A Multi-Channel Low-Power System-on-Chip for in vivo Neural Spike Recording

Andrea Bonfanti, Guido Zambra, Alessandro Sottocornola Spinelli, Andrea Lacaita
Dipartimento di Elettronica e Informazione, Politecnico di Milano

***Abstract* – This paper reports a multi-channel neural spike recording system-on-chip (SoC) with digital data compression and wireless telemetry. The circuit consists of a 64-channel low-power low-noise analog front-end, a single 8-bit analog-to-digital converter (ADC), followed by digital signal compression and transmission units. The 400-MHz transmitter employs a Manchester-Coded Frequency Shift Keying (MC-FSK) modulator with low modulation index. In this way a 1.25-Mbit/s data rate is delivered within a band of about 3 MHz. Compression of the raw data is implemented by detecting the action potentials (APs) and storing 20 samples for each spike waveform. The choice greatly improves data quality and allows single neuron identification. A larger than 10-m transmission range is reached with an overall power consumption of 17.2 mW. This figure translates into a power budget of 269 μW per channel, which is in line with the results in literature but allowing a larger transmission distance and more efficient wireless link bandwidth occupation. The implemented IC was mounted on a small and light printed circuit board to be used during neuroscience experiments with freely-behaving rats. Powered by 2 AAA batteries the system can work continuously for more than 100 hours allowing long-lasting neural spike recordings.**

I. Introduction

Advanced research in electrophysiology and behavioral neuroscience aims at better understanding brain operation by exploring the complex neural networks in more detail. These investigations are generating an increasing demand for wireless microsystems capable to record neural signals from a large number of implanted electrodes and to deliver data in real time to a remote processing unit. Moreover, as today's scientific instruments become tomorrow's medical devices, these systems are seen as a step towards devices for assisting humans with disabilities (Lebedev et al., 2006). However, as the number of electrodes increases a huge data throughput is generated calling for a corresponding increase of processing frequency, power and RF spectral bandwidth. To cope with this issue two design trends have been followed so far:

- drastically reducing the throughput, detecting just the occurrence time of action potential spikes (Harrison et al., 2007);
- pushing the throughput to the limits to preserve the entire data content, and either transmitting ultra-wide band pulses (UWB-IR) in the $3.1-10.6$ GHz with low spectral efficiency (Chae et al., 2008), or using pulse-width modulated (PWM) signals (Lee et al., 2010) to reach a better spectral efficiency at lower RF frequencies.

In (Chae et al., 2008) a 1-GHz band around 5 GHz is adopted to transmit short pulses and delivers up to 90-Mbit/s data rate for a 128-channel system, while in (Lee et al., 2010) a FSK modulation at 915-MHz carrier frequency transmits 5.5-Mbit/s equivalent throughput within 38-MHz band (14% of spectral efficiency) for a 32-channel device. In this frame our approach was intermediate. The idea of the presents work was to investigate whether data compression can be improved thus making possible to preserve the information needed for single neuron identification while keeping the throughput and the bandwidth occupation limited in a few MHz. In addition the transmission range was pushed well beyond the 1-m value reported in above mentioned works to make possible realistic *in-vivo* experiments. A valid compression algorithm is the key to reduce the system power consumption allowing long lasting experiments since the needed transmitted power is directly proportional to the bit rate and to the square of transmission distance (Liu et al., 2009).

II. System Desription

The system consists of a home-made wireless recording unit, a receiver built with off-the-shelf modules plus and a remote host complete of a graphical user interface (GUI) to allow neural signal visualization during *in-vivo* experiments.

