

Article

A Multi-Channel Low-Power System-on-Chip for *in Vivo* Recording and Wireless Transmission of Neural Spikes

Andrea Bonfanti ^{1,*}, Maria Ceravolo ^{1,†}, Guido Zambra ^{1,†}, Riccardo Gusmeroli ¹, Gytis Baranauskas ², Gian Nicola Angotzi ², Alessandro Vato ², Emma Maggiolini ², Marianna Semprini ², Alessandro Sottocornola Spinelli ¹ and Andrea Leonardo Lacaita ¹

¹ Dipartimento di Elettronica e Informazione, Politecnico di Milano, Via Golgi 40, Milano 20133, Italy

² Italian Institute of Technology (IIT), Department of Robotic, Brain and Cognitive Science, Genova 16163, Italy

[†] Present address: STMicroelectronics, Via Tolomeo, Cornaredo 20010, Italy.

* Author to whom correspondence should be addressed; E-Mail: bonfanti@elet.polimi.it; Tel.: +39-02-2399-6146; Fax: +39-02-2367-604.

Received: 11 June 2012; in revised form: 6 September 2012 / Accepted: 21 September 2012 /

Published: 28 September 2012

Abstract: This paper reports a multi-channel neural spike recording system-on-chip with digital data compression and wireless telemetry. The circuit consists of 16 amplifiers, an analog time-division multiplexer, a single 8 bit analog-to-digital converter, a digital signal compression unit and a wireless transmitter. Although only 16 amplifiers are integrated in our current die version, the whole system is designed to work with 64, demonstrating the feasibility of a digital processing and narrowband wireless transmission of 64 neural recording channels. Compression of the raw data is achieved by detecting the action potentials (APs) and storing 20 samples for each spike waveform. This compression method retains sufficiently high data quality to allow for single neuron identification (spike sorting). 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 limited band of about 3 MHz. The chip is realized in a 0.35 μm AMS CMOS process featuring a 3 V power supply with an area of $3.1 \times 2.7 \text{ mm}^2$. The achieved transmission range is over 10 m with an overall power consumption for 64 channels of 17.2 mW. This figure translates into a power budget of 269 μW per channel, in line with published results but allowing a larger transmission distance and more efficient bandwidth occupation of the wireless link. The integrated circuit was mounted on a small and light board to be used

during neuroscience experiments with freely-behaving rats. Powered by 2 AAA batteries, the system can continuously work for more than 100 hours allowing for long-lasting neural spike recordings.

Keywords: neural recording; action potential; low-noise amplifier; wireless transmitter; bio-telemetry; FSK; Manchester-encoding

1. Introduction

There is a growing need for wireless implantable neural recording systems that can simultaneously acquire neural signals from a large number of channels. Recording from hundreds of neurons may allow neurophysiologists to study brain function in freely-behaving animals, and in future will provide the signals necessary to control neural prosthetic devices [1]. Compared to wired systems, wireless neural recording devices provide a number of benefits: reduction of motion artifacts and freedom of animal movements in neuroscience experiments, as well as reduced infection risks and higher patient mobility in neuroprosthetic applications. However, these systems entail severe technical challenges, such as the robustness of the wireless link and the need of a large data transmission band, coupled to low-power dissipation and small size in case of brain-implantable devices. The main issue is related to the huge amount of data associated with a large number of recording channels: in a 64 channel system featuring a 20 kHz sampling frequency (the neural signal band extends to about 7–10 kHz) with 8 bit resolution, the required data rate of the wireless link is 10.24 Mbit/s. Achieving such a high bit rate is not straightforward either in fully implantable devices where power consumption is limited to about 10 mW to avoid tissue necrosis [2] and in battery-powered systems for maintaining a reasonable battery life. Hence, most integrated circuits (ICs) reported in the literature that simultaneously support neural recording and wireless telemetry must trade-off power consumption against the amount of data to be wirelessly sent.

Data compression was adopted to solve this problem in earlier systems. For example, since in a typical extracellular electrode trace most of the useful information is contained in spikes (or action potentials, APs), Harrison *et al.* [3] enable only the transmission of the spike occurrence information, achieving a low bit rate of 330 kbit/s for 100 channels. However, since the examination of the raw traces may improve the selection of the spike detection threshold, the system reported in [4] allows to collect raw data from 2 out of 64 channels present in the system, while only spike occurrence times are sent for the remaining 62 channels via a 2 Mbit/s wireless link. On the other hand, spike waveforms are used in neuroprosthetic applications to assign spikes to specific neurons improving the prediction of an intended movement [5]. Hence, during the last four years, wireless systems capable of sending all raw data have been reported. In [6], Chae *et al.* describe a 128 channel system able to transmit all raw data using an impulse-radio ultra-wide band (IR-UWB) transmitter with up to 90 Mbit/s rate. In such system, problems may arise from skull and skin absorption at high frequencies (signal spectrum is in the 3.5–4.5 GHz range) and from pulse distortion and the need of an external broadband antenna. As a matter of fact, no *in vivo*

tests are reported for this kind of system. In [7], Ghovanloo *et al.* report a system that transmits raw data from 32 channels, encoded in a pulse-width modulated (PWM) signal via a 640 kS/s frequency-shift keying (FSK) wireless link. In this system, the transmitted information relies on the time width of each symbol, making this transmission prone to the noise of the wireless link and of the receiver. Moreover, the system in [7] requires a relatively large bandwidth (38 MHz), which means higher probability of RF interferences, and does not meet the band requirements of any medical communication standard.

We propose here a solution that is intermediate to the two extremes described above. The idea was to employ a data compression algorithm in order to:

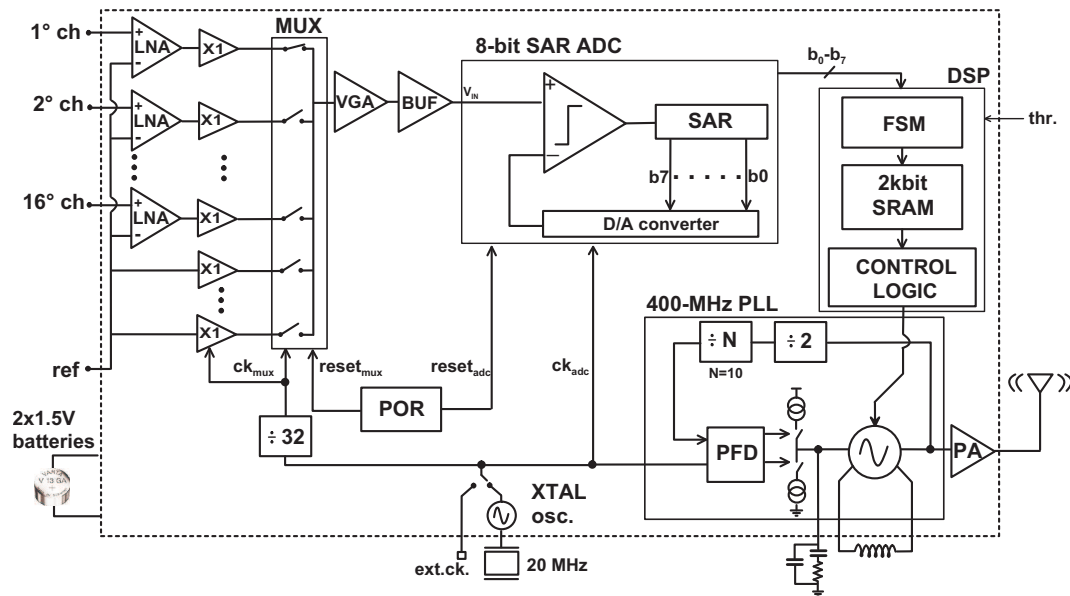
- preserve the information needed for single neuron identification;
- reduce the throughput and thus the power consumption, since the power needed to the antenna is directly proportional to the bit rate [8];
- keep the bandwidth limited to few MHz, thus reducing the probability of RF interference and verifying the possibility to make in the future the system compliant to Medical Implanted and Communication Service (MICS) or Industrial Scientific and Medical (ISM) bands, in the 402–405 MHz and 902–928 MHz frequency range, respectively.

In particular, having reduced the data rate, it was possible to extend the transmission distance well beyond the 1 m value reported in the above-mentioned works to make possible realistic *in vivo* experiments with a small penalty in terms of power consumption. The proposed integrated circuit, already presented in [9], has been included in a complete system, made of a transmission and a receiving unit, that is in use in a neuroscience laboratory on freely-behaving animals in place of commercial wired systems to avoid tethering effects and to provide electrical isolation. In this paper we describe in detail the architecture and circuit design choices, the electrical performance of the IC and the *in vivo* tests on small-laboratory animals.

The paper is organized as follows. Section 2 describes the overall system made of a signal conditioning and transmitting unit and a receiver built with off-the-shelf components. Section 3 discusses in detail the implementation of the system-on-chip (SoC) building blocks and validates the data compression algorithm by means of ad-hoc simulations, while Section 4 shows the electrical characterization of the integrated circuit and reports the results of *in vivo* tests on freely-behaving laboratory animals. Finally, a comparison of the implemented SoC with the state-of-the-art is presented in Section 5.

2. System Architecture

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

Figure 1. Block diagram of the wireless neural recording IC.

2.1. Wireless Recording Unit

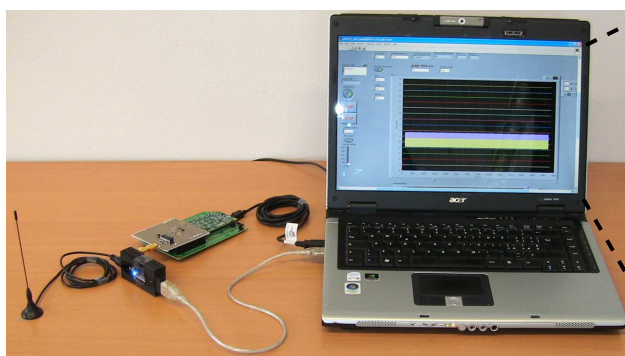
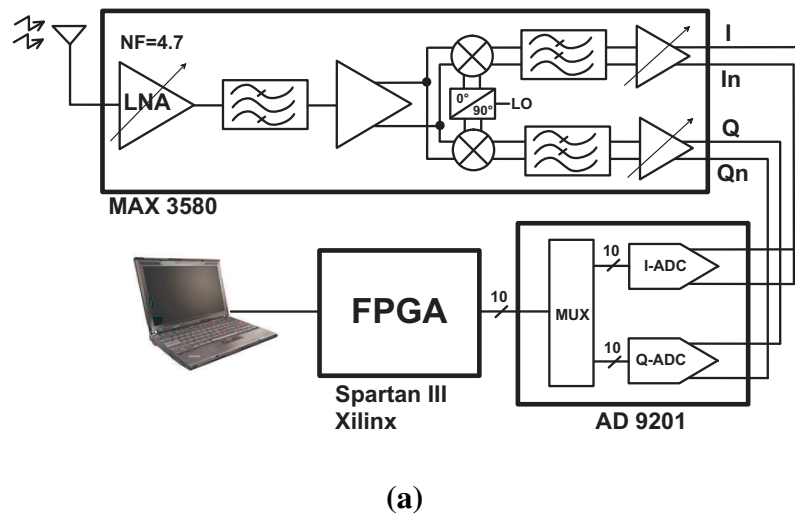
Figure 1 shows the block diagram of the IC which acts as recording and transmitting module. The circuit consists of 16 low-noise amplifiers (LNAs), which perform low-power and low-noise amplification of neural signals. The amplifier outputs are sampled by a time-division multiplexer (TDM) at 20 kS/s onto one data lead before being further amplified by a variable-gain amplifier (VGA) and converted by an 8 bit successive approximation register (SAR) analog to digital converter (ADC). The digital data are then processed by a logic block, performing data compression, assembly and Manchester encoding of the bit stream. Finally, the digital signal is sent to a 400 MHz binary FSK wireless transmitter based on direct modulation of a Voltage Controlled Oscillator (VCO), enabling a 1.25 Mbit/s data rate. A crystal Clapp oscillator employing an external 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. It is worth pointing out that, to reduce silicon area and costs, only 16 amplifiers have been integrated. The remaining 48 channels have been added by connecting a voltage reference to the multiplexer input. In this way all the stages following the amplifier array are designed to process signals from a 64 channel front-end, thus making it possible to reliably assess the actual performance of a complete 64 channel system.

