2

FULLY INTEGRATED SYSTEMS FOR NEURAL SIGNAL RECORDING: TECHNOLOGY PERSPECTIVE AND LOW-NOISE FRONT-END DESIGN

Andrea Bonfanti, Tommaso Borghi, Guido Zambra, and Andrea L. Lacaita

2.1 INTRODUCTION

Neurological disorders like epilepsy, migraines, multiple sclerosis, and Parkinson's disease represent 35% of all diseases in Europe [1], accounting for about 46 million cases over 466 million EU residents, for an aggregate cost of more than 80 billion euros. On a more global scale, 450 million people are affected by neural disorders and 6.8 million people are estimated to die every year as a result of these pathologies worldwide [2]. Moreover, this toll is expected to quickly rise since disorders such as Parkinson's disease and Alzheimer's are forecasted to massively spread in our aging society.

In this frame, electronics can play an important role contributing to developing new monitoring and diagnostic devices and new rehabilitation and therapeutic tools to restore lost skills or to deliver effective treatment of neural diseases. For example, in the last decade, advances in electroencephalogram (EEG) acquisition and processing made it possible to introduce neural activity monitors in

CMOS Biomicrosystems: Where Electronics Meet Biology, First Edition. Edited by Krzysztof Iniewski.

the clinical practice [3], and a number of therapies and prosthetic systems gradually entered hospitals and clinics to help patients live a better life: deep brain stimulation for Parkinson's disease [4] and cochlear implants [5] are just some examples of neurotechnologies today available in health-care centers worldwide. Despite their widespread adoption, these tools are still in an early stage of development and significant room for improvement is left [6]. This chapter offers an introductory overview of microelectronic systems for neural probes, pointing out the major issues met in deciding their architectures and function partitioning. Then the discussion will focus on the integration of low-power low-noise multichannel amplifiers for extracellular action potential (AP).

2.2 NEURAL SIGNALS AND FRONT-END REQUIREMENTS

Interfacing brain and electronic circuits implies the effective transduction of ionic-driven potential spikes into electronic signals. Requirements on the electronic side are basically set by the anatomy and physiology of the brain cells. A comprehensive description of the neuronal cells, of their morphology, of their working principles, and of their interaction dynamics would certainly be an intriguing journey through a wide spectrum of different topics and disciplines, but it clearly falls outside of the scope of the chapter. Therefore, in following, we will just recall some few key facts and figures, ultimately setting the specs of an implantable neural probe and the requirements for any system for neural monitoring and therapeutics.

Neurons comprise three different parts: a cell body called soma, a number of input branches (or processes) called dendrites that collect information coming from other neurons, and an output process called axon that connects each neuron to other nerve cells or muscles. The soma contains everything, is usually inside a cell body (nucleus, Golgi apparatus, endoplasmic reticulum, mitochondria, liposome, and ribosome), and has a characteristic diameter of 20 μm. Dendrites and axons can reach lengths up to hundreds of centimeters and their diameter defines the speed of information propagation along the nerves. As a matter of fact, neurons can communicate with each other by means of electrical pulses called APs or spikes. As all other cells, they contain high concentrations of K^+, while keeping high concentrations of Na^+ and Ca^{2+} outside the soma. The unbalance of charge across the cell membrane generates the so-called resting potential of –60 mV.

Such a distribution of positive and negative ions is at the basis of cell-to-cell communication: The reason behind this arrangement dates back to the origin of life when the first small single-cell organisms arose in the sea, a Na^+-rich environment. In such a hostile habitat, the development of a Na^+-proof membrane was the only choice nature had to defend the living cells from the environment. On the other hand, Na^+ abundance suggested using this ion charge for signaling. As a matter of fact, the generation of each spike starts with a flow of Na^+ into the cell. The flow is controlled by ion channels, proteins that fold in small pore structures and regulate the flow of different ions across the cellular membrane.

The dynamics of spike generation can be divided into three steps: the initial inflow of Na^+ that depolarizes the neuron (i.e., the inner potential of the cell becomes positive by a few millivolts); an outflow of K^+ ions that hyperpolarizes (i.e., makes the inside potential negative again); and finally, a refractory period, preventing the neuron to fire a new spike for a few milliseconds [7]. Na^+ ions have mobility six orders of magnitude lower than electrons in Si. Their intrinsic low speed as charge carriers has impact on both the treelike morphology of the neurons and the way they connect to each other to give rise to intelligent behavior. Nature chose to link these slow devices in complex redundant networks. If the details of the brain processing algorithm are still to be fully understood, its architecture and morphology were clear since the early experiments at the beginning of the last century: The brain is a highly parallel processor with a high density of slow devices. It follows that a deep investigation of neural mechanisms requires simultaneous sensing of many neurons from a variety of different brain regions.

To really appreciate the dynamics of a network and the role of each neuron, investigators need different tools to zoom in and out at different temporal and spatial resolutions. The overall activity of the neuron population can be readily detected using a few-centimeter-wide electrode on the surface of the skull. As a matter of fact, this was the first kind of information ever recorded from the brain [8]. Unfortunately, these signals (called EEG) are loosely representative of the activity of each individual cell since they are a superposition of thousands or millions of signals from firing neurons filtered and attenuated by the five cortical layers, the skull, the skin, and finally, the sensor, or electrode. Even if EEG signals have been successfully used to study important cerebral processes (such as epileptic seizure recognition [9], anesthesia monitoring, and sleep disorders analysis [3]), they cannot provide key information on smaller-scale phenomena. The highest temporal and spatial resolutions can be achieved with an invasive recording technique referred to as single-unit extracellular AP recording. This method was first pioneered by David Hubel in the late 1950s and has led to the discovery of a lot important properties [10]. In these measurements, a needlelike electrode with a small conductive tip is inserted into the cortex. The tip diameter is about 20 μm close to the size of a single cell. The electrode is usually fixed to a microdrive that precisely controls its position: The closer the tip to a cell, the larger the signal collected as the neuron fires.