A. Wireless recording unit

Figure 1 shows the block diagram of the integrated circuit (IC) which acts as recording and transmitting module. The chip, fabricated in $0.35-\mu$m CMOS AMS process, occupies a 3.1×2.7mm^2 area, and is descripted in details in (Bonfanti et al., 2010). The circuit consists of a 64-channel low-noise amplifier array, featuring a pass-band transfer function (tunable high-pass and 10-kHz low-pass cut-off frequencies) and about $3-\mu V_{rms}$ input referred noise. The amplifier outputs are sampled by a time-division multiplexer (TDM) at 20 kS/s onto one data lead before being further amplified by a variable-gain amplifier (VGA) and converted by an 8-bit successive approximation register analog to digital converter (SAR ADC). The data throughput at the ADC output is 10.24 Mbit/s. The digital data are then processed by a logic block, performing data compression, assembly and Manchester encoding of the bit stream. Data compression, that is briefly described in Sect. III, allows a reduction of data throughput by a factor of 10, generating a bit stream of 1.25 Mbit/s that is sent to the modulator. The transmitter consists of a voltage-controlled oscillator (VCO) directly modulated by the digital data. To squeeze spectrum occupation into 3-MHz bandwidth, a Manchester-coded FSK modulation with low modulation index (peak-to-peak frequency deviation of 800 kHz) has been adopted. In order to prevent drifts of the 400-MHz RF carrier

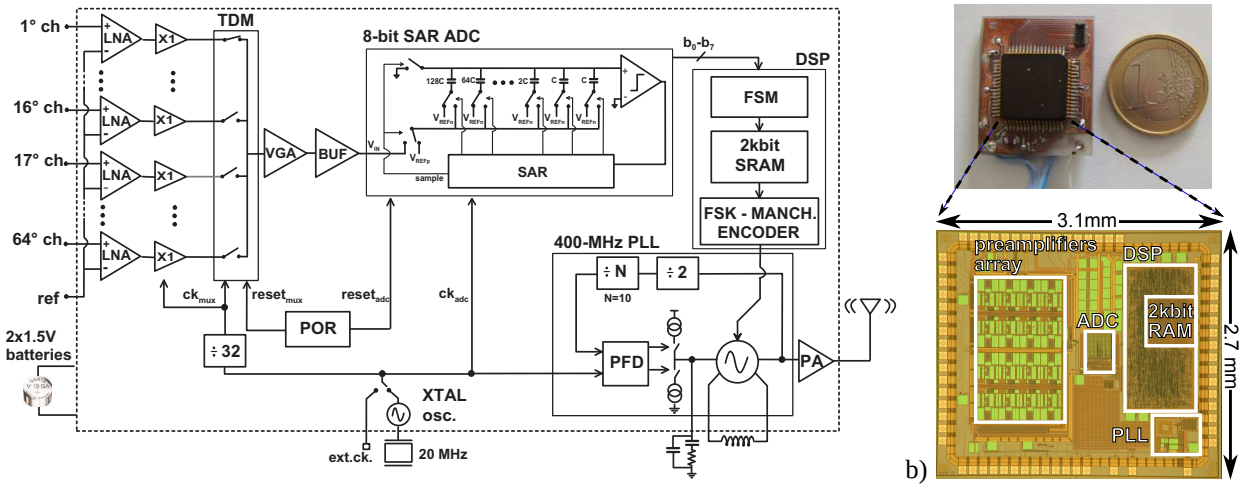

Fig. 1. Block diagram of the wireless neural recording integrated circuit (a) and picture of the headstage with the die microphotograph (b).

frequency due to temperature and supply variations, the VCO is inserted within an N-integer phase-locked loop, with a 50-kHz bandwidth set by an off-chip R-C filter (see Fig. 1). The transmitter is completed by an open-drain class-AB power-amplifier (PA) with external tapped-C resonant network and antenna. The PA is able to deliver an output power of 0 dBm with an efficiency of 12% to a 50-ohm antenna. A crystal Clapp oscillator employing an off-chip 20-MHz quartz and a power-on/reset (POR) circuit complete the system, providing a clean 20-MHz frequency clock and a synchronization signal at the start-up. The overall power consumption of the chip is 17.2 mW. 60% of power is due to the power amplifier that was sized to reach enough transmission range (larger than 10 m) not to pose issues during *in-vivo* experiments.