2.2. Receiver and Graphical User Interface

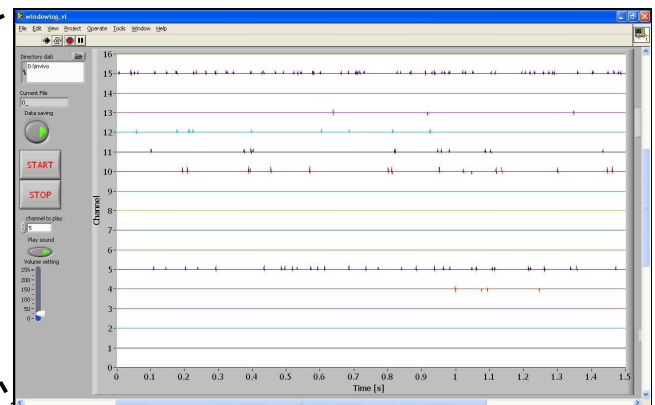
The receiver (see Figure 2) consists of a quadrature zero-IF down-converter (MAX3580) using an LNA with a noise figure of 4.7 dB at 400 MHz, RF and baseband tunable filters and automatic gain control circuits. The quadrature signals are converted by a dual channel, 20 MSps, 10 bit ADC (AD9201) and fed to a Xilinx FPGA module (OpalKelly, XEM3005), which performs frequency demodulation and Manchester decoding and sends the received data to a PC via a USB link. In addition, a

Graphical User Interface (GUI) based on Labview/C software was implemented to allow data saving and on-line elaboration.

Figure 2. Block diagram of the receiver (a) and photo of the remote receiver with the detail of the graphical user interface (b).



(b)



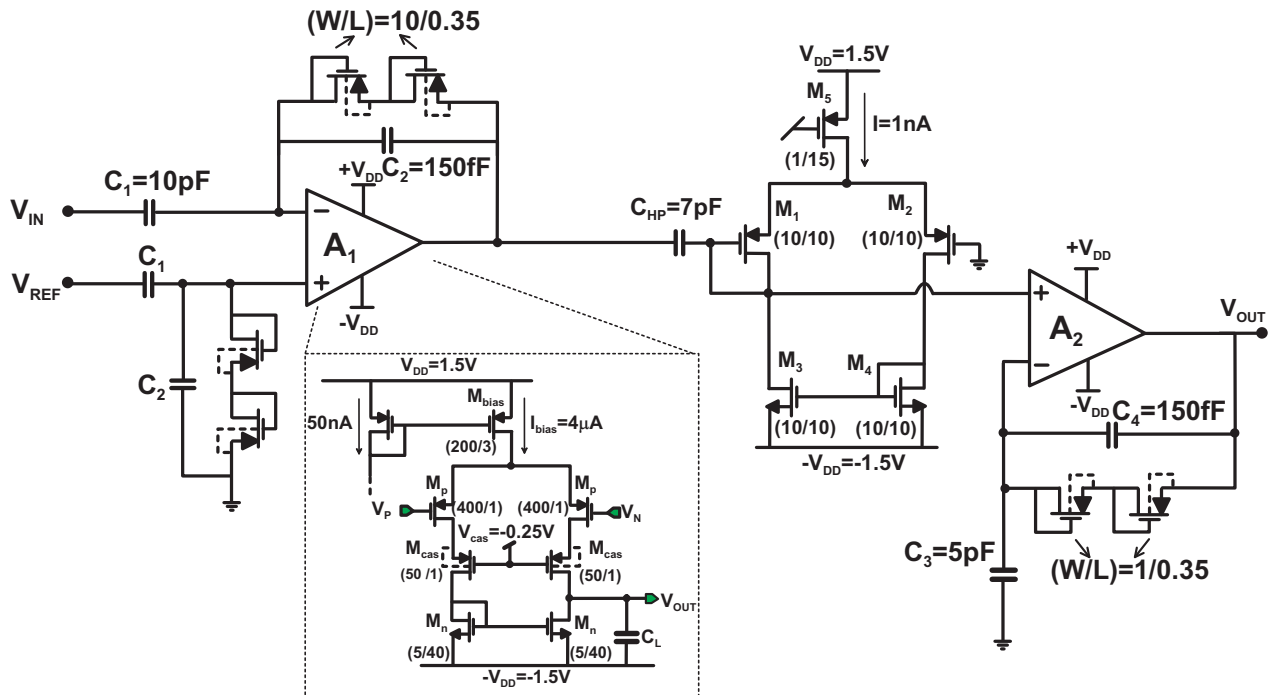
3. Circuit Design

3.1. Analog Front-End

Amplification and filtering of the input signal are performed by a three-stage circuit shown in Figure 3. The first stage is an AC-coupled high-pass filter, using two MOS-bipolar pseudo-resistors as feedback elements [10], which enables the synthesis of high-value resistances without using large-area components. The mid-band amplifier gain is given by $G_1 \cong -C_1/C_2$, set to -67 by taking $C_1 = 10$ pF and $C_2 = 150$ fF. The high-pass pole frequency is placed below 10 Hz to reject the offset and the slow voltage drift of the electrode. A tunable G_M -C high-pass filter with a cut-off frequency f_{HP} of about 300 Hz is introduced after the first stage. It is designed to reject the low-frequency signals, such as Local Field Potentials (LFPs) in the 1–100 Hz frequency range that can prevent a correct detection of the neural spikes or even saturate the amplifier, and cuts off the input-referred $1/f^2$ noise due to the pseudo-resistors of the first amplification stage. After the selective high-pass filter, a second non-inverting gain stage is

added to provide further signal amplification and to define the high frequency cut-off, f_{LP} . The stage is a single-ended capacitive-coupled voltage amplifier with a gain $G_2 \approx 33$, achieved by using $C_3 = 5$ pF and $C_4 = 150$ fF. A capacitive-coupled structure was preferred to a purely resistive feedback amplifier to minimize the current drawn by the OTA output stage. The DC voltage at the amplifier input is determined by the pseudo-resistor elements in the feedback path, while the low-pass cut-off frequency is set by the gain-bandwidth product of the operational amplifier ($GBWP_2$) to about $GBWP_2/G_2 \cong 10$ kHz. Note that at very low frequencies this non-inverting stage has a unity gain in order not to amplify the offset of the operational amplifier.

Figure 3. Schematic of the low-noise band-pass neural pre-amplifier.



3.1.1. Noise Analysis and First Stage Sizing

In order to minimize the amplifier array power consumption while maintaining the input-referred noise specifications, a rigorous noise analysis of the preamplifier and filter has to be carried out following the approach adopted in [10]. The noise performance of the whole LNA depends on the design of the first-stage OTA, whose input noise power density is mainly determined by a compromise between its thermal and flicker noise. To explain this point, let us denote as C_p the input parasitic capacitance to ground of each input terminals of the first-stage OTA [11]. If we neglect the noise contribution due to the two pseudo-resistors (this term will be addressed in a next step), we can write the input-referred noise of the overall amplifier as:

$$\overline{E^2}_{n_{eq}} = \left(\frac{C_p + C_1 + C_2}{C_1} \right)^2 \overline{E^2}_{n_{eq,OTA}} = \left(1 + \frac{1}{|G_1|} + \frac{C_p}{C_1} \right)^2 \overline{E^2}_{n_{eq,OTA}} \quad (1)$$

where $\overline{E^2}_{n_{eq,OTA}}$ is the input-referred voltage noise of the first-stage OTA. Equation (1) highlights that C_p , which is mainly determined by the input transistor dimensions, cannot be increased too much

without compromising the overall input-referred noise. On the other hand, it is well known that large-area input transistors are required to minimize the OTA flicker noise. Therefore, to achieve the best overall performance these two contradictory requirements have to be balanced. Furthermore, according to Equation (1), the first stage gain, G_1 , can be increased maximizing C_1 . Its value is limited by the electrode impedance, whose capacitive component is in the 160–320 pF range. The choice of $C_1 = 10$ pF, $C_2 = 150$ fF, thus $G_1 \approx -67$, results in an upper limit of 0.5 pF for the parasitic capacitance C_p , which guarantees sufficient margin to keep under control the $1/f$ noise at a cost of a marginal 10% increase in the overall amplifier noise with respect to the ideal OTA [see Equation (1)]. In such a design, $\overline{E^2}_{n_{eq}}$ and $\overline{E^2}_{n_{eq,OTA}}$ are almost the same; therefore, both terms will be denoted with the same symbol, $\overline{E^2}_{n_{eq}}$.

Table 1. Transistor operating points.

Transistor	$V_{GS} - V_T$ [mV]	IC	g_m [μ A/V]	r_o [M Ω]
M_p	110	0.075	52.3	5.1
M_n	490	51	7.76	108
M_{cas}	24	0.36	43.7	8.2

To minimize the noise, the first OTA stage has been designed as a simple telescopic cascode amplifier (see Figure 3). This well-known configuration guarantees excellent noise performance thanks to its few transistors, while cascoding and proper transistor sizing ensure enough gain. Straightforward analysis relates the input-referred noise power spectral density to the noise sources of the transistors, leading to:

$$\overline{E^2}_{n_{eq}} \cong \frac{8kT\gamma}{g_{mp}} + \frac{8kT\gamma g_{mn}}{(g_{mp})^2} + \frac{2K_p^{(1/f)}}{C'_{ox} W_p L_p} \frac{1}{f} + \frac{2K_n^{(1/f)}}{C'_{ox} W_n L_n} \left(\frac{g_{mn}}{g_{mp}} \right)^2 \frac{1}{f} \quad (2)$$

The result has been derived by taking into account that the transistors M_{bias} and M_{cas} do not significantly contribute to the overall noise. To minimize the noise, the condition $g_{mp} \gg g_{mn}$ must be fulfilled, *i.e.*, $(W/L)_p \gg (W/L)_n$. Under this assumption the total noise power density results:

$$\overline{E^2}_{n_{eq}} = \frac{8kT\gamma}{g_{mp}} + \frac{2K_p^{(1/f)}}{C'_{ox} W_p L_p} \frac{1}{f} \quad (3)$$

with γ equal to $2/3$ for transistors working in strong inversion or $1/(2\kappa)$ in weak-inversion ($\kappa \approx 0.7$) [12]. In order to minimize the thermal noise without increasing the current consumption, the M_p transistors have to work in the weak inversion region, where the transconductance is maximum. Its value can be estimated from the EKV model [13], valid in all regions of inversion:

$$g_m \cong \frac{\kappa I_D}{U_T} \frac{2}{1 + \sqrt{1 + 4IC}} \quad (4)$$

U_T and IC being the thermal voltage and the inversion coefficient [12], respectively. The latter parameter is defined as:

$$IC = \frac{\kappa I_D}{2\mu C'_{ox} \left(\frac{W}{L} \right) U_T^2} \quad (5)$$

and is much less than 1 in weak-inversion or subthreshold region. From Equation (4), the overall-amplifier input-referred thermal noise power spectral density results:

$$\overline{E^2}_{n_{th}} = \frac{8kTU_T}{\kappa^2 I_{bias}} \quad (6)$$

with I_{bias} being the bias current of the telescopic cascode amplifier. Assuming a first-order roll-off of the frequency response, the noise bandwidth is $\frac{\pi}{2} \cdot 10 \text{ kHz} \cong 15.7 \text{ kHz}$. For an upper limit of $3 \mu\text{V}_{rms}$ for the thermal noise contribution in this noise band, a minimal I_{bias} value of about $3 \mu\text{A}$ is required; to be conservative, we set $I_{bias} = 4 \mu\text{A}$.

With regard to the flicker noise, its contribution can be reduced by increasing the PMOS transistor area ($W_p \times L_p$). Considering that for the adopted technology $K_p^{(1/f)} \approx 2 \times 10^{-26} \text{ V}^2\text{F}$ and $C'_{ox} \approx 5 \text{ fF}/\mu\text{m}^2$ and setting a noise corner frequency lower than 100 Hz , the input PMOS transistors were sized with $W_p \times L_p = 400 \times 1 \mu\text{m}^2$. This sizing results into a stray OTA input capacitance C_p of approximately 500 fF , which does not excessively impair the equivalent input noise. Moreover, in order to fulfill the requirement $(W/L)_p \gg (W/L)_n$, the settings $W_n = 5 \mu\text{m}$ and $L_n = 40 \mu\text{m}$ were applied, thus forcing these devices to work in the strong inversion region (see Table 1). For this choice of the transistor lengths, cascoding of the input differential pair is needed to not to degrade the output resistance and, thus, to not to lower the amplifier gain. The introduction of M_{cas} guarantees an output resistance $R_{out} = r_{0n} || (g_{mcas} r_{0cas} r_{0p}) \approx r_{0n}$ and an amplifier gain $g_{mp} r_{0n} \approx 75 \text{ dB}$.

