An auto-calibrated neural spike recording channel with feature extraction capabilities

Alberto Rodríguez-Pérez, Jesús Ruiz-Amaya, Manuel Delgado-Restituto and
Ángel Rodríguez-Vázquez

Institute of Microelectronics of Seville and University of Seville.
Avda. Américo Vespucio s/n. 41092 – Sevilla (SPAIN).
{alberto, ruiz, mandel, angel}@imse-cnm.csic.es

ABSTRACT

This paper presents a power efficient architecture for a neural spike recording channel. The channel offers a self-calibration operation mode and can be used both for signal tracking (to raw digitize the acquired neural waveform) and feature extraction (to build a PWL approximation of the spikes in order to reduce data bandwidth on the RF-link). The neural threshold voltage is adaptively calculated during the spike detection period using basic digital operations. The neural input signal is amplified and filtered using a LNA, reconfigurable Band-Pass Filter, followed by a fully reconfigurable 8-bit ADC. The key element is the ADC architecture. It is a binary search data converter with a SC-implementation. Due to its architecture, it can be programmed to work either as a PGA, S&H or ADC. In order to allow power saving, inactive blocks are powered off depending on the selected operation mode, ADC sampling frequency is reconfigured and bias current is dynamically adapted during the conversion. Due to the ADC low input capacitance, the power consumption of the input LNA can be decreased and the overall power consumption of the channel is low. The prototype was implemented using a CMOS 0.13 μ m standard process, and it occupies 400 μ m x 400 μ m. Simulations from extracted layout show very promising results. The power consumption of the complete channel for the signal tracking operations is 2.8 μ W, and is increased to 3.0 μ W when the feature extraction operation is performed, one of the lowest reported.

Keywords: Neural recording, spike detection, low-power, low-voltage, ADC, low-noise amplifier.

1. INTRODUCTION

The design of implantable neural recording microsystems for the wireless transmission of brain activity has focused the attention of many research groups during the last years¹⁻⁴. These microsystems can be very useful both in clinical (as part of therapeutic procedures in patients with neurological diseases) and neuroscience applications. Compared to wired systems, these solutions provide untethered communications with the external world, allowing for more freedom of movement and avoiding the risk of infections associated with percutaneous plugs.

Wireless connectivity implies two major challenges on the design of implantable biosensors. First, as microsystems are powered by small batteries or even rely on transcutaneous inductive links, the need for ultra-low power design strategies is mandatory. Second, the data rate that can be transmitted by wireless telemetry circuits is limited and, therefore, some data reduction mechanisms must be performed in the implanted device. This is particularly true in the case of microsystems mounted on Multi-Electrode Arrays (MEAs) in which, the transmission bandwidth requirement linearly increases with the number of recording channels. Then, the objective of these systems is to transmit the minimum information necessary about the recorded neural data that allows a reliable spike sorting processing. Most of the works reduce the channel data bandwidth by the transmission of just the information of the spikes after performing a spike detection processing at cell level¹⁻⁴. Although it introduces a high reduction in the amount of data transmitted, it can be insufficient in those cases where the system is integrated in a big MEA where the transmission channel is completely saturated. In these situations, a big reduction of the sent data is needed, and some additional spike feature extraction operations are added to the neural channel sensor³.

In this paper we present a low power neural spike recording channel suitable for raw brain activity acquisition at 22.5kS/s and on-chip detection/characterization of firing action potentials. This channel is intended to be part of a large multi-channel array and uses first-order Piecewise-Linear (PWL) approximations for spike feature extraction. This data

compression mechanism could be useful in cortically-controlled neuroprosthetics for spinal cord injured patients. The proposed solution employs mixed-signal techniques, and uses a foreground digital calibration process to counteract technology process variations and to adaptively calculate the voltage threshold voltage for the spike detection operation. The complete channel has been integrated in a CMOS standard 130nm technology and it occupies 400umx400um, thus fitting in the pitch of commercially available MEAs¹.

The paper is structured as follows. Section 2 describes the architecture of the channel and its operation modes, while Section 3 and Section 4 describe, respectively, two of its basic building blocks, the band-limited low noise amplifier and the data converter. Section 5 shows post-layout simulations of the recording channel and summarizes the achieved performance. Finally, Section 6 gives some conclusions.