B. Receiver and graphical user interface

The receiver (see Fig. 2) consists of a quadrature zero-IF down-converter (MAX3580) using an LNA with a noise figure of 4.7 dB at 400 MHz, RF and baseband tunable filters and automatic gain control circuits. The quadrature signals are converted by a dual channel, 20 MSps, 10-bit ADC (AD9201) and fed to a Xilinx FPGA module (OpalKelly, XEM3005), which performs frequency demodulation and Manchester decoding and sends the received data to a PC via an USB link. Adopting a monopole antenna, the sensitivity of the receiver was estimated to be -74 dBm for a BER of 10^{-5} allowing a transmission range larger than 30 m in free space and larger than 10 m in a hostile environment, like a neuroscience laboratory. Moreover, a Graphical User Interface (GUI) based on Labview/C software was implemented to allow data saving, on-line elaboration and audio output generation.

III. Data compression algorithm

The implemented data reduction system takes advantage of the low duty cycle of the neural activity to eliminate the transmission of noise. Spike duration and firing rate are typical lower than 1 ms and 100 Hz, respectively. A finite state machine (FSM in Fig. 1) compares each incoming digital sample with a user programmable threshold. When the threshold is crossed, 20 samples of the signal registered on the same channel are recorded, covering a 1-ms time frame (Fig. 3). This strategy was chosen by taking into account that a clear identification of AP's requires the detection of three features, namely both peak and trough amplitudes together with the time interval between them (Vibert and Costa, 1979). Taking 20 samples makes possible to extract these features with enough resolution even for the fastest spikes. The samples are stored in an embedded memory together with the channel address and the timing stamp. Each spike is then compressed in 22 bytes. The SRAM is continuously read at 1-Mbit/s rate; the bit stream is completed by adding service bits for synchronization at the receiver and then encoded according to a Manchester code. The final data rate reaching the transmitter is 1.25 Mbit/s. The memory read speed was sized taking into account to aspects:

- the lower the read speed the higher the compression factor is. 1-Mbit/s read speed results in a compression factor of 10 (from 10.24-Mbit/s to 1.25-Mbit/s data rate) and allows saving a considerable amount of power and bandwidth at the transmitter side;
- considering the worst case scenario of 64 channels firing at 100 spike/s, 20 samples per spike with a resolution of 8 bit, the average data throughput is about 1 Mbit/s.

The memory was sized to reduce missed spikes when burst activity is present. Monte-Carlo simulations were performed considering that neural firing rate follows a Poisson statistic. Figure 4 shows the missing spike percentage as a function of the average firing rate per each channel, for different memory sizes. Opting for 1-Mbit/s read speed and 2-kbit RAM size, the missing spike percentage is less than 0.1% for 64 channels firing at 50 spikes/s and less than 10% for all channels firing at 100 spikes/s. With these choices, the memory is able to store spikes that occur in the same 1-ms frame in 11 consecutive channels before missing a spike. Each AP is transmitted, once completely stored, in 176 μs and thus the maximum latency time (time difference between detection and transmission of an AP) is less than 3 ms. The validity of the data-reduction algorithm was tested running a custom clustering software that employed fuzzy c-means spike sorter (Lewicki, 1998; Zouridakis et al., 2000) on a data set from a public source