In summary, by careful transistor sizing, the noise within the amplifier band may be reduced to the sole thermal noise. Since in a generic amplifier a trade-off exists between noise and current consumption, its efficiency may be evaluated referring to the noise efficiency factor (NEF), which is a parameter commonly adopted in literature, proposed in [14] and defined as:

$$NEF = V_{in,rms} \sqrt{\frac{2I_{tot}}{\pi U_T 4kT BW}} \quad (7)$$

where $V_{in,rms}$ is the input-referred rms noise, I_{tot} is the total supply current and BW is the bandwidth of the amplifier. In this limit, the minimum NEF achievable with a simple differential stage can be estimated from Equations (6) and (7) leading to a value of $\sqrt{2}/\kappa = 2.02$ [15]. However, in our design the contribution to the thermal noise of the current mirror transistors cannot be completely neglected. In fact, even if M_p transistors work in weak-inversion and M_n FETs are biased in strong inversion region, the transconductance ratio is only about 7 (see Table 1). Considering the general expression of the transconductance given by Equation (4) and the total input-referred thermal noise given by the first two terms of Equation (2), the theoretical NEF for this preamplifier becomes:

$$NEF_{th} = \frac{\sqrt{2}}{\kappa} \cdot \sqrt{1 + \frac{2\gamma\kappa}{\sqrt{IC_n}}} \cong 2.15 \quad (8)$$

where $\gamma \approx 2/3$, $\kappa \approx 0.7$ and IC_n is the inversion coefficient of current mirror transistors. In reality, the NEF of the proposed amplifier is slightly larger compared to this theoretical limit for three main reasons:

- The transconductance of the input transistors is slightly lower than $\kappa I_D/U_T$ since their inversion coefficient is larger than zero. For the present design, the input transistor IC is 0.075 (see Table 1), and consequently the rms input noise and the NEF increase by a factor of 1.034 .

- The input-referred noise of the overall amplifier is larger by a factor $(1 + 1/G_1 + C_p/C_1)^2$ than the one of the first operational amplifier, as stated by Equation (1). Taking into account that the parasitic input capacitance is approximately 0.5 pF, mainly due to the OTA input transistors, this factor is equal to 1.065. A further contribution derives from the strays associated with the input capacitor plates that was drastically reduced by connecting the capacitor bottom plate (which has the largest parasitism) to the amplifier input and by connecting the top plate to the OTA terminal. In this way, a parasitic capacitance larger than 1.5 pF was avoided.
- The current drawn by the second amplifying stage contributes to the total current in Equation (7) but does not lower the input-referred thermal noise. Therefore, the NEF increases by a factor of $\sqrt{(1 + I_2/I_1)} \cong 1.037$, where I_1 and I_2 are the currents drawn by the first and the second operational amplifier respectively.

These three factors increase the estimated NEF of the overall amplifier to about 2.45.

Finally, note that the adoption of a telescopic cascode amplifier, known to have a small output dynamic, is not a limiting factor for the input signal amplitude. In fact, the positive voltage swing is ≈ 450 mV ($|V_{T,p}| + V_{cas}$) while the negative swing is about 1 V (1.5 V $- V_{ov,n}$), thus assuring a maximum amplitude for the input signal of about 7 mV.

3.1.2. Second Amplifying Stage

A second gain stage is needed to increase the amplitude level of the input signal before multiplexing. Since this stage requires a relative large output swing, a two-stages OTA can be employed. Provided that the first stage has enough gain, the impact of this second amplifier on the input-referred noise is negligible, and its power dissipation can be reduced without affecting the noise performance. In fact, the noise of the second op-amp is dominated by the input transistors, designed to operate in weak-inversion region. If we limit the contribution to the equivalent input noise added by the second stage OTA to less than $0.3 \mu V_{rms}$ (i.e., 1/10 of the dominant contribution), a minimum current of 100 nA is needed in the input differential pair of the second op-amp. To be conservative, we set the bias current to 200 nA in the first stage, while a current of 100 nA is drawn by the second stage.

Finally, nonlinear distortion may be of some concern in the second stage. Distortion arises from the non-linear high-resistance pseudo-resistors placed in the feedback path, which is driven by a large output voltage swing. For this reason the subthreshold MOS transistors were sized to have resistance values an order of magnitude higher than those in the first stage. This choice makes the signal current flowing through them always several orders of magnitudes lower than the signal through the capacitors, limiting the measured Total Harmonic Distortion (THD) generated by the stage below 5% even if the pseudo-resistors were modulated by a 1.5 V peak-to-peak 1 kHz sinusoid.

3.1.3. High-Pass Filter Design and Optimization

Let us now consider the noise due to the MOS-bipolar pseudo-resistors that was neglected in Equation (2). It was experimentally verified that the noise spectral density of the pseudo-resistors complies with the usual equation:

$$\overline{S^2_R} = \frac{4kT}{R} \quad (9)$$

where R is the small signal incremental resistance. This noise contribution to the OTA input-referred noise power spectral density follows a $1/f^2$ law and is given by:

$$\overline{E^2_{n_R}} = 2 \frac{4kT}{R} \frac{1}{(C_1 2\pi f)^2} = \frac{8kT f_1}{2\pi G_1^2 C_2 f^2} \quad (10)$$

where f_1 is the cut-off frequency of the high-pass filter set by the first amplifier stage. The factor of 2 in Equation (10) accounts for the two pseudo-resistors, one in the feedback path and one connected to the positive input terminal of the operational amplifier. For simplicity, let us now neglect the presence of the selective high-pass filter following the first stage. By integrating Equation (10) from f_1 to the low-pass cut-off frequency of the overall amplifier, f_{LP} , we get:

$$\int_{f_1}^{f_{LP}} \overline{E^2_{n_R}} df \cong \frac{8kT}{2\pi G_1^2 C_2} \quad (11)$$

which is independent of the cut-off frequency f_1 . From Equation (11) the contribution of the pseudo-resistors to the OTA input-referred noise is about $2.79 \mu V_{rms}$. This value is not negligible and explains why it is not convenient to implement the selective high-pass filter (*i.e.*, the 300 Hz cut-off frequency high-pass filter) in the first stage, as this solution would degrade the noise performance. The G_M - C high-pass filter is instead able to cut off most of the noise from the pseudo-resistors of the first stage, but a careful choice of the filter capacitor is needed to reduce its noise contribution. We will discuss this point in detail. Let us now suppose that $f_1 \ll f_{HP}$, where f_{HP} is the cut-off frequency of the G_M - C high-pass filter. For frequency above f_1 , the input-referred power spectral density due to both the pseudo-resistors of the first stage and to the G_M - C filter can be written as:

$$\overline{E^2_{eq(1/f^2)}} \cong \frac{8kT}{R} \frac{1}{(C_1 2\pi f)^2} + \overline{S^2_{n_I}} \frac{1}{(G_1 C_{HP} 2\pi f)^2} \quad (12)$$

where $\overline{S^2_{n_I}}$ is the output current noise of the high-pass filter. For simplicity, let us assume that the current noise generated by the G_M cell ($G_M = 1/R_{HP}$) can be written as $4kT/R_{HP}$. Equation (12) becomes:

$$\overline{E^2_{eq(1/f^2)}} = \frac{8kT f_1}{2\pi G_1^2 C_2 f^2} + \frac{4kT f_{HP}}{2\pi G_1^2 C_{HP} f^2} \quad (13)$$

By integrating Equation (13) over the amplifier band, *i.e.*, from f_{HP} to f_{LP} , we obtain:

$$\int_{f_{HP}}^{f_{LP}} \overline{E^2_{eq(1/f^2)}} df = \frac{8kT}{2\pi G_1^2 C_2} \frac{f_1}{f_{HP}} + \frac{4kT}{2\pi G_1^2 C_{HP}} \quad (14)$$

The result suggests to reduce the first term by minimizing the f_1/f_{HP} ratio while a large capacitor value C_{HP} is needed to reduce the second term. In practice, the G_M cell is a simple differential stage

(see Figure 3) whose bias current can be externally tuned. Because a small bias current ($I \cong 1$ nA) is needed to synthesize a high value resistance, all the transistors work in weak-inversion region and their transconductance is $G_M = g_{m1} \cong \kappa I / U_T$. The current noise of this configuration is therefore:

$$\overline{S_{nI}^2} = 4 \frac{2kT}{\kappa} g_{m1} = \frac{8kT}{\kappa R_{HP}} \quad (15)$$

The current noise of the G_M cell is therefore a factor of $2/\kappa$ larger than that of a resistor R_{HP} and this factor has to be added in the second term of Equation (14).

In summary, in order to keep the input-referred noise due to the first stage pseudo-resistors and to the G_M cell in the amplifier band lower than $1 \mu\text{V}_{rms}$, f_1 has to be at least a factor of 10 smaller than f_{HP} and C_{HP} must be greater than 2 pF. The two pseudo-resistors in the first stage of the circuit in Figure 3 were designed to have $f_1 < 10$ Hz (with $f_{HP} = 300$ Hz) and $C_{HP} = 7$ pF.

3.1.4. Line Buffer and Multiplexer

A circular shift register, controlled by a 625 kHz clock (ck_{mux} in Figure 1), sequentially enables each amplifier to access the common data lead; each pass gate switches on both the rising and the falling edge of the multiplexer clock, avoiding clock transitions in the middle of the sampling window. As in this topology each amplifier has to drive a long routing line connecting all the pass-gate outputs, a class-AB buffer is needed to boost the line. To save power, each buffer is switched off when the corresponding amplifier is not selected and turned on when the previous amplifier is selected (*i.e.*, one clock edge before). In this way, at any time, only two buffers are powered, draining about $25 \mu\text{A}$ each from the supply. Further amplification is provided before sampling by a variable-gain amplifier (VGA) added to the amplification chain at the multiplexer output. The gain can be digitally varied between 1 to 8 by means of two auxiliary control bits.

3.2. Analog to Digital Converter

The multiplexed and amplified signal is then converted into a digital form by an 8 bit SAR ADC (see Figure 1). This topology is well-suited for low power and small-area applications since it requires a minimal amount of analog circuitry. To save area, analog multiplexing of input signals, requiring only one AD converter, was employed even though it resulted in a small power penalty with respect to the use of one converter per each amplifier and digital multiplexing of data [16].

The number of the converter bits was chosen taking into account the following considerations:

- The maximum amplitude for an extra cellular action potential is ~ 1 mV [17] while the minimum signal is about $10 \mu\text{V}$, the latter being determined by the typical input noise due to neural background activity and electrode impedance [18]. Therefore, the ratio between the Full Scale Range (FSR) and the ADC least significant bit, *i.e.*, the dynamic range of the converter, should be better than $1 \text{ mV} / 10 \mu\text{V} \approx 100$. This results in a converter resolution larger than 6.35 bit.
- The ADC quantization noise has to be kept much lower than the minimum detectable signal, *i.e.*, the rms input noise. This requirements translates into:

$$\frac{FSR}{2^n G \sqrt{12}} < 10 \mu\text{V} \quad (16)$$

where G is the amplifier gain and n is the number of converter bits. The worst-case condition occurs for the minimum gain of the amplification chain, which is set by the ratio between the FSR and the maximum amplitude of the input signal, A_{max} :

$$G_{min} < \frac{FSR}{A_{max}} \cong \frac{3 \text{ V}}{1 \text{ mV}} = 3000 \quad (17)$$

In the present design we set $G_{min} \cong 2000$ and thus:

$$n > \log_2 \left[\frac{FSR}{\sqrt{12} \cdot 10 \mu\text{V} \cdot G_{min}} \right] \cong 5.4 \quad (18)$$

In practice, taking into account that the effective number of bit (ENOB) is always lower than the nominal figure, the use of an 8 bit ADC will provide a safe margin for the design. A higher resolution is useless and harmful since it would increase the ADC power consumption.

The ADC consists of three main parts: the capacitor array, the comparator and the logic block performing the successive-approximation algorithm. The capacitor array, which acts as sampling capacitance during the sampling interval and as charge-sharing DAC during the conversion period, has been implemented as a binary array. The value of the unit capacitance, C_{LSB} , was chosen considering two requirements:

- The noise in the sampling phase has to be significantly smaller than the quantization noise. Thus, the total capacitance of the array, C_{TOT} , must satisfy the condition:

$$\frac{KT}{C_{TOT}} < \frac{(V_{REF})^2}{12} \quad (19)$$

which clearly is not a limit factor, resulting in C_{TOT} much larger than about 1 fF.