2. RECORDING CHANNEL ARCHITECTURE

2.1 General architecture

Figure 1 shows the block diagram of the proposed neural spike recording channel. It consists of a band-limited fully-differential Low-Noise Amplifier (LNA), a reconfigurable Analog-to-Digital Converter (ADC) which embeds a Programmable Gain Amplifier (PGA), and some digital circuitry for control, spike detection and feature extraction purposes.

Figure 1: Architecture of the neural recording channel

Three different operation modes can be programmed in the recording channel. They are the calibration, signal tracking and feature extraction modes. Depending on the operation mode, idle blocks are total or partially powered off for power saving. Active blocks are operated by dedicated clock signals (derived from a master 1.6MHz reference) at frequencies that depend on the particular system configuration.

2.2 Calibration mode

Calibration is necessary in a multichannel neural sensor in order to counteract process variations, the difference between the signal levels of the electrodes and the unequal noise level of the neural signals. It means that in the proposed channel we need to calibrate both the low-pass and high-pass frequencies of the BP-LNA, the gain of the PGA, and the threshold voltage of the spike detection circuitry.

Due to the fact that each channel is affected by different process variation parameters and that the signal level of the sensed neural signals will depend on the position of the electrodes respect to the neurons, the calibration process has to be done on each channel separately.

We can distinguish between three different calibrations in the neural channel; the calibration of the PGA, the calibration of the LNA and the automatically detected threshold voltage for the spike detection^{5,6}. The first two are performed at the beginning of the cell operation and its calibration values are stored to be used during the rest of the operation of the neural channel.

The calibration process will be split in three steps, following the flow shown in Figure 2. First, the input of the system is connected to a frequency-controlled signal generator to calibrate the low-pass and high-pass frequencies of the BP-LNA. After that, the input of the system is switched to the electrode and the PGA is calibrated to maximize the dynamic range of the ADC. During the detection process the adaptive neural threshold detection algorithm is run as a background process. These calibrations are explained in detail in the following sub-sections.

Figure 2: Digital flow of the calibration process

BP-filter calibration

The objective of this calibration is to set the correct BP-band of the filter shown in Figure 1. As the positions of the zeros and the poles depend on the value of pseudo-resistors, (built using pMOS transistors) and capacitors, respectively, process variations produce a displacement on the position of the HP and LP filters. In order to compensate these shifts, we introduced calibration both in the pseudo-resistors (through 3-bits) and the capacitor (2-bits) to adjust their respective values.

In this step of the calibration flow we will adjust these control bits to set the BP-band at 200Hz-7kHz. For that, we will employ the frequency-controlled signal generator shown in Figure 3 as the input of our system. The output frequency of this signal generator is adjusted by the input clock signal of the block. This clock signal is used to control the switches that charge and discharge a load capacitor using a certain bias current, which is controlled in function of the desired output frequency. The circuit generates a triangular wave signal which is passed through an attenuator that divides its amplitude by 1000 in order to be used as the input of our neural signal interface.

The schematic of the signal attenuator is shown in Figure 3a. The attenuation is done by means of a resistive divisor, which is preceded by a buffer stage, as Figure 3b shows. The schematic of the charge pump can be seen in Figure 3c, and it is basically formed by a biasing circuitry which is adjusted in function of the desired frequency, and a load capacitor that is charged or discharged using the generated bias current.

The process flow of the BP-filter calibration is shown in Figure 4. First, the signal generator is programmed to work at the band-pass frequency (around 2kHz). Setting the BP-filter to the widest band, the PGA is calibrated to maximize the input dynamic range of the ADC. Once it is finished, the system stores the maximum value of the output signal under this configuration, that is, the reference of the BP-band.

The next step is the calibration of the LP-frequency. The signal generator is configured to create a 7kHz signal and the control bits of the LP-filter are configured to set the lowest frequency. The calibration is performed by comparing the maximum output value of the system at 7kHz input with the stored BP-band reference. If the drop of the first is equivalent or lower than 3dB (70%), the LP-frequency is correct and the calibration ends. In other case, the LP-frequency is rose to the next step and the comparison is performed again. This routine is repeated either until the drop

voltage is lower than 3dB with respect to the stored BP-band reference or the LP-frequency is configured at its widest value.

