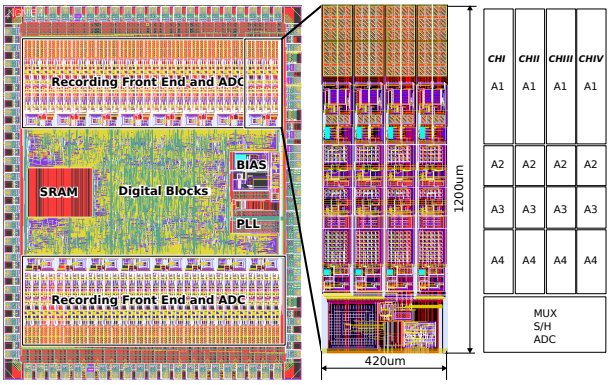

Fig. 2. Chip implementation in AMS 2P4M 0.35μm CMOS technology. Overall chip dimensions are 3.9 mm × 4.9 mm.

AFE and digital filter settings, and provides a data/command interface to the chip level FSM and SRAM.

Each block can be individually addressed and configured for different recording purposes, as listed in Table I. During streaming, the sampled data is directly read from memory and sent out continuously; in spike event mode, the detection results will determine whether there is a valid spike, which will trigger the entire spike window to be sent out. The analogue filter settings and in-channel logic are adjusted accordingly by the FSM. Individual channels can also be disabled to reduce data bandwidth and power consumption.

All recorded data is directly stored in the SRAM after conversion and filtering. During data output all 16 blocks share the SRAM and access the SPI under the administration of an internal scheduler. The chip-level FSM manages the data/command interface between the system and external controller via SPI, and latches the data transmission clock to blocks only when there is valid data.

III. CIRCUIT IMPLEMENTATION

A. Overall System

The system has been implemented in AMS 0.35 μm 2P4M CMOS technology with a total area of 19.1 mm^2 (0.3 mm^2 gross per channel). The layout and annotated floorplan for the entire system and individual blocks are shown in Fig. 2.

B. Analogue Front-End (AFE)

The 1^{st}-stage low noise amplifier is based on a source-degenerated folded-cascode topology [7], with common-mode feedback provided through triode-biased devices. The gain is set to 75 by the ratio of capacitances (input:feedback = 7.5 pF:100 fF), with a pseudo-resistor element connected in parallel with feedback capacitor. The capacitors have been implemented by combining overlapping Poly-Insulator-Poly (PIP) and Metal-Insulator-Metal (MIM) capacitors. The 2^{nd} and 3^{rd} stage amplifiers are based on a fully-balanced symmetric OTA topology, with programmable feedback networks as shown in Fig. 3. The feedback networks combine two pseudo-resistor topologies to provide alternative highpass poles. This uses a diode-connected PMOS pseudo-resistor to provide the lower (<1 Hz) corner frequency and a NMOS/PMOS pair biased in triode to provide a tunable higher corner frequency from 30 Hz – 600 Hz. The bias voltage for the NMOS/PMOS pair is in turn provided by a diode-connected NMOS/PMOS [8] with tunable current, controlled by a 5-bit current DAC. A LP filter with a selectable corner frequency (either 220 Hz or 5 kHz) is implemented in the 3^{rd} stage by adjusting the bias current to the input pair. The 4^{th} stage amplifier provides tunable gain by changing the capacitor ratio [9]. The transconductor uses a 2-stage Miller topology to provide sufficient speed to drive the following sampling and hold stage.

C. Analogue-to-Digital Converter (ADC)

A 4-to-1 multiplexer connects the required FEA analogue output to the sample and hold capacitor. This is connected to a single-ended 10-bit charge-redistribution Successive Approximation Register (SAR) ADC as shown in Fig 3.e. A split capacitor array with unit Poly-Insulator-Poly capacitors of 33fF is used to reduce the total capacitance with a split ratio of 9:1 [10]. A rail-to-rail comparator with dynamic biasing is used to accommodate for a wide analogue input range, whilst also saving power when idle. In order to save area and power, all the ADC were driven by the same trigger bus and operated at the same timing.

D. IIR HPF and Threshold Crossing Detection

The IIR filter is based on a direct form II realization with the adders and dedicated delay elements shared between each four input channels. Tuning of the highpass corner frequency is achieved by shifting a different number of bits, resulting in a tuning step of 6 dB over a 50 dB range. The threshold crossing detection is based on absolute value threshold, which can detect both positive and negative spike.