- The LSB capacitance has to be sufficiently accurate and it has to fulfill the following requirement [19]:

$$2 \frac{\Delta C_{LSB}}{C_{LSB}} \cdot \sqrt{2^{N-1}} = 2 \frac{K_c}{\sqrt{A_c}} \cdot \sqrt{2^{N-1}} < 0.5 \quad (20)$$

where $K_c = 0.5\% \cdot \mu\text{m}$ and A_c are the Pelgrom mismatch parameter and the least significant bit (LSB) capacitor area, respectively. From Equation (20) A_c has to be larger than $(4.5 \mu\text{m})^2$, which implies $C_{LSB} > 19.5 \text{ fF}$, considering a specific capacitance of $1 \text{ fF}/\mu\text{m}^2$.

Taking into account also the effect of parasitic capacitance that can degrade the achievable accuracy, $C_{LSB} = 80 \text{ fF}$ was employed as unit capacitor, resulting in a total capacitance of $2^N \cdot C_{LSB} = 20.48 \text{ pF}$.

The comparator was implemented as a classical two-stage topology, as depicted in Figure 4. A 20 dB gain preamplifier is followed by a PMOS latch stage and a couple of inverters are connected at the latch outputs to regenerate the signals. The two-stage architecture was preferred to a dynamic latch in order to avoid kickback noise and metastability [20,21].

The successive approximation logic has been implemented as a synchronous full-custom logic. The A/D converter works with a clock frequency of 20 MHz, providing 1.25 MS/s, *i.e.*, about 64 channels \times 20 kS/s per channel. Hence, a conversion period lasts 16 clock cycles (see Figure 5): The sampling phase lasts 7 clock periods to relax the specs of the input signal buffer, while every bit decision takes

one clock period, starting from the most significant bit (*MSB*). The last clock cycle is reserved for the end-of-count (EOC) operations. The current consumption of input buffer and A/D converter are 250 and 410 μA , respectively.

Figure 4. Schematic of the two-stage comparator. A preamplifier is followed by a cascoded semi-latch to avoid kickback noise and metastability.

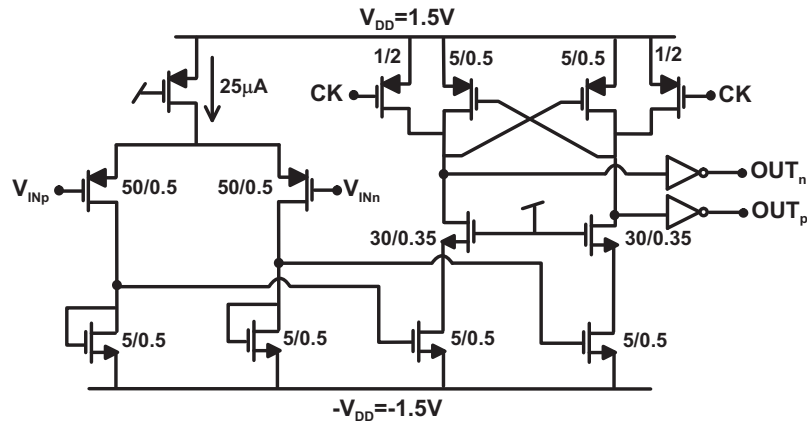
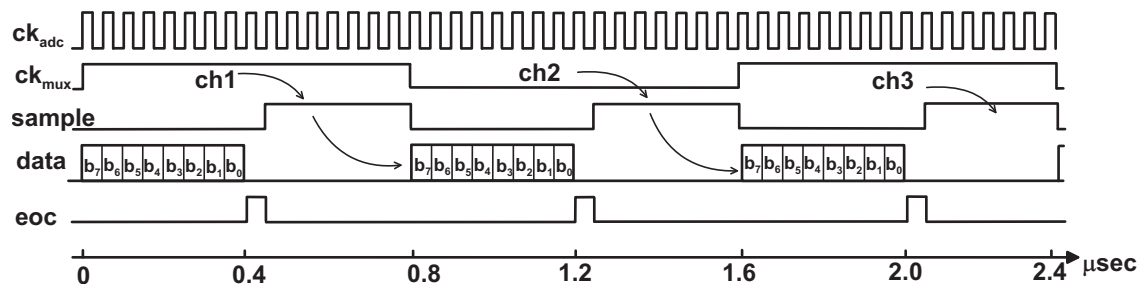


Figure 5. Time diagram of conditioning circuit and ADC.



3.3. Digital Signal Processing

The implemented data reduction system takes advantage of the low duty cycle of the neural activity to eliminate transmission of useless signal containing only noise. Spike duration and firing rates are typically lower than 1 ms and 100 Hz [22], respectively. A finite-state machine (FSM in Figure 6) compares each incoming digital sample with a user-programmable digital threshold, *i.e.*, an 8 bit word stored in the embedded SRAM. When the threshold is crossed, 20 samples of the signal from the same channel are recorded, covering a 1 ms time frame (see Figure 7). This strategy was chosen taking into account that a clear identification of APs requires the detection of three features, namely the amplitude of the peak and the trough and the time interval between them [22], which can be extracted with enough resolution by using 20 samples, even for the fastest spikes. The threshold can be set by the user before the recording starts, choosing among four predefined digital words: “x1000000”, “x0100000”, “x0010000” and “x0001000” (corresponding to 1/2, 1/4, 1/8 and 1/16 of half the ADC full scale range). Since the spike can have a first peak either positive or negative and the baseline is at the middle of the ADC range, the threshold is either positive and negative for each sample. For example, considering the first threshold,

the FSM is triggered if the incoming sample is larger than “11000000” or lower than “01000000”. If the gain of the amplification chain is set to the maximum value ($83\text{ dB} \cong 14000$), these four thresholds correspond to an input-referred amplitude of $\pm 54\text{ }\mu\text{V}$, $\pm 27\text{ }\mu\text{V}$, $\pm 13\text{ }\mu\text{V}$ and $\pm 7\text{ }\mu\text{V}$.

Figure 6. Block diagram of Digital Signal Processing.

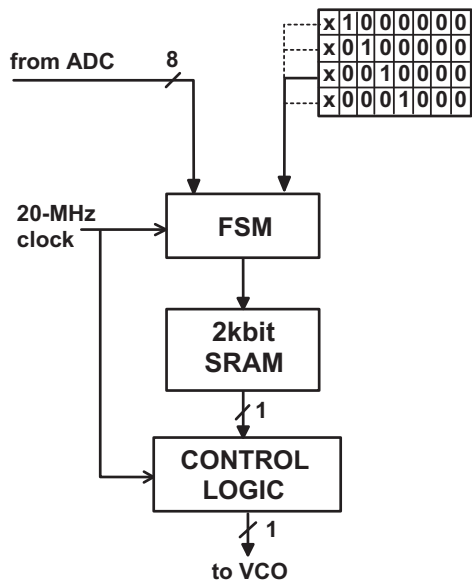
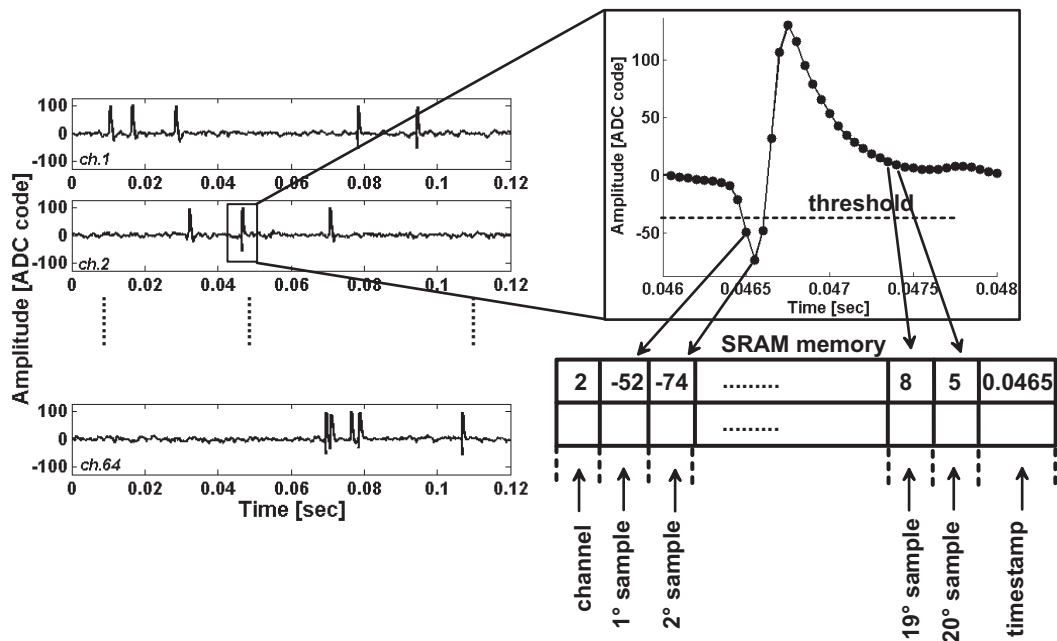


Figure 7. Data compression algorithm: when a spike is detected, 20 digital samples are stored in the embedded memory, together with channel address and timing stamp information.



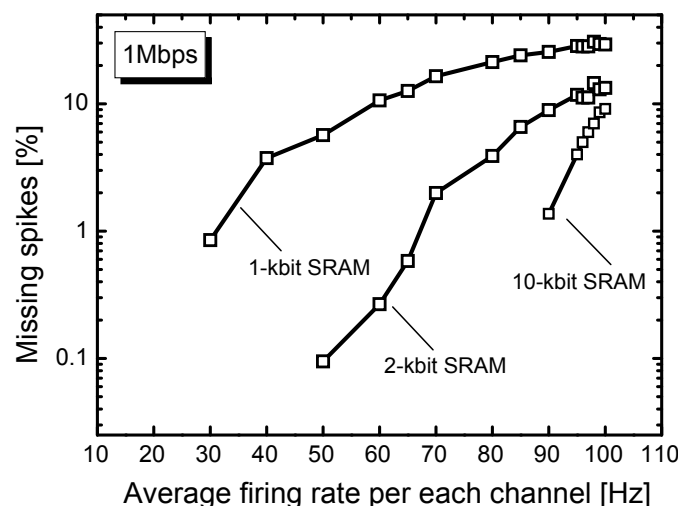
The recorded samples are stored in a 2 kbit embedded memory (SRAM in Figure 6) together with the channel address and the timing stamp, both with a 8 bit resolution. Each spike is then compressed in 22 bytes. The SRAM is continuously read at 1 Mbit/s rate by the control logic block, which adds service

bits for synchronization at the receiver and then performs Manchester encoding of the bit stream. The final data rate reaching the transmitter is 1.25 Mbit/s. The memory read speed was determined by taking into account two 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, corresponding to 64 channels sampled at 20 kHz per channel and 8 bit per sample, 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 and 20 samples per spike with an 8 bit resolution, 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 the neural firing rate follows a Poisson statistics. Figure 8 shows the missing spike percentage as a function of the average firing rate for each of the 64 channels and 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 μ s and thus the worst-case latency time (time difference between the detection and the transmission of an AP) is about 2.9 ms, which is not an issue both in neuroscience experiments and in future prosthetic applications [23].

Figure 8. Missing spike percentage as a function of the average firing rate per channel for different memory sizes and for a read speed of 1 Mbit/s.



The validity of the data-reduction algorithm was tested by running a custom clustering software that employed fuzzy c-means spike sorter [24,25] on a data set from a public source [26] with spike waveforms hardly distinguishable one from each other (see Figure 9). 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 9 shows that clustering on the reduced data set leads to clear separation of the three spike families, with identification errors reported in Table 2. Two types of error occur in a 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 [22], more than 80% of neural spike information is carried by three analog features of the AP waveform, namely the peak and the trough amplitudes and the time between them. The results of spike sorting using these analog features [27] and performed on the same trace are also reported in Table 2. Note that the compression method employed in this work always outperforms the clustering obtained using only these three features. Finally, it is worth nothing that the current consumption of the overall digital signal processing is limited to about 400 μA , most of which consumed by the finite-state machine.

Figure 9. Waveforms of neural spikes and PCA-based clustering results employing the proposed windowing strategy, *i.e.*, processing 20 samples after the threshold crossing.