Figure 3: Schematic of a) Frequency-controlled signal generator, b) Attenuator, c) Charge pump

Figure 4: Digital flow of the BP-LNA calibration process

PGA calibration

The calibration of the PGA starts connecting the signal from the electrode directly to the input of the calibrated BP-LNA of the recording system. Then, the system follows the calibration flow shown in Figure 5, in which the gain of the PGA is calculated using a successive approximation algorithm similar to that of the SAR ADCs. The 3-bits of control are successively set starting with the most significant one and continuing with the rest until the least significant.

In this case, the system sets the MSB of the PGA gain control bits to '1', while the others remain to '0'. Then, the output of the recording channel is analyzed during a certain time in order to detect if the signal is saturated (it exceeds the 90% of the full-scale) or not. In the first case, the MSB is switched to '0'; otherwise, it is kept to '1'. The next iteration starts

switching the next bit to '1' and, hereinafter, the same analysis is performed. The calibration ends repeating the explained algorithm for the third bit.

Figure 5: Digital flow of the PGA calibration process

Adaptive neural threshold voltage

The threshold voltage of the neural signal for spike detection is continuously calculated during the operation of the neural spike recording channel.

The objective of the algorithm is to measure the mean of the noise level and multiply it by a certain factor in order to get the desired threshold voltage. There is a big number of works that present different solutions⁵⁻¹⁰. One of the methods that have been proven as efficient and easily implementable is the mean deviation (MD)¹⁰, which can be calculated as:

$$MD(x) = \frac{1}{N} \sum_{i=1}^N |x_i|$$

According to a published study⁵, it is more efficient to calculate a double threshold (positive and negative) rather than a global one. Then, the system evaluates if the input values are either negative or positive, calculates the mean using N elements for both cases and multiplies it by a factor of 4.5, which has been calculated as the optimum one⁶.

Both the number of elements in the summary and the multiplying factor are programmable in order to have a better control during the detection process.

2.3 Tracking mode

During the tracking mode, the signal captured by the neural electrode is conditioned by the LNA and 8-bits digitized by the ADC at a 22.5kS/s throughput rate. The output of the ADC is serially read out at a rate of 1.6MHz to allow for real-time external reconstruction of the neural data. This mode is always preceded by the calibration process in order to adjust the correct band-pass frequency of the filter and the gain of the PGA. As the spike detection is not performed during this mode, the adaptive threshold voltage calculation is disabled.

2.4 Feature extraction mode

This mode provides a data compressing mechanism to reduce the bandwidth of the information transmitted from the channel. Two phases can be distinguished in this operation mode which are referred to as spike detection and spike processing.

During spike detection, the ADC is configured to sample the neural data at 22.5kS/s. The digital logic detects if there is any spike in function of the calculated adaptive threshold voltage. Once the input voltage goes over the threshold voltage, the system assumes that there is a spike, switches the sampling rate of the ADC to 90kS/s and starts the spike processing phase. The objective of this phase is to create a PWL approximation biphasic peak by means of two time

segments (one for the peak position, Δ_1 , and the other for the position of the second peak, Δ_2) and the magnitude of these two peaks (V_{pmax} and V_{pmin}). Additionally, information about the threshold voltage (V_{th}) and the magnitude of the end of the spike, V_{end} (which is supposed to have a fixed duration from the first peak, given by a programmable variable t_{spike}) is also provided (see Figure 6). All the variables are codified into 8-bit digital vectors, which mean a total of 48-bits to characterize each spike. As the sampling rate of the ADC is four times larger during this mode than on the normal operation, only one of each four samples are taken to calculate the threshold voltage for spike detection.

Figure 6: PWL approximation performed by the feature extraction

3. LOW-NOISE AMPLIFIER

The LNA must be able to boost the weak signals detected by the microelectrodes and filter out the undesired frequency components¹¹⁻¹³. Our proposed LNA follows the capacitive feedback structure¹² but it introduces three main modifications for improved performance. First, it uses a fully-differential architecture, shown in Figure 7a, to increase the dynamic range of the amplifier and improve PSRR. Second, the Operational Transconductance Amplifier (OTA) uses a high output swing two-stage topology with feedforward capacitive compensation, shown in Figure 7b, to obtain a higher low-pass filter rejection of -40dB/dec. Third, the first stage of the OTA uses a complementary input differential pair to reuse the tail current and nearly double the achieved transconductance¹⁴.