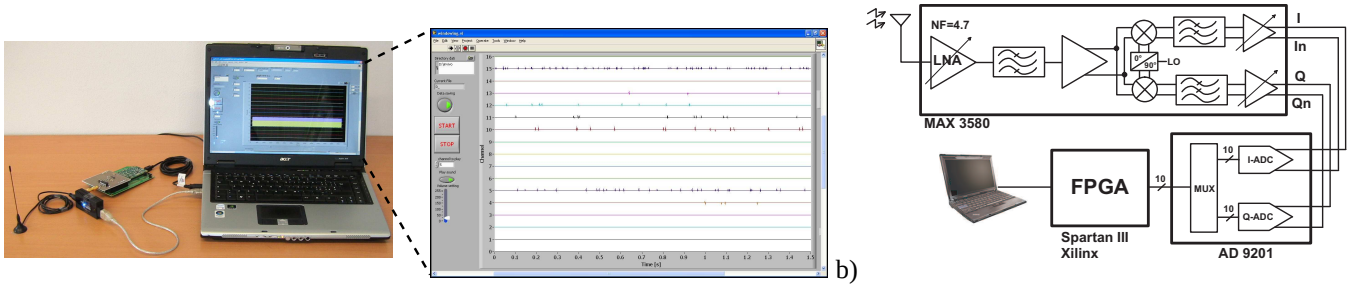

Fig. 2. Receiver unit with the detail of the graphical user interface (a) and its block diagram (b).

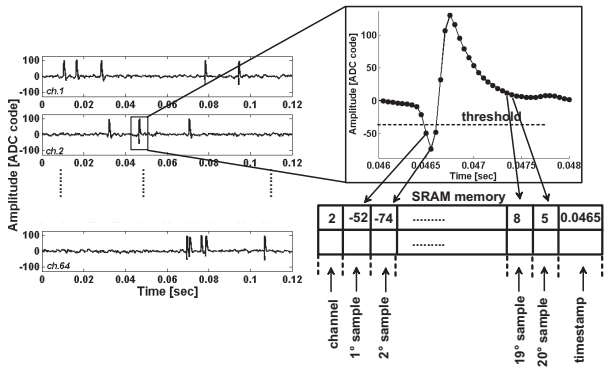

Fig. 3. Data compression algorithm: when a spike is detected, 20 digital samples are stored in the embedded memory, together with channel address and timing stamp information.

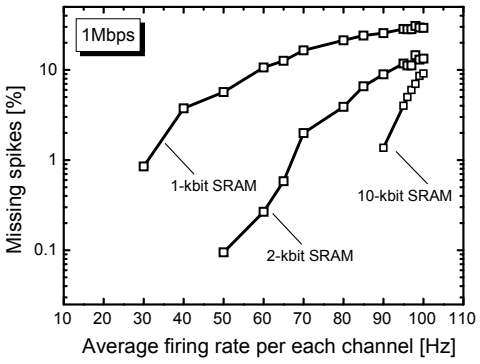

Fig. 4. Missing spike percentage as a function of average firing rate per channel for different memory sizes and for a read speed of 1 Mbit/s.

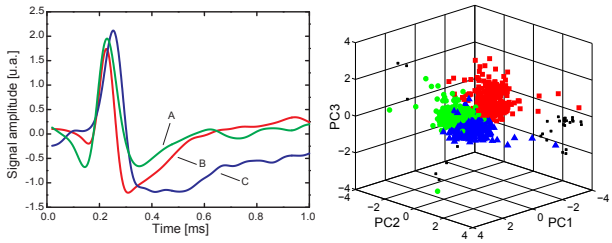

Fig. 5. Waveforms of neural spikes and PCA-based clustering results obtained processing data (difficult01.mat) from a public website (www.vis.caltech.edu/∼rodri).

(difficult01.mat from www.vis.caltech.edu/∼rodri) with spike waveforms hardly distinguishable one from each other (Figure 5). The ratio between the peak-to-peak spike amplitude and the rms noise was approximately 10. Clustering based on Principal Component Analysis (PCA) was performed on original data containing spikes from three different neurons and on the same data but reduced according to the above windowing strategy. Figure 5 shows that clustering on the reduced data set leads to clear separation of the three spike families. To be more quantitative, identification errors have been computed and the results are reported in Table I. Two types of error occur in the PCA-based clustering: type I errors occur when APs from two different neurons are grouped together (false positives), whilst type II errors occur when not all APs generated by one neuron are grouped together (false negatives). When applied to the original dataset, the PCA-based identification procedure leads to Type I/Type II mean error rates of $4.1\% - 5.2\%$, respectively. The reduced data sets showed $4.6\% - 5.3\%$ values, thus demonstrating that the reduction strategy preserves the quality needed for effective neuron identification. These results are not surprising considering that, as stated in (Vibert and Costa, 1976), more than 80% of neural spike information are carried by three analog features of the AP waveform, namely peak and through amplitudes and time width between peak and through. The results of spike sorting using these analog features (Bonfanti et al., 2008) and performed on the same trace is also reported in Table I. Note that the compression method adopted in the presented work always outperforms the clustering obtained using only these three features.