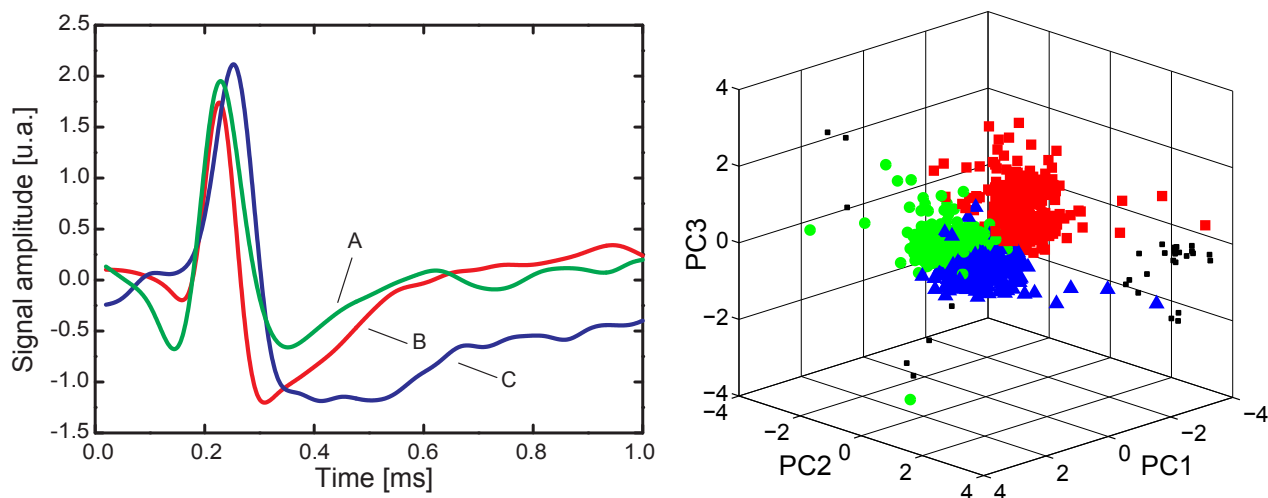


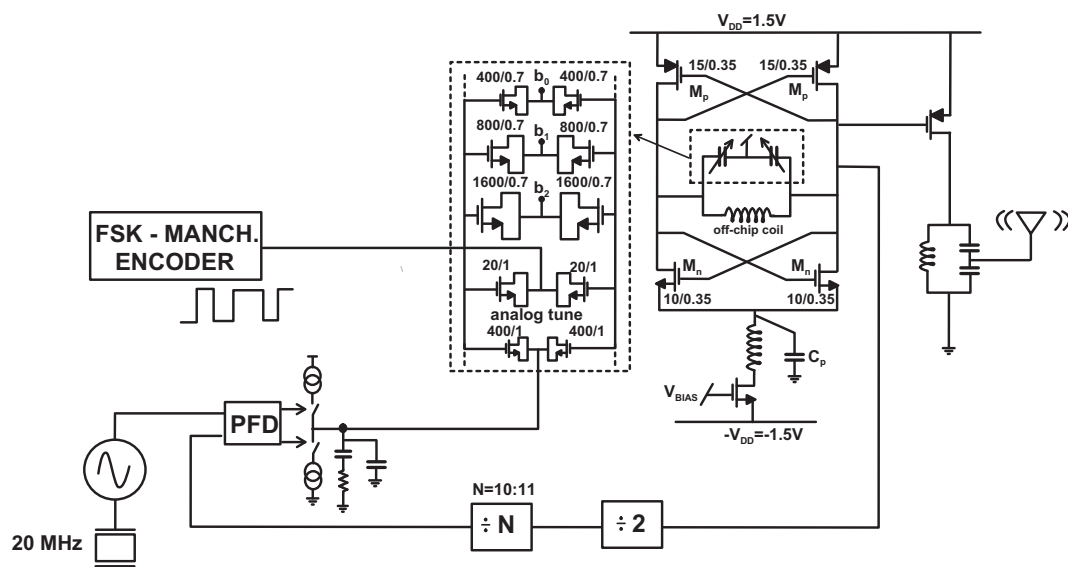
Table 2. Sorting results.

AP class	Original		Reduced		Peak-trough-width	
Error type	I	II	I	II	I	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%

3.4. Manchester-Coded FSK Modulator

The 1.25 Mbit/s bit stream is transmitted using a Manchester-coded binary Frequency Shift Keying (MC-2FSK) modulation at a carrier frequency of 400 MHz. The transmitter (see Figure 10) consists of a voltage-controlled oscillator (VCO) directly modulated by the digital data, inserted in a Phased-Locked Loop (PLL). Since the modulator is a power-hungry system, the design of the whole PLL, and in particular of the high-frequency oscillator, has to be carried out with great care.

Figure 10. Schematic of the Phased-Locked Loop adopted as frequency synthesizer and modulator. The Manchester-coded data stream directly modulates the VCO. The Power Amplifier adopts an external resonant filter to efficiently deliver power to the antenna allowing long transmission range.



The oscillator has a current-biased topology with double cross-coupled pair transconductor. This oscillator architecture allows to minimize the current consumption for a given phase noise with respect to the topology featuring a single differential pair transconductor [28]. In order to accurately set the PLL output frequency at the desired value of 400 MHz and to compensate for process, temperature and power-supply variations, the oscillator has to span over $\pm 10\%$ of its central frequency, *i.e.*, a band of about ± 40 MHz. This means that the capacitance has to be varied as:

$$\frac{|\Delta C|}{C} = 2 \frac{\Delta f_0}{f_0} = 40\% \quad (21)$$

where f_0 is the central oscillation frequency (400 MHz), Δf_0 the covered band of 80 MHz and C the tank capacitor. To make the oscillation frequency insensitive to the parasitic capacitance of the active devices, routing, pads and package pins (the inductor is off-chip), the tank capacitance has to be set in the pF range. Thus, we designed the oscillator with C tunable from 6 pF to 11 pF, determining an inductance value of about 20 nH. The tank capacitance is implemented by three digitally-controlled binary-weighted NMOS inversion-mode varactors that perform coarse tuning of the oscillation frequency and an analog NMOS inversion-mode varactor for fine-tuning of the frequency, enabling to close the loop of the PLL. Another small NMOS varactor is also adopted for frequency modulation.

Once the tank parameters have been determined, the current I_{BIAS} of the tail generator and the (W/L) ratio of the pair transistors have to be chosen taking into account the following considerations:

- The small-signal loop gain or excess gain [29] of the oscillator ($g_m R_{TANK}$, where g_m is the small-signal transconductance of the double differential-pair and R_{TANK} is the tank equivalent parallel resistance) has to be larger than 1 in order to assure the oscillation start-up. Since the common-mode output voltage of the oscillator is set to the half of the power supply to maximize the oscillation swing, *i.e.*, $V_{DD}/2 = 1.5$ V, we have:

$$g_m R_{TANK} = \frac{I_{BIAS} R_{TANK}}{V_{ov}} > 1 \quad (22)$$

corresponding to a current larger than $\cong 220 \mu\text{A}$. Note that in Equation (22) it was assumed that the overall transconductance g_m of the double differential-pair is equal to the single MOS transconductance and that NMOS and PMOS transistors have the same overdrive voltage, $V_{ov} = V_{GS} - V_T = (1.5 - 0.7) \text{ V} \cong 0.8 \text{ V}$.

- The differential oscillation amplitude, A_0 , has to be sufficiently large (>1.5 V) to drive the subsequent stages, *i.e.*, the frequency dividers and the power amplify. Since

$$A_0 \cong \frac{4}{\pi} I_{BIAS} R_{TANK} = \frac{4}{\pi} I_{BIAS} \cdot \omega_0 L Q \quad (23)$$

where Q is the tank quality factor [30], the current I_{BIAS} is minimized adopting an inductor with the maximum quality factor. We opted for an off-chip inductor of 21 nH featuring a Q of about 70 at 400 MHz. This choice results in a current larger than 320 μA to achieve the desired voltage amplitude.

- The VCO phase noise, which mainly determines the phase noise of the whole synthesizer at a frequency offset larger than the PLL loop bandwidth, has to be sufficiently lower in order not to worsen the Signal-to-Noise-Ratio (SNR) of the modulated signal. This can be evaluated as Single-Sideband to Carrier Ratio, *i.e.*, as the ratio of the power in a 1 Hz bandwidth at an offset ω_m from the fundamental angular frequency ω_0 and the power of the carrier, giving [31]:

$$SSCR(\omega_m) = \frac{kT}{C} \frac{\omega_0}{Q} \frac{1}{\omega_m^2} \frac{1+F}{A_0^2} \quad (24)$$

where F is the noise factor of the transconductor ($F \cong 1$). From behavioral simulations, the phase noise at 1 MHz offset from the carrier has to be lower than -120 dBc/Hz to not to degrade the frequency modulated signal.

- A resonant filter (at twice the oscillation frequency) connected to the source of the NMOS differential pair transistors (see Figures 10 and 11) can be added to reduce flicker noise up-conversion into phase noise, which can severely limit the VCO performance [29].

For $I_{BIAS} = 400 \mu\text{A}$, C in the 6–11 pF range and $Q \cong 70$, the excess gain is approximately 2 and the oscillation amplitude is larger than 2 V, preventing the tail current generator to enter the ohmic region

and avoiding a phase noise degradation. Moreover, these choices ensure that the phase noise is not an issue being better than 1–50 dBc/Hz at 1 MHz frequency offset.

The VCO is followed by a cascade of a divide-by-2 dynamic logic prescaler and a static modulo-10 CMOS counter, which further divides the frequency by 10. The PLL reference is the same 20 MHz master clock, allowing the loop to set the oscillation frequency at 20 times the reference clock, *i.e.*, 400 MHz. The PLL also includes a phase-frequency detector with a charge-pump driving an RC - C off-chip filter (see Figure 10), which sets the loop bandwidth to 50 kHz. Excluding the VCO, the power consumption is mainly due to the high-frequency prescaler, which draws more than 200 μ A. The current consumption of the whole synthesizer is about 700 μ A.

The synthesizer frequency is modulated in an open-loop mode by the data at 1.25 MHz, *i.e.*, at 1.25 Mbit/s rate with Manchester encoding. Thus, the loop is open for the frequency modulation since the PLL bandwidth is significantly lower than the bit-rate. This prevents the PLL to filter out the induced frequency modulation. The only purpose of the loop is to fix the central oscillation frequency in order to avoid frequency drift due to temperature and power supply variations. To limit the occupied bandwidth, a frequency deviation of about ± 400 kHz was adopted. Considering a modulation frequency of 1.25 MHz, this value corresponds to a modulation index of 0.64. The bandwidth of the transmitted signal is determined by the Carson rule [32]:

$$BW = 2 \cdot (\Delta f + f_m) \cong 3.3 \text{ MHz} \quad (25)$$

where Δf and f_m are the peak frequency deviation (400 kHz) and the frequency modulation (1.25 MHz), respectively.

3.5. Power Amplifier

The transmitter is completed by a class-AB power amplifier (PA) that drives a 50- Ω antenna via an off-chip resonant filter. The PA consists of an open-drain PMOS transistor directly connected to the oscillator output, sized in order to draw about 3.5 mA. Thus, the power delivered to the antenna is about 0 dBm considering a reasonable efficiency of 10%. The output power, and thus the power consumption of the PA, were set after estimating the sensitivity of the receiver, which depends on the modulation type, the signal bandwidth and the noise factor of the receiver itself. The receiver input noise in the signal bandwidth of 3 MHz is:

$$N_{in} = kTF \cdot BW = -99 \text{ dBm} \quad (26)$$

where F is the noise factor of the receiver, which was assumed to be equal to 10 dB in the above equation. For a low modulation index Manchester-encoded modulation, the SNR required at the receiver input to ensure a bit error rate (BER) lower than 10^{-5} is $SNR_{min} \cong 20$ dB [33]. Thus the input power needed at the receiver is:

$$P_{min} = N_{in} + SNR_{min} = -79 \text{ dBm} \quad (27)$$

Once the receiver sensitivity has been computed, it is possible to set the PA power to achieve a desired transmission range. Taking into account the Friis transmission equation, the ratio of power available at the input of the receiving antenna, P_R , to the output power of the transmitting antenna, P_T , is:

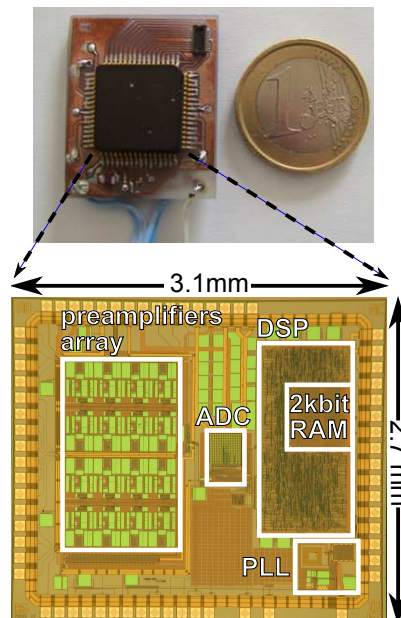
$$\frac{P_R}{P_T} = G_T G_R \left(\frac{\lambda}{4\pi d} \right)^2 \quad (28)$$

where G_T and G_R are the gains with respect to an isotropic radiator of the transmitting and receiving antennas, respectively, λ is the wavelength, and d is the distance between the antennas. While the receiver antenna is a dipole antenna and its gain can be estimated to be 0 dB, the transmitter antenna is a whip antenna with a gain that can be as low as −20 dB. Note that Equation (28) represents the power at the receiver antenna that is supposed to be matched with the receiver. Thus, the receiver input power is 3 dB lower. Thus, considering Equation (28) and setting the maximum transmission distance to 10 m, the power needed at the transmitter antenna is 11 dBm, where a 3 dB factor accounting for the matching between the receiver antenna and the receiver itself is considered. Finally, to have a safety margin, a PA featuring a 0 dBm output power was designed.