The high-pass and low-pass frequency corners of the characteristic are 2-bits adjustable, by programming the value of the feedback resistors (control word [b2r,b1r]) and the load capacitor (control word [b2c,b1c]), respectively. Figure 7c shows the structure of the feedback resistors. They are formed by as many pseudo-resistors in series as determined by the 1-out-4 code $c = [c1, \dots, c4]$ defined, in turn, by the control word [b2r,b1r]. The schematic of the pseudo-resistors is shown in Figure 7c¹⁴. They are formed by eight MOS-bipolar devices in series to reduce distortion for large output signals. Mismatch effects and leakage currents have been carefully simulated to ensure feasibility.

Figure 8a shows the transfer characteristics of the LNA for different programming configurations. A mid-band gain of about 47.5dB is achieved by means of the capacitors ratio C_f/C_i . Figure 8b depicts the input-referred voltage noise spectrum for a nominal passband of 217Hz - 7.16kHz. Integration under this curve yields an rms noise voltage of 2.84uVrms. Considering that the power consumption of the amplifier is 1.92uW including CMFB circuits, the Noise Efficiency Factor (NEF)¹² of the LNA is 1.62 (1.75 when the biasing circuitry is also considered).

4. PGA AND ADC

The proposed ADC follows a binary search algorithm for data conversion but, instead of using a SAR architecture¹⁻³, it employs the SC structure of Figure 9a¹⁵. The circuit contains two active blocks, an S&H programmable gain amplifier/integrator and a comparator, as well as a conventional SAR register (not shown in the figure). Circuit operation is as follows. The signal is first sampled and stored in the integrator by transferring charge from capacitor C_{in} to C_{int} .

Figure 7: a) Schematic of the BP-LNA, b) Schematic of the OTA, c) Schematic of the programmable resistor chain

Figure 8: a) Transfer characteristic of the BP-LNA from post-layout simulations, b) Input-referred voltage noise spectrum of the BP-LNA

Afterwards, the conversion phase starts. It is simply realized by successively comparing the integrated voltage with $V_{ref}/2^n$, where V_{ref} is the full-scale reference voltage of the converter, and index $n = 1, \dots, N$, indicates the conversion step (N represents the converter resolution). The weighted voltages $V_{ref}/2^n$ are generated by a passive SC arrangement (shaded area in Figure 9a). If the integrated voltage is larger (alt., smaller) than $V_{ref}/2^n$ the comparator sets to '1' (alt., '0') the n -th bit of the conversion, and the integrator is updated by subtracting (adding) such increment. As illustrated in the timing diagram of Figure 9b, conversion takes $N+1$ clock cycles. Depending on the operation mode, the circuit is clocked at 100, 200 or 800kHz.

The value of the input capacitor C_{in} is 3-bits digitally programmable, so it implements a Programmable Gain Amplifier (PGA) during the sampling of the input voltage. The gain of the PGA can be adjusted from 0 to 18dB at discrete steps of 3dB (this gives a total gain for the channel front-end of 48-66dB, including the LNA contribution) by 3-bits programming the input capacitance.

Figure 10a shows the schematic of the fully differential OTA in the ADC+PGA. It is a folded-cascode topology with a SC-based common-mode feedback circuit (not shown). The OTA uses a dynamic biasing scheme in which the tail current of the block is adjusted according to the slew-rate demand (it decreases along the conversion process, as shown in Figure 9b) and the selected operation mode. This is accomplished by controlling the widths of transistors M3-M7.

Figure 10b shows the schematic of the comparator. It is a current-controlled dynamic-latch buffered with class-A amplifiers. As in the OTA, the bias current is adapted according to the selected sampling frequency.

Figure 9: a) SC-based reconfigurable ADC, b) ADC signal waveform

Figure 10: a) Schematic of the FD-OTA used in the SC-ADC, b) Schematic of the comparator

Figure 11: FFT spectrums of the SC-ADC at sampling rates of: a) 22.5kS/s, b) 90kS/s

Figure 11 shows two FFT analyses of the ADC at sampling rates of 90kS/s and 22.5kS/s obtained from a post-layout analysis. The SFDR and SNDR of the converter are approximately 62dB and 50dB along the whole Nyquist band, respectively. This corresponds to an ENOB of almost 8 bits. These features remain unaltered regardless of the gain

setting of the embedded PGA. The average power consumption of the ADC is 500nW at 22.5kS/S sampling mode and 1.8uW at 90kS/s.