IV. In-vivo recording experiments

Figure 6 shows the system mounted on a rat for *in-vivo* signal acquisition. It consists of two parts: the wireless neural recording headstage and a backpack. To limit disturbs picked-up from the environment, a small board (3×3 cm^2 and 4.5g of weight) is directly connected to a 16-channel microelectrode array (Tucker-Davis), showing an impedance in the $20-60$ kΩ range at 1 kHz and implanted into the somatosensory cortex of an adult rat. The board mounts the packaged chip and 10 external small surface-mount components (0805 or 0603 footprint): the VCO inductor, the power-amplifier resonant filter components (2 capacitors and 1 inductor), the 20-MHz quartz, the PLL loop-filter (1 resistor and 2 capacitors) and 2 decoupling capacitors between power-supplies. The antenna is a quarter-wavelength whip antenna: it consists of a piece

Table I: sorting results

AP class (n^o of APs)	original		reduced		peak-trough-width	
	Type I	Type II	Type I	Type II	Type I	Type II
A (379)	4.1%	6.9%	3.1%	2.9%	23%	22%
B (389)	2.1%	3.3%	3.7%	6.9%	12%	18%
C (401)	6.2%	5.5%	6.9%	6.0%	17%	30%

Fig. 6. Arrangement for *in-vivo* measurements. Note the headstage, the battery back-pack and the whip antenna placed along the back of the animal.

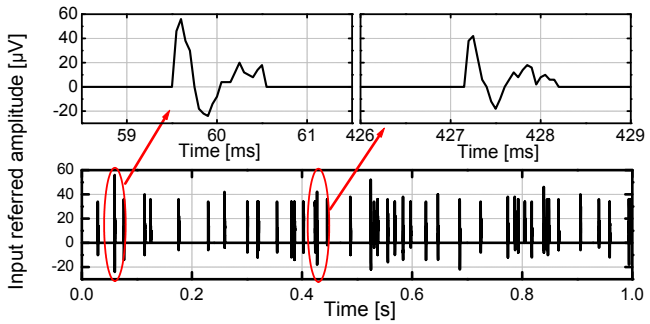

Fig. 7. Example of a reconstructed trace registered in a *in-vivo* experiment.

of wire about 17-cm long, easily placed along the back of the rat. The backpack has a weight of 40 g and includes two AAA batteries with 1000 mA/h capacity that may allow neural activity recordings for more than 100 hours. The two batteries are connected via three 10-cm long wires to the headstage. At the beginning of the *in-vivo* experiment, the quality of the signals from the implanted electrodes was analyzed using a commercial acquisition system recording the raw signals and noise from the 16 channels. The recorded traces featured a noise of about 10 μV_{rms} on all channels. The threshold of the digital peak processor was therefore set to an input-referred amplitude of ± 30 μV, i.e. ± 3 times the rms noise. The high-pass cut-off frequency of the front-end amplifiers was set to about 300 Hz to properly reject the low-frequency signals and the power-line noise, enabling correct spike detection. The gain of the overall amplifying chain was set to the maximum value (76 dB) since the peak-to-peak spike amplitude was always lower than 100 μV. The receiver monopole antenna was placed at a couple of meters from the rat and neural activity was detected on 3 of the 16 channels. Figure 7 shows a single trace registered during an experiment and the detail of two spike waveforms, while in Fig. 8 a collection of spikes registered on the same channel is aligned.