4. Experimental Results

4.1. Electrical Characterization

Figure 11. Photo of the board adopted during the neuroscience experiments with rats and microphoto of the die.



The chip was fabricated in AMS 0.35 μm , 3 V 4M2P CMOS process and occupies a total area of $3.1 \times 2.7 \text{ mm}^2$, pads included (see Figure 11). The preamplifier has a mid-band gain of about 65 dB, a high frequency cut-off of 10.5 kHz and a low-frequency cut-off tunable from 1 Hz to 1 kHz (see Figure 12). The input-referred noise, for $f_{HP} = 300 \text{ Hz}$, is 3.05 μV , while the current consumption is 4 μA for the first stage, 0.001 μA and 0.3 μA for the $G_M - C$ high-pass filter and the second amplifying

stage respectively, resulting in an overall NEF of 2.5. The input-referred power spectral density is shown in Figure 13 for different values of the G_M - C filter high-pass corner frequency. The plateau of -152 dBV²/Hz corresponds to an input-referred noise of about $(25$ nV)²/Hz, close to the expected value. Note that, for $f_{HP} = 300$ Hz, the $1/f^2$ input-referred noise is mainly due to the G_M - C high-pass filter, as predicted by Equation (14). The common-mode rejection ratio ($CMRR$) and the power supply rejection ratio ($PSRR$) of the overall pre-amplifier are higher than 65 dB and 50 dB, respectively, while the cross-talk between adjacent channels due to the analog multiplexing is lower than -40 dB.

Figure 12. Measured transfer function for different high-pass corner frequencies set by the G_M - C filter.

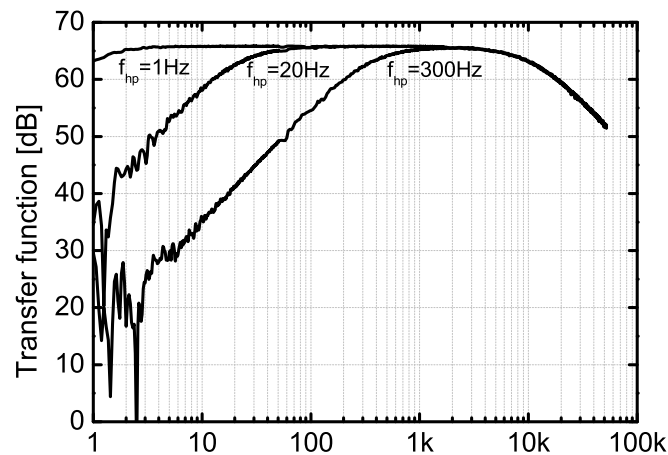
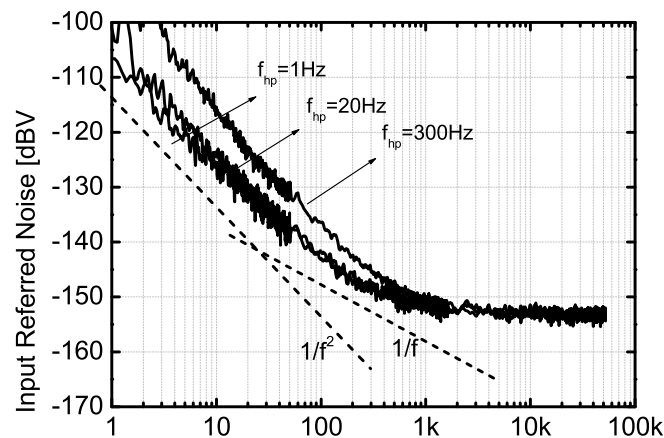
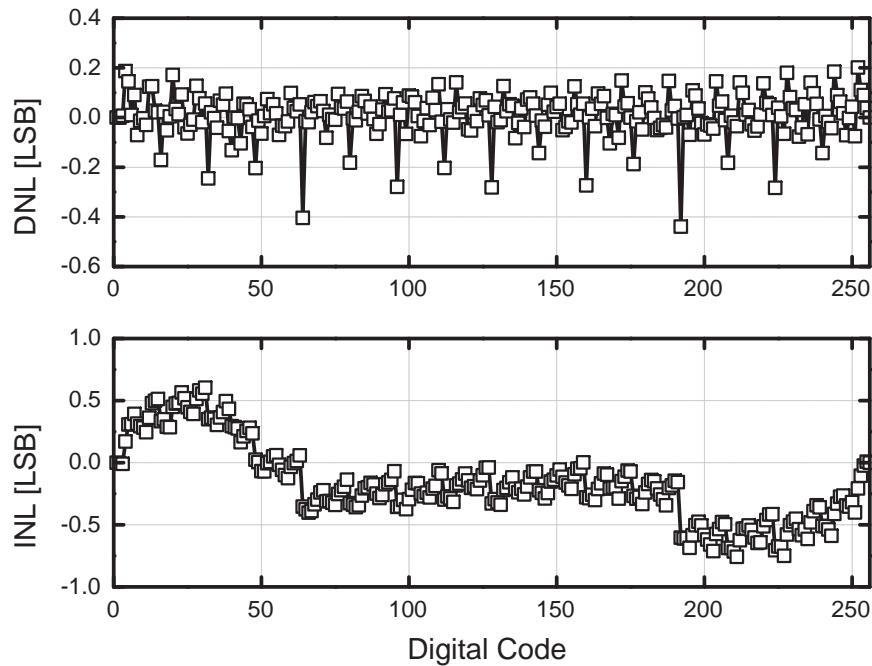
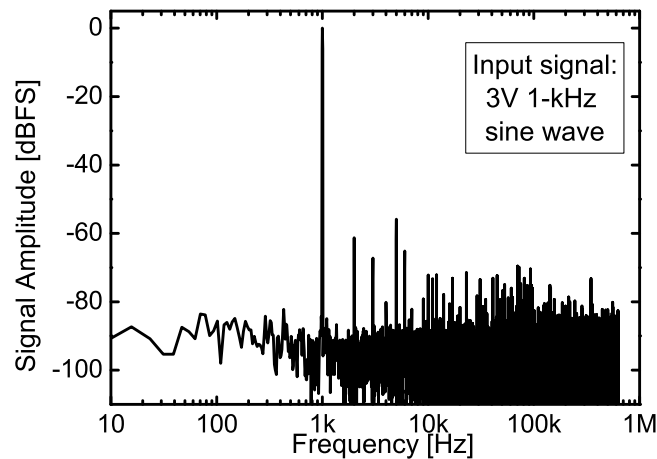


Figure 13. Measured input-referred noise for different high pass-corner frequency.



With regard to the AD converter, Figure 14 shows the results of static measurements. The converter features a DNL less than 0.45 LSB and an INL less than 1 LSB. Figure 15 shows the output spectrum of the ADC for a full-scale sine wave at 1 kHz frequency as input signal. The signal to noise plus distortion ratio (SNDR) is close to 45 dB, and the number of effective bit (ENOB) is thus about 7.2.

Figure 14. Measured DNL and INL of the integrated AD converter.**Figure 15.** Spectrum of the ADC output for a 3 V sine wave at 1 kHz frequency as input signal.

Considering the modulator, the frequency range covered by the VCO spans from 310 to 460 MHz (see Figure 16), allowing to lock the PLL loop to a frequency of 400 MHz frequency even with a large spread of tank component values and for sizable variations in power supply and temperature. When no data modulation is performed, the modulator phase noise (see Figure 17) is less than -123 dBc/Hz at 1 MHz offset from the carrier. The phase noise spectrum has a typical low-pass shape, with 50 kHz cut-off frequency, since it is completely dominated by the charge-pump noise [34], while the VCO phase noise is much lower. The rms integral frequency noise of the modulator in the 10 Hz to 10 kHz band is less than 3 kHz (see the residual frequency FM in Figure 17), whereas the output spectrum of the synthesizer when acting as frequency modulator is depicted in Figure 18: more than 98% of the signal power is included in a 3 MHz band.

Figure 16. Oscillation frequency varying the analog tuning voltage and for different coarse tuning bit configuration.

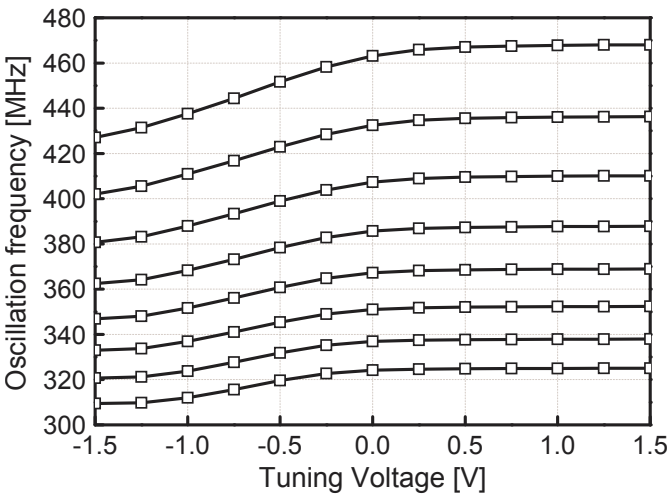


Figure 17. Measured phase noise of the phased-locked loop when the modulation is off.