5. POST-LAYOUT SIMULATION RESULTS

Figure 12 shows the layout of the complete neural spike recording cell. It has been implemented in a 130nm CMOS technology and occupies 400um x 400um.

Figure 12: Layout of the neural recording cell

Figure 14 illustrates the operation of the proposed neural recording channel in feature extraction mode. The plot at the top is an exemplary synthesized spike-controlled neural sequence which has been built using three different spikes from a real database as the measured local activity. The background noise is generated using a bigger database to simulate the behavior of the distant neurons¹⁶, while the local spikes are positioned using a Poisson distribution. The SNR of this 60 seconds generated sequence is 10dB.

Figure 13: Quality of the spike detection in function of the SNR of the input signal

The plots at the bottom illustrate the result of the spike sorting operation performed using the information provided by the system after the feature extraction. The spikes are sorted comparing their features by means of the Euclidean distance¹⁷. When a new spike comes, its Euclidean distance is evaluated with the rest of the stored spikes. If some of them are below a certain threshold, the spike is sorted with it. If not, it is classified as a new spike.

As can be extracted from the presented analysis, the proposed feature extraction method is able to provide the enough information to perform a spike sorting operation.

The performance of the spike detection operation and the adaptive neural threshold voltage is illustrated in Figure 13 through the probability of detection and false alarm for different SNRs. The first one is defined as the ratio between the number of true detected spikes and the total number of spikes introduced in the processed sample. The second one is defined as the ratio between the number of false spikes detected and the total number of spikes detected by the system. We can see that the performance of the spike detection algorithm is good while the SNR of the neural data keep above 0dB, with a high probability of detection and a low probability of false alarm.

Table 1 summarizes the performance of the presented neural spike recording channel. The power consumption depends on the selected operation mode. During signal tracking, only the band-limited LNA and the ADC with a throughput rate of 22.5kS/s are enabled, and the overall power consumption of the channel is 2.8uW, also including the dissipation of the biasing circuitry. During the feature extraction mode, the average power consumption is of about 3uW, assuming that the spike detection phase takes the 95% of the total evaluation time, while the remaining 5% is employed for spike processing.

6. CONCLUSIONS

A neural spike recording channel with feature extraction capabilities has been presented. The complete channel has been integrated in a CMOS standard 130nm technology and it occupies 400umx400um. The circuit is reconfigurable and offers different operation modes, including foreground self-calibration, adaptive threshold voltage, signal tracking and feature extraction using first-order PWL approximations. The performed feature extraction reduces the data bandwidth by nearly the 90% and simulation results confirm that the chosen PWL approximation reconstruction allows an effective spike sorting. Power-down strategies and frequency adaptation techniques are adopted in the chip for power saving. The channel consumes 2.8uW during signal tracking at 22.5kS/s, and 3.0uW average when performing feature extraction operation.

Table 1. Performance of the neural spike recording channel.

Operation Mode	Spike detection	Feature Extraction
Band-Pass Low-Noise Amplifier		
Input-referred noise	2.84uVrms	
Nominal passband	217Hz – 7.16kHz	
Low frequency corner	20-400Hz (3-bits tunable)	
High frequency corner	6.25-12.6kHz (2-bits tunable)	
THD (2mVpp input)	0.04%	
Power Consumption	2.17uW (including biasing circuitry)	
NEF	1.75 (including biasing circuitry)	
PGA and Analog-to-Digital Converter		
Sampling Rate	22.5kS/s	90kS/s
ENOB	7.98-bits	7.96-bits
SNDR	49.8dB	49.68dB
Power Consumption	500nW	1.8uW
Neural Recording Channel		
Process	Standard CMOS 0.13um	
Supply Voltage	1.2 V	
Overall Gain	47.5 – 65.5dB (3dB step)	
Power consumption	2.8uW	4.1uW
Die Area	0.0160mm ² (400um x 400um)

Figure 14: Spike sorting classification using the PWL approximation provide by the neural cell. The first row illustrates the raw generated spike-controlled data, while the second row shows the three different spike shapes used in the generated data. Third row plots the PWL approximation of the detected spikes and the last row shows the result of the spike sorting algorithm.