V. Conclusion

A single-chip neural interface employing digital data compression and wireless transmission was presented. The system is able to process the data coming from 64 channels and to reduce the data stream from 10.24 to 1.25 Mbit/s, i.e. with a 10x bandwidth reduction. The data compression algorithm

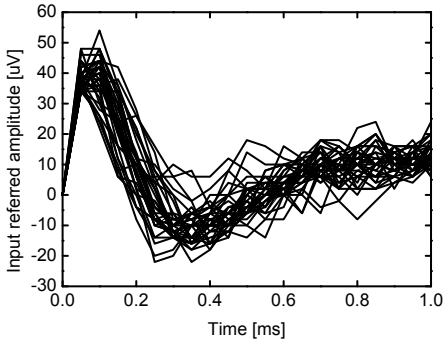

Fig. 8. Action potentials registered on a same channel and aligned.

allows the use of a narrowband FSK transmitter and to extend the transmission range to some meters with a reasonable power consumption. A comparison with the energy efficiency of other published neural wireless systems may be elaborated by taking into account that these latter have a transmission range of about 1 m. Scaling the transmitter power to get 1-m transmission range would result in a power consumption per channel of about 110 μW. This figure compares well with the 135 μW, 47 μW and 220 μW results reported for the systems in (Harrison et al., 2007; Chae et al., 2008; Lee et al., 2010), respectively.

Acknowledgment

The authors would like to thank G. Baranauskas, G.N. Angotzi, A. Vato, M. Semprini and E. Maggiolini for the *in-vivo* measurements performed at Department of Robotic, Brain and Cognitive Science of Italian Institute Technology (Genova), Prof. Luciano Fadiga for valuable discussions and Dr. A. Oleynjk for providing us the clustering software.

References

- Bonfanti A. et al., "A low-power integrated circuit for analog spike detection and sorting in neural prosthesis systems," *Proc. of BioCAS*, Baltimore (USA), pp. 257-260, Nov. 2008.
- Bonfanti A. et al., "A Multi-Channel Low-Power IC for Neural Spike Recording with Data Compression and Narrowband 400-MHz MC-FSK Wireless Transmission," *Proc. of ESSCIRC*, Sevilla (Spain) pp. 330-333. Sept. 2010.
- Chae M. et al., "A 128-channel 6 mW wireless neural recording IC with on the-fly spike sorting and UWB transmitter," *IEEE Int. Solid- State Circ. Conf.*, San Francisco (USA), pp. 146-148, Feb. 2008.
- Harrison R. R. et al., "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuit*, no. 1, pp. 123-133, 2007.
- Lee S.B., Lee H.-M. Kiani M., Jow U.-M., and Ghovanloo M., "An inductively powered scalable 32-channel wireless neural recording system-on-a-chip for neuroscience applications," *IEEE Int. Solid-State Circ. Conf.*, San Francisco (USA), pp. 120-121, Feb. 2010.
- Lebedev M. and Nicolelis M., "Brain-machine interfaces: past, present and future," *TRENDS in Neuroscience*, vol. 29, no. 9, pp. 536-546, 2006.
- Lewicki M. S., "A review of methods for spike sorting: the detection and classification of neural action potentials," *Network*, vol. 9, no. 4, pp.R53-R78, 1998.
- Liu Y.-H., Li C.-H. Lin T.-H., "A 200-pJ/b MUX-Based RF transmitter for implantable multichannel neural recording," *IEEE Trans. Micr. Th, Tech*, vol. 57, no. 10, pp. 2533-2541, 2000.
- Vibert J. and Costa J, "Spike separation in multiunit records: a multi-variate analysis descriptive parameters," *Electroenceophal. Clin.Neurophys.*, no. 47, pp. 172.182, 1979.
- Zouridakis G. and Tam DC., "Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering," *Comp. Meth. Prog. Biomed.*, vol. 61, pp. 91-98, 2000.