E. Spike Processing

Within each channel, a dedicated spike processing engine is implemented to achieve robust spike detection, based on the threshold detection results. This classifies a spike when three consecutive samples are above threshold. Then, a spike window is determined by allocating a preset number of samples before and after this validation point. In this design, we select 4 samples before and 12 samples after. A request signal is then generated after all samples have been stored, to inform the top level FSM to transmit the spike data off-chip. This is illustrated in Fig. 4.

If the channel is configured in streaming mode, this request signal is generated each time there is valid sample available. For LFP recordings a simple decimation scheme is implemented, i.e. data is output every eight samples.

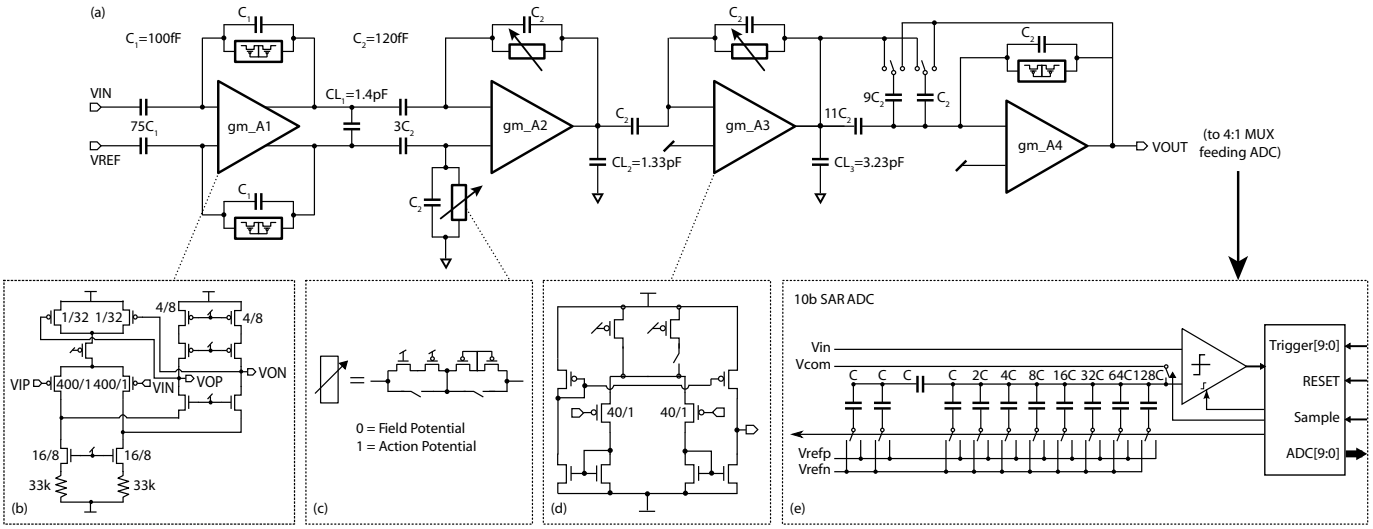

Fig. 3. Circuit schematic of the in-channel block. Shown are: (a) the Analogue Front-End (AFE); (b) The first stage gain; (c) tunable pseudo-resistor (d) The third stage gain; and (e) ADC. Key device dimensions are annotated as Width/Length in μm.

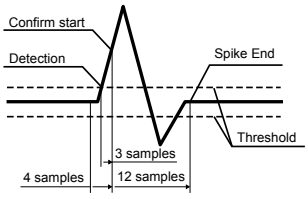

Fig. 4. Spike window with detection threshold for spike only output.

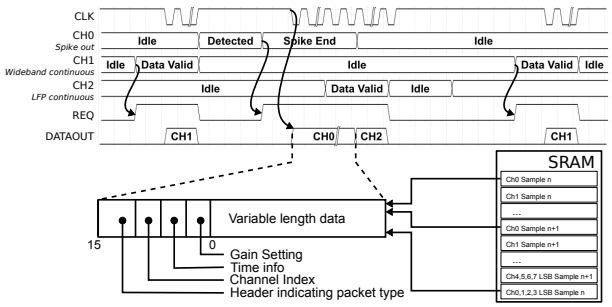

Fig. 5. Indicative timing diagram for data output, SRAM configuration and packet format.