2.2.1 Signals and Noise

Extracellular APs are biphasic pulses with a typical duration ranging from 750 and 1000 μsec and a power spectrum residing in a frequency band from a few hertz to 5–10 kHz (see Fig. 2.1). The shape and amplitude of each pulse depend on the relative tip-to-cell position. Biophysical arguments and experimental recordings demonstrated that spikes collected near the soma show a negative trough followed by a positive peak, while extracellular APs collected in proximity of dendrites have peaks followed by a trough [7]. The peak amplitude of extracellular APs ranges from several tens to few hundreds of microvolts, the specific

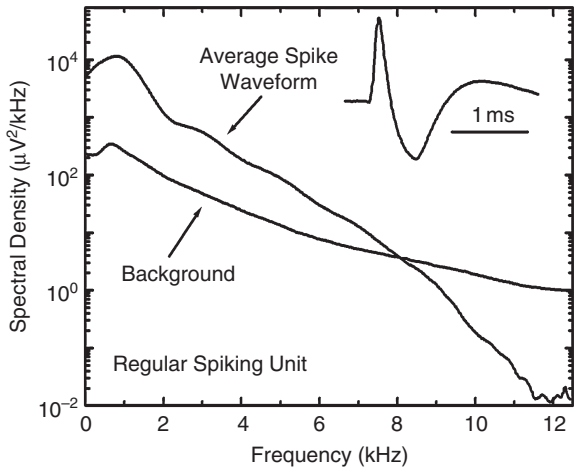

Figure 2.1. Spectral components of the average action potential waveform and power spectrum of background signals and noise collected by neural electrodes. *Inset*: Schematic waveform of action potential pulses.

shape being a complicated function of the electrode materials and geometry as well as the tip-to-cell distance.

This variable is of extreme importance since the spike amplitude rapidly decays as a function of the distance from the cell. In Henze et al. [11] and Pettersen and Einevoll [12], a $1/r^2$ decay was estimated. It follows that a single electrode can collect signals from a small volume around the tip. An experiment performed by Henze et al. [11], using simultaneous recording of intracellular and extracellular signals, demonstrated that the largest spikes are fired at distances <50 μm from the electrode, while at distances >140 μm no distinguishable waveforms are collected.

Given these numbers, each single electrode probing a thin cortical layer (60-μm thick) should receive signals from all the neurons within a 50-μm radius. Taking into account the typical neuron density, the theoretical total number of cells sensed by each electrode should be more than 140. This number is much higher than what is sensed in practice. As a matter of fact, a single electrode usually receives distinctive signals from just three to four neurons. The remaining hundreds or even thousands in its proximity contribute with their tiny APs to the background noise that is the most important source of noise in extracellular recordings [13].

2.2.2 Electrode, Its Noise Source, and General Specs for Front-End Design

In contrast to the broad information content of EEG signals, extracellular recordings offer a window on single cell behavior. The two approaches lay at the oppo-

site, looking either at too big or too small cell populations. A possible solution to this mismatch was proposed in the mid-1960s by Prof. Moll at Stanford who envisioned the use of highly scalable manufacturing techniques such as lithography to produce array of extracellular electrodes [14]. Such an invention offered not only high resolution but also the valuable possibility to simultaneously record from a number of different cells. Since then, silicon-based microelectrode array evolved in two different architectures, the so-called Michigan probe and the Utah array.

The Michigan probe is the natural evolution of the original idea incubated at Stanford. Deep boron diffusion defines the electrode shank, and recording spots are exposed along the shank, thus sensing at different depths in the tissue. Each electrode can be coated with different materials (iridium, gold, and platinum), and up to 64 channels can be packed in a single device. The Utah array is instead a three-dimensional (3D) array made of a matrix of needlelike 1.5-mm-long electrodes [15]. In contrast to the Michigan probe, the recording tips of this array can be either placed all on the same plane or in a 3D arrangement depending on the size of each needle. The structure is obtained by dicing and chemically etching an n-type silicon wafer. The conductive tips are coated with Pt and then insulated with polyamide. Whatever the electrode structure is, the ultimate goal in the design of neural recording systems is the integration of these sensors with receiving and processing circuits on the same substrate in order to reduce the lengths of interconnects that can collect additional noise and interferences. A reasonable volume for such a system can be $4 \times 4 \times 1.5\,\text{mm}^3$. The contact surface with the surrounding tissues is an important parameter. Many studies [16, 17] showed that the power dissipation density of an implantable neural system should not exceed $800\,\mu\text{W/mm}^2$ to prevent tissue overheating and necrosis. For a 16-mm^2 surface, this corresponds to about a 12.5-mW power budget for a complete system.

Electrodes are the physical interface between ionic currents and integrated electronics, and thus their specific characteristics influence the front-end design in terms of noise performance, input impedance, and filter sizing. Only a small number of materials are available for neural electrodes applications: They have to be neurocompatible and conductive and just a handful of materials can match these requirements—stainless steel, platinum and iridium are the most commonly used. Often the exposed tip is processed with other materials (black Pt, iridium oxide, polymers) to increase the surface and thus reducing the overall impedance.

The impedance of the electrodes is not directly defined by the electrical properties of the materials. It rather depends on the particular chemical reactions taking place at the interface between the tip and the electrolyte filling the extracellular matrix