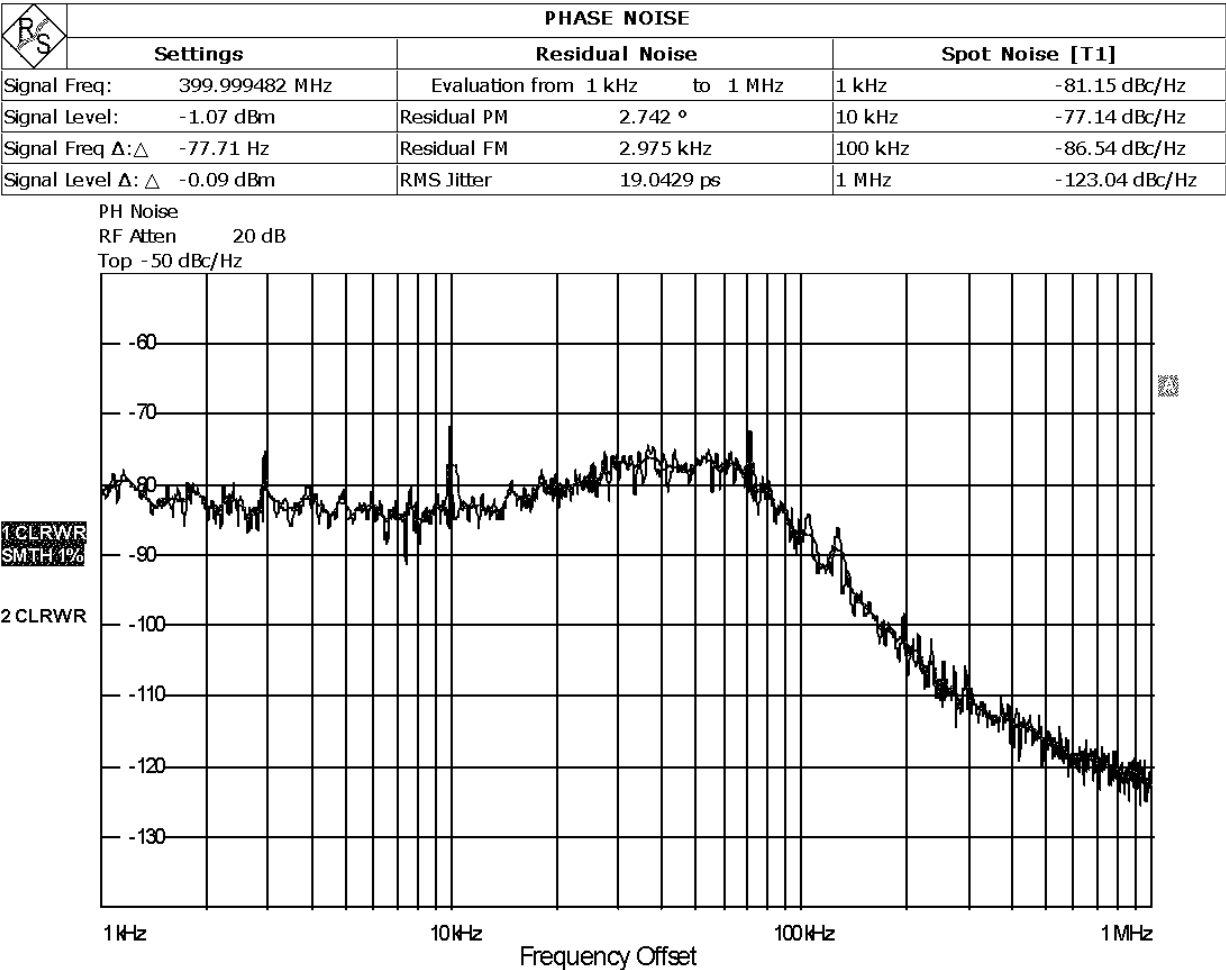
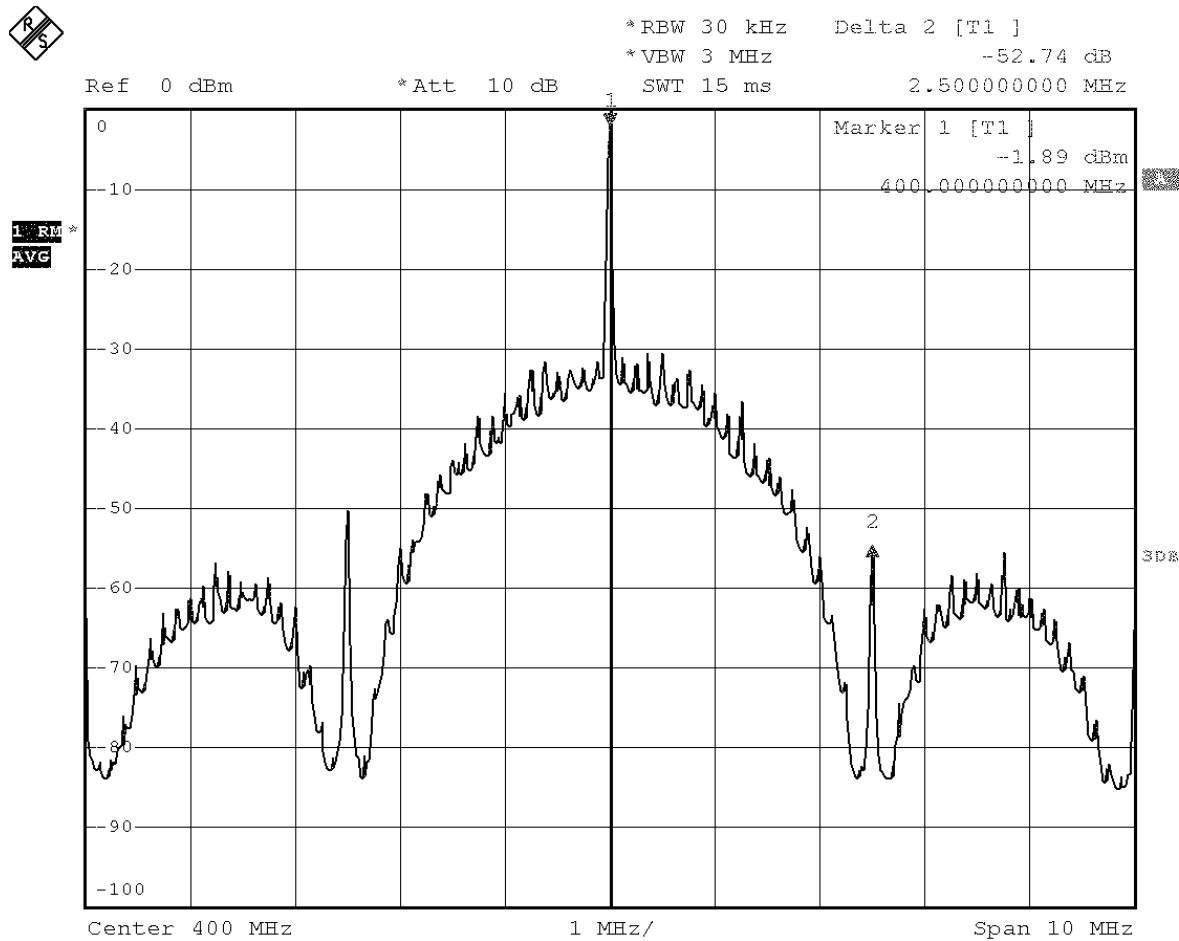


Figure 18. Spectrum of the modulated signal.

For a $\lambda/4$ whip antenna at the transmitter and a dipole antenna at the receiver side, the received power in a free space is shown in Figure 19. The sensitivity of the receiver has been measured to be -74 dBm for a BER of 10^{-5} , close to the estimated theoretical value predicted by Equation (27) of -79 dBm, allowing a transmission range larger than 30 m. The system has also been successfully tested for a 10 m distance between transmitter and receiver in a hostile environment as a neuroscience laboratory, giving a PA efficiency of about 7% while delivering a -2 dBm at a $50\text{-}\Omega$ output load. Disabling the PA, the tank inductor may act as transmitting antenna; in such measurements about -70 dBm was received by a loop antenna placed at approximately 5 cm from the tank inductor.

The overall performance of the implemented system is provided in Table 3. The power consumption of the presented circuit has been computed by projecting the performance to a complete 64 channel system. To this end, the additional budget needed for the missing 48 amplifiers has been added to the actual power consumption (16.6 mW) leading to 17.2 mW (6.7 mW with the PA turned off). The estimate is fair since all the other circuit blocks work with the same data rate of a full 64 channel system.

Figure 19. Received power as a function of the distance between transmitter and receiver antenna in free-space. A dipole antenna was adopted at the receiver side.

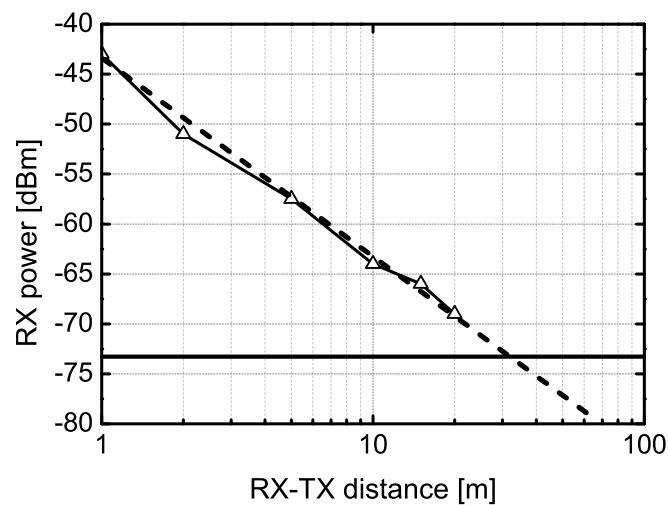


Table 3. Performance summary.

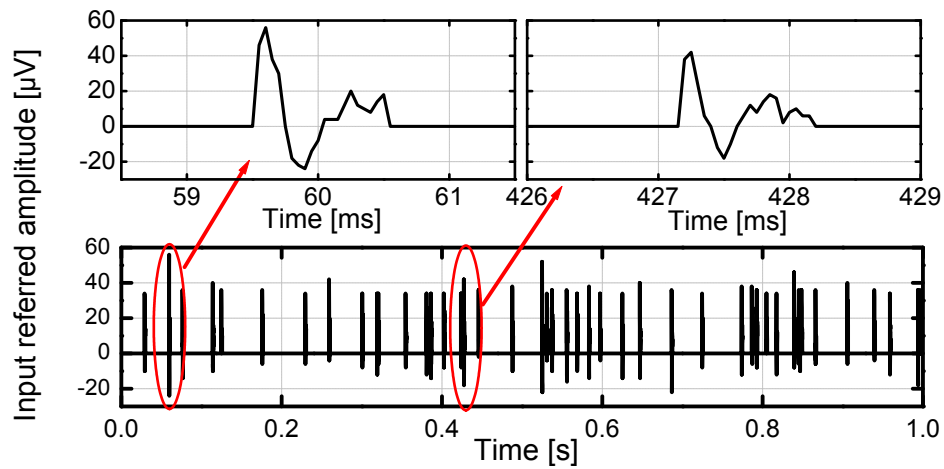
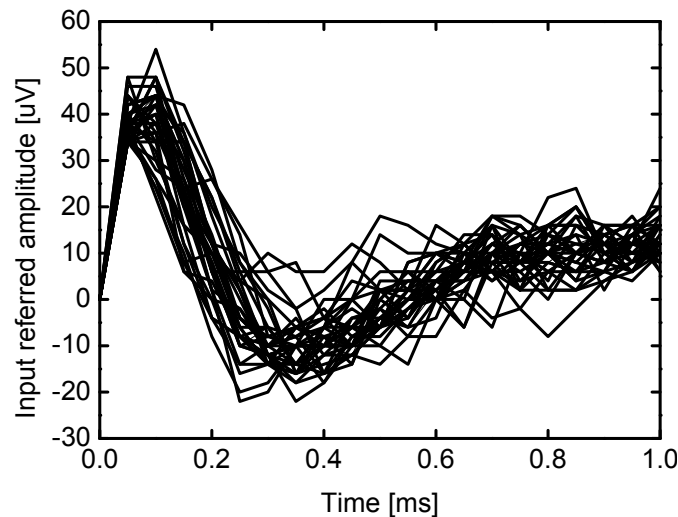
Technology	0.35 μm AMS
Supply Voltage	3 V
Number of channels	64 (16 available)
Pre-amplifier gain	65 dB
Pre-amplifier band	HP tunable-10.5 kHz
Input-Referred Noise	$3.05 \mu\text{V}_{\text{rms}}$ ($f_{\text{HP}} = 300 \text{ Hz}$)
ADC DNL-INL	$< 0.45 \text{ LSB}; < 1 \text{ LSB}$
ADC ENOB	7.2
Transmission frequency	400 MHz
Bit-rate	1.25 Mbit/s
Current Consumption	
20-MHz crystal oscillator	90 μA
Pre-amplifiers	$4.3 \mu\text{A} \times 64 = 275 \mu\text{A}$
Line buffers	$25 \mu\text{A} \times 2 = 50 \mu\text{A}$
VGA	50 μA
ADC buffer	250 μA
ADC	410 μA
DSP+2 kbit RAM	400 μA
Modulator (PLL)	700 μA
Power amplifier	3.5 mA
Total power consumption	6.7 mW (17.2 mW with PA enabled)

4.2. In Vivo Experiments

Figure 20 shows the system mounted on a rat for *in vivo* signal acquisition. It consists of two parts, a wireless neural recording headstage and a backpack. To limit disturbances collected from the environment, a small board ($3 \times 3 \text{ cm}^2$ and 4.5 g of weight) is directly connected to a 16 channel microelectrode array (Tucker-Davis), with an impedance in the 20–60 k Ω range at 1 kHz, implanted into the somatosensory cortex of an adult rat. The board includes 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 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 by recording the raw signals and noise from the same 16 electrode channels. The recorded traces featured a noise of about $10 \mu\text{V}_{rms}$ on all channels. The threshold of the digital peak processor was set to “ $\times 1000000$ ” and the gain of the overall amplifying chain to the maximum value (83 dB) since the peak-to-peak spike amplitude was always lower than $100 \mu\text{V}$. Thus, the digital threshold corresponds to an input-referred amplitude of about $\pm 30 \mu\text{V}$, *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 receiver dipole antenna has been placed at two meters from the rat while neural activity was recorded. Figure 21 shows a single trace recorded during the experiment and two spike waveforms extracted from the data stream, while in Figure 22 a large number of spikes recorded on the same channel are shown aligned in time.

Figure 20. Arrangement for *in vivo* measurements. Note the headstage, the battery backpack and the whip antenna placed along the back of the animal.



Figure 21. Example of a reconstructed trace registered in an *in vivo* experiment.**Figure 22.** Action potentials registered on a same channel and aligned in time.