ACKNOWLEDGEMENTS

This work has been supported by the Spanish Ministry of Science & Innovation under grant TEC2009-08447, the Junta de Andalucía under grant TIC-02818 and the 2007-2013 FEDER Programme.

REFERENCES

- [1] R. R. Harrison, *et al.*, "A Low-Power Integrated Circuit for a Wireless 100-Electrode Neural Recording System," *IEEE J. Solid-State Circ.*, vol. 42(1), pp. 123-133, Jan. 2007.
- [2] A. M. Sodagar, *et al.*, "An Implantable 64-Channel Wireless Microsystem for Single-Unit Neural Recording," *IEEE J. Solid-State Circ.*, vol. 44(9), pp. 2591-2604, Sept. 2009.

- [3] M.S. Chae, *et al.*, "A 128-Channel 6mW Wireless Neural Recording IC With Spike Feature Extraction and UWB Transmitter," IEEE T. on Neural Systems and Rehabilitation Eng., vol. 17(4), pp. 312-321, Aug. 2009.
- [4] B. Gosselin, *et al.*, "A Mixed-Signal Multichip Neural Recording Interface with Bandwidth Reduction," IEEE T. Biomed. Circ. Sys., vol. 3(3), pp. 129-141. June 2009.
- [5] A. M. Kambow and J. A. Mason, "On-chip feature extraction for spike sorting in high density implantable neural recording systems," in Biom. Circ. and Systems Conf., BioCAS 2010, pp. 13-16.
- [6] S. Gibson, J. W. Judy and D. Markovic, "Technology-Aware Algorithm Design for Neural Spike Detection, Feature Extraction and Dimensionality Reduction," in IEEE Trans. On Neural Systems and Rehabilitation Eng., vol. 18(5), pp. 469-478. October 2010.
- [7] R.R. Harrison, "A low-power integrated circuit for adaptive detection of action potential in noisy signals," Proc. IEEE EMBS, Cancun, pp. 3325-3328, Sept. 2003.
- [8] P. H. Thakur, H. Lu, S. S. Hsiao and K. O. Johnson, "Automated optimal detection and classification of neural action potentials in extra-cellular recordings," J. Neuroscience Methods, vol. 162, pp. 364-376, 2007.
- [9] H. L. Chan, M. A. Lin, Y. T. Tsai and P. K. Chao, "Detection of neural spikes using an adaptive threshold based on the max-min spread sorting method," J. Neuroscience Methods, vol. 172, pp. 112-121, 2008.
- [10] D. L. Donoho, "De-noising by soft-thresholding," IEEE Trans. Inf. Theory, vol. 41(3), pp. 613-627, 1995.
- [11] B. Gosselin, M. Sawan and C. A. Chapman, "A low-power integrated bioamplifier with active low-frequency suppression," IEEE T. Biomed. Circ. & Syst., vol. 1, pp. 184-192, Sep. 2007.
- [12] R. R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," IEEE J. Solid-State Circ., vol. 38, no. 3, pp. 958-965, June 2003.
- [13] W. Zhao, H. Li, Y. Zhang: "A Low-Noise Integrated Bioamplifier with Active DC Offset Suppression", IEEE T. Biomed. Circ. & Syst., pp. 5-8, Nov. 2009.
- [14] S. Rai *et al.*, "A 500uW neural tag with 2uVrms AFE and frequency-multiplying MICS/ISM FSK transmitter," in IEEE Int. Solid-State Circuits Conference (ISSCC 2009), pp. 212-214, 2009.
- [15] A. Rodríguez-Pérez, *et al.*, "A low-power reconfigurable ADC for biomedical sensor interfaces," in IEEE Biomedical Circ. Sys. Conf. 2009, pp. 253-256, Nov. 2009.
- [16] J. Martínez, C. Pedreira, M. J. Ison and R. Q. Quiroga, "Realistic simulation of extracellular recordings," in Journal of Neuroscience Methods, vol. 184, pp. 285-293, 2009.
- [17] C. C. Peng, P. Sabhrwal and R. Bashirullah, "Neural Cache: A Low-Power Online Digital Spike-Sorting Architecture," IEEE EMBS Conference 2008, pp. 2004-2007.