F. Data Out and Packet Structure

The output data is then stored in SRAM, with format shown in Fig. 5. It should be noted that in order to utilise a standard SRAM (i.e. 8-bit word length), the LSBs of channels within the same block are stored separately from the MSBs. After the master controller receives and acknowledges the request signal, a clock signal for the data transmission is asserted. Then depending on the type of data being sent out (e.g. spike, stream, etc), an appropriate header and data packet is constructed, as shown in Fig. 5.

An internal scheduler is implemented to ensure only channels with a valid request are read out, in order of index. Furthermore, to avoid losing any timing information, a counter is used to track the latency between when the data is originally marked as ready, and when it is actually sent out.

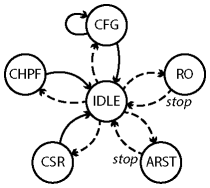

Fig. 6. State transition diagram for chip-level FSM. Dashed lines indicate transitions that are triggered by a command, with solid lines indicating automatic transitions.

TABLE II
SET OF COMMANDS FOR TOP-LEVEL FSM

Commands	States	Function
Analogue reset	ARST	Analogue reset all channels on chip or in a selected block or a selected channel
Change HPF corner	CHPF	Tune global HPF cut-off frequency
Change sampling rate	CSR	Change sample rate of the entire chip
Configure	CFG	Configure a selected bank
Enter readout	RO	Enter state for sending out recorded data
STOP	IDLE	Stop current operation and return to IDLE
NOP	any	NOP command

G. Chip-level Control

The top-level FSM has 6 states (see Fig. 6). A transition from idle state is always triggered by receiving a command from the external SPI master, as shown in Table II. After successfully receiving a command, the chip acknowledges this by replying with an inverted copy of the previous transaction such that the external controller can verify data integrity.

In ARST state, all AFEs are reset until a STOP command is received. This provides the flexibility when using probes with different impedances to control the settling times during reset. In CFG mode, the SPI command is followed by a 16-bit word to update the register within the selected block. The old register values are shifted back to the SPI master at the same time. It returns to IDLE state once all 64-bits are written. In CHPF state, the global HPF corner frequency can be tuned between 33.59 Hz to 507.8 Hz with a resolution of 5-bit. In

TABLE III
SUMMARY OF ACHIEVED (SIMULATED) SPECIFICATIONS

	EAP	LFP
Input-referred noise 300-3 kHz (μV_{rms})	1.92	
Input-referred noise 0.1-200 Hz (μV_{rms})	1.26	
Highpass corner (Hz)	33-600	0.02
Lowpass corner (Hz)	5.37 k-8 k	218-356
Gain	190-4254	190-4254
LSB resolution input-referred (μV)	0.65-13.33	0.65-13.33
CMRR	74 @1K	75 @10
PSRR	75 @1K	69 @10
THD	-39dB	-42dB
Analogue front-end power (μW)	23	24
Analogue front-end supply voltage (V)	3.0	
ADC ENOB	9.1	
ADC power (μW)	33.6	
ADC digital supply voltage (V)	3.3	

CSR state, the sampling rate for all the 16 ADCs can be configured between 13.245 kHz and 85 kHz with a resolution of 8-bit. In RO mode, the chip will send out recorded data corresponding to the configuration. Any data received from the external controller will be ignored in RO except for a STOP command which returns the FSM to IDLE state.

IV. Simulation Results

The simulated performance of the AFE and ADC are summarised in Table III. The implemented system has an input range from 0.7 mV to 15 mV without saturating the amplifiers and/or ADC, and a worst-case noise level less than 3 μV. The overall power consumption (per block) is 130 μW from a 3 V supply voltage.

A complete transient simulation of a typical spike waveform with four channels is shown in Fig. 7. This demonstrates the circuit operation within a single block of four channels during spike output mode. The threshold value for spike detection here is individually configured for each channel after first streaming out training data.

V. Conclusion

In this paper we presented a 64-channel fully-integrated low power neural recording ASIC with the ability to record both LFPs and EAPs in a variety of different capture modes. The design features a richly-configurable AFE with 4-stage amplification, in-channel analogue and digital filtering, data conversion, threshold detection and spike processing for a data-driven output. The circuit has been implemented in a commercially-available 0.35 μm CMOS technology, occupying a gross area of 0.3 mm^2/channel and requiring an average power consumption of 32 μW/channel.

Acknowledgement

This work was supported by UK EPSRC grant refs: EP/I000569/1, EP/K015060/1, and EP/M020975/1.

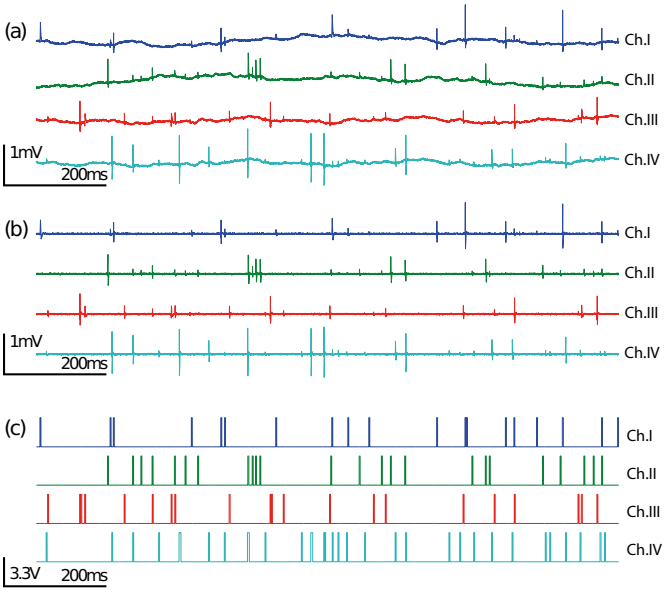

Fig. 7. Complete system transient simulation of the proposed system. Shown are: (a) raw neural recording including both the LFP and EAP; (b) filter output using a 200 Hz analogue HPF and a 100 Hz Digital HPF; and (c) spike output data packet.

References

[1] I. H. Stevenson and K. P. Kording, "How advances in neural recording affect data analysis," *Nature neuroscience*, vol. 14, no. 2, pp. 139–142, 2011. [Online]: http://dx.doi.org/10.1038/nn.2731

[2] D. Y. Barsakcioglu *et al.*, "An analogue front-end model for developing neural spike sorting systems," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 8, no. 2, pp. 216–227, 2014. [Online]: http://dx.doi.org/10.1109/TBCAS.2014.2313087

[3] A. Eftekhar *et al.*, "Towards a next generation neural interface: Optimizing power, bandwidth and data quality," in *Biomedical Circuits and Systems Conference (BioCAS), 2010 IEEE*, 2010, pp. 122–125. [Online]: http://dx.doi.org/10.1109/BIOCAS.2010.5709586

[4] I. Williams *et al.*, "Live demonstration: A scalable 32-channel neural recording and real-time fpga based spike sorting system," in *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*, 2015, pp. 1–5. [Online]: http://dx.doi.org/10.1109/BioCAS.2015.7348330

[5] T. M. Hall *et al.*, "A common structure underlies low-frequency cortical dynamics in movement, sleep, and sedation," *Neuron*, vol. 83, no. 5, pp. 1185–1199, 2014. [Online]: http://dx.doi.org/10.1016/j.neuron.2014.07.022

[6] T. M. Hall *et al.*, "Real-time estimation and biofeedback of single-neuron firing rates using local field potentials," *Nature communications*, vol. 5, 2014. [Online]: http://dx.doi.org/10.1038/ncomms6462

[7] W. Wattanapanitch *et al.*, "An energy-efficient micropower neural recording amplifier," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 1, no. 2, pp. 136–147, 2007. [Online]: http://dx.doi.org/10.1109/TBCAS.2007.907868

[8] M. Yin and M. Ghovanloo, "A low-noise preamplifier with adjustable gain and bandwidth for biopotential recording applications," in *Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on*, 2007, pp. 321–324. [Online]: http://dx.doi.org/10.1109/ISCAS.2007.378400

[9] X. Zou *et al.*, "A 1-v 450-nw fully integrated programmable biomedical sensor interface chip," *Solid-State Circuits, IEEE Journal of*, vol. 44, no. 4, pp. 1067–1077, 2009. [Online]: http://dx.doi.org/10.1109/JSSC.2009.2014707

[10] H. Gao *et al.*, "Hermese: A 96-channel full data rate direct neural interface in 0.13 m cmos," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 4, pp. 1043–1055, 2012.