5. Discussion and Conclusions

Table 4 compares the implemented system to other wireless neural recording ICs with respect to the three most important features, *i.e.*, the transmitted data type (raw data, spike detection or spike waveform), the transmission range and the power consumption per channel. For a fair comparison, since state-of-the-art wireless neural recording systems feature a transmission range always lower than 1 m, the proposed circuit can be thought sized to get a similar transmission range. Since the power needed at the antenna is inversely proportional to the square of the transmission distance, the PA power consumption could be scaled down by a factor of 900, *i.e.*, $(30 \text{ m}/1 \text{ m})^2$. In this case the equivalent power consumption of the whole system reduces to about 6.7 mW, corresponding to less than 105 μW per channel.

Table 4. Comparison of wireless neural recording ICs.

Parameter	[3]	[6]	[7]	[35]	This work
Technology	0.5 μm	0.35 μm	0.5 μm	0.13 μm	0.35 μm
Power source	inductive link	battery	battery	NA	battery
Number of channels	100	128	32	64	64 (16 avail.)
Overall gain	60 dB	57–60 dB	68–78 dB	54–60 dB	65–83 dB
Input noise	5.1 μV	4.9 μV	9.3 μV	6.5 μV	3.05 μV
TX frequency	433 MHz	4 GHz	915 MHz	915 MHz	400 MHz
Modulation	2FSK	IR-UWB	2FSK (PWM)	FSK-OOK	2MC-FSK
Data type	spike detection	raw data	raw data	raw data	AP waveform
Data rate	330 kbit/s	90 Mbit/s	640 kbaud/s	1.5 Mbit/s	1.25 Mbit/s
Bandwidth	0.8 MHz	1 GHz	38 MHz	3 MHz	3 MHz
TX range	13 cm	1 m	1 m	1 m	30 m
Power per channel	135 $\mu\text{W}/\text{ch}$	47 $\mu\text{W}/\text{ch}$	220 $\mu\text{W}/\text{ch}$	80 $\mu\text{W}/\text{ch}$	269 $\mu\text{W}/\text{ch}$

By comparing our system with the one described in [3], working in the same frequency range, the proposed system features a lower power consumption per channel. Also the data quality is greatly improved since spike detection is replaced by waveform detection, making possible single spike identification over a large microelectrode array. Compared with the IC in [7], our system shows an improvement of a factor larger than 2 in terms of power consumption per channel. It is true that the data rate does not allow all raw data acquisition as in [7], but our design optimizes information transfer and is potentially compliant with the narrowband spectral limitations in the Medical Implant Communication Service (MICS) band range (402–405 MHz). In the future, this system could also be designed to work in the Industrial Scientific and Medical (ISM) band (902–928 MHz).

The system in [6] is better in term of power per channel but, relying on an UWB transmitter, it can suffer from interference. Moreover, this system has never been tested in *in vivo* experiments and no information on the adopted antenna or on the power delivered by the PA is given. Regarding the system presented in [35], its power consumption per channel (80 $\mu\text{W}/\text{ch}$) was evaluated considering the minimum transmitter power, assuming that it allows a 1 m transmission distance. However, the data rate is limited to 1.5 Mbit/s, translating into data acquisition at a sampling rate <3 kS/s from all 64 channels. Higher sampling rates are possible only for a limited number of channels: assuming that 20 kS/s is the minimum sampling rate to acquire good quality neural signals, the maximum number of channels that can be recorded and transmitted simultaneously is only 9.

In conclusion, thanks to the implemented compression algorithm that allows to preserve the information needed for single neuron identification while limiting the data throughput and the bandwidth, we achieved a good balance between power consumption and quality of transmitted data and transmission range. In addition, the power consumption per channel was lowered by a careful sizing of each individual circuit block.

Acknowledgment

The authors would like to thank L. Fadiga (IIT, Genova) for useful discussions and A. Oleynik (Universita' di Ferrara) for providing us the clustering software.

References and Notes

1. Olsson, R.H.; Buhl, D.L.; Sirota, A.M.; Buzsaki, G.; Wise, K.D. Brain-machine interfaces: Past, present and future. *Trends Neurosci.* **2006**, *29*, 536–546.
2. Seese, T.; Harasaki, H.; Saidel, G.; Davies, C. Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of muscle tissue to chronic heating. *Labs Invest.* **1998**, *78*, 1553–1562.
3. Harrison, R.R.; Watkins, P.T.; Lovejoy, R.O.; Kier, R.J.; Black, D.J.; Greger, B.; Solzbacher, F. A low-power integrated circuit for a wireless 100-electrode neural recording system. *IEEE J. Solid State Circuits* **2007**, *1*, 123–133.
4. Sodagar, A.M.; Perlin, G.E.; Yao, Y.; Najafi, K.; Wise, K.D. An implantable 64-channel wireless microsystem for single-unit recording. *IEEE J. Solid State Circuits* **2009**, *9*, 2591–2604.
5. Hochberg, L.R.; Serruya, M.D.; Friebs, G.M.; Mukand, J.A.; Saleh, M.; Caplan, A.H.; Branner, A.; Chen, D.; Penn, R.D.; Donoghue, J.P. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **2006**, *442*, 164–171.
6. Chae, M.S.; Liu, W.; Yang, Z.; Chen, T.; Kim, J.; Sivaprakasam, M.; Yuce, M.A. 128-Channel 6 mW Wireless Neural Recording IC with On-the-Fly Spike Sorting and UWB Transmitter. In *Proceedings of the IEEE International Solid-State Circuits Conference (ISSCC 2008), Digest of Technical Papers*, San Francisco, CA, USA, 3–7 February 2008; pp. 146–148.
7. Lee, S.B.; Lee, H.M.; Kiani, M.; Jow, U.M.; Ghovanloo, M. An inductively powered scalable 32-channel wireless neural recording system-on-a-chip for neuroscience applications. *IEEE Trans. Biomed. Circuits Syst.* **2010**, *4*, 360–371.
8. Liu, Y.-H.; Li, C.-H.; Lin, T.-H. A 200-pJ/b MUX-Based RF transmitter for implantable multichannel neural recording. *IEEE Trans. Microw. Theory Tech.* **2009**, *57*, 2533–2541.
9. Bonfanti, A.; Ceravolo, M.; Zambra, G.; Gusmeroli, R.; Borghi, T.; Spinelli, A.S.; Lacaita, A.L. A multi-channel low-power IC for neural spike recording with data compression and narrowband 400-MHz MC-FSK wireless transmission. In *Proceedings of the 2010 European Solid-State Device Research Conference (ESSCIRC)*, Seville, Spain, 14–16 September 2010; pp. 330–333.
10. Harrison, R.R.; Charles, C. An implantable 64-channel wireless microsystem for single-unit recording. *IEEE J. Solid State Circuits* **2003**, *6*, 958–965.
11. C_p is not drawn in Figure 3. It can be easily verified that these strays do not affect, in principle, the differential gain of the first stage. Only their mismatch causes a pole-zero doublet which, however, has no practical impact since it lays at very low frequency, below the band of interest.
12. Vittoz, E.; Fellrath, J. CMOS analog integrated circuits based on weak inversion operations. *IEEE J. Solid State Circuits* **2003**, *3*, 224–231.

13. Enz, C.C.; Krummenacher, F.; Vittoz, E.A. An analytical MOS transistor model valid in all region of operation and dedicated to low-voltage and low-current applications. *Analog Integr. Circuits Signal Process.* **1995**, *8*, 83–114.
14. Steyaert, M.S.J.; Sansen, W.M.C. A micropower low-noise monolithic instrumentation amplifier for medical purposes. *IEEE J. Solid State Circuits* **1987**, *22*, 1163–1168.
15. This limit can be overwhelmed adopting a current-reuse input stage that makes use of both NMOS and PMOS transistors [6,36]. However, in this case, the impact of flicker noise and of the input capacitance [see Equation (1)] may be detrimental.
16. Chae, M.S.; Liu, W.; Sivaprakasam, M. Design Optimization for Integrated Neural Recording Systems. *IEEE J. Solid State Circuits* **2008**, *49*, 1931–1939.
17. Mollazadeh, M.; Murari, K.; Cauwenberghs, G.; Thakor, N. From Spikes to EEG: Integrated Multichannel and Selective Acquisition of Neuropotentials. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2008)*, Vancouver, Canada, 20–25 August 2008; pp. 2741–2744.
18. Joye, N.; Schmid, A.; Leblebici, Y. An electrical model of the cell-electrode interface for high-density microelectrode arrays. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS 2008)*, Vancouver, Canada, 20–25 August 2008; pp. 559–562.
19. Agnes, A. Very low power Successive Approximation A/D Converter for Instrumentation Applications. Ph.D. Dissertation, Università di Pavia, Pavia, Italy, 2009.
20. Razavi, B.; Wooley, B.A.; Leblebici, Y. Design techniques for high-speed, high-resolution comparators. *IEEE J. Solid State Circuits* **1992**, *27*, 1916–1926.
21. Figueiredo, P.M.; Vital, J.C. Kickback noise reduction techniques for CMOS latched comparators. *IEEE Trans. Circuits Syst. II Express Briefs* **2006**, *53*, 541–545.
22. Vibert, J.F.; Costa, J. Spike separation in multiunit records: a multivariate analysis of spike descriptive parameters. *Electroencephal. Clin. Neurophys.* **1979**, *47*, 172–182.
23. Oweiss, K. A Systems Approach for Data Compression and Latency Reduction in Cortically Controlled Brain Machine Interfaces. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1364–1377.
24. Lewicki, M.S. A review of methods for spike sorting: The detection and classification of neural action potentials. *Comput. Meth. Prog. Biomed.* **1998**, *9*, 53–78.
25. Zouridakis, G.; Tam, D.C. Identification of reliable spike templates in multi-unit extracellular recordings using fuzzy clustering. *Comput. Meth. Prog. Biomed.* **2000**, *61*, 91–98.
26. The public database is available at Professor Rodrigo Quiroga's website, <http://www2.le.ac.uk/departments/engineering/research/bioengineering/neuroengineering-lab/spike-sorting/> [37]. The file adopted to validate the compression algorithm is *C-difficult1-noise01.mat*.
27. Bonfanti, A.; Borghi, T.; Gusmeroli, R.; Zambra, G.; Oliyink, A.; Fadiga, L.; Spinelli, A.S.; Baranauskas, G. A low-power integrated circuit for analog spike detection and sorting in neural prosthesis systems. In *Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS 2008)*, Baltimore, MD, USA, 20–22 November 2008; pp. 257–260.
28. Andreani, P.; Fard, A. More on the $1/f^2$ Phase Noise Performance of CMOS Differential-Pair LC-Tank Oscillators. *IEEE J. Solid State Circuits* **2006**, *41*, 2703–2712.

29. Bonfanti, A.; Pepe, F.; Samori, C.; Lacaita, A.L. Flicker noise up-conversion due to harmonic distortion in Van der Pol oscillators. *IEEE Trans. Circuit Syst. I* **2012**, *59*, 1418–1430.
30. The quality factor of the tank is approximately determined by the inductor loss at 400 MHz frequency, giving $Q \cong 70$.
31. Samori, C.; Lacaita, A.L.; Villa, F.; Zappa, F.; Spectrum folding and phase noise in LC tuned oscillators. *IEEE Trans. Circuit Syst. I* **1998**, *45*, 781–790.
32. Carson, J.R. Notes on the theory of modulation. *Proc. Inst. Radio Eng.* **1922**, *10*, 57–64.
33. Tan, C.; Tjhung, T.; Singh, H. Performance of Narrow-Band Manchester Coded FSK with Discriminator Detection. *IEEE Trans. Commun.* **1983**, *31*, 659–667.
34. Lacaita, A.; Levantino, S.; Samori, C. *Integrated Frequency Synthesizers for Wireless Systems*; Cambridge University Press, Cambridge, UK, 2007.
35. Abdelhalim, K.; Genov, R. 915-MHz Wireless 64-Channel Neural Recording SoC with Programmable Mixed-Signal FIR Filters. In *Proceedings of the 2011 European Solid-State Device Research Conference (ESSCIRC)*, Helsinki, Finland, 12–16 September 2011; pp. 223–226.
36. Liew, W.-S.; Zou, X.; Lian, Y. A 0.5-V 1.13- μ W/channel neural recording interface with digital multiplexing scheme. In *Proceedings of the 2011 European Solid-State Device Research Conference (ESSCIRC)*, Helsinki, Finland, 12–16 September 2011; pp. 219–222.
37. Martinez, J.; Pedreira, C.; Ison, M.J.; Quiroga, R.Q. Realistic simulation of extracellular recordings. *J. Neurosci. Meth.* **2009**, *184*, 285–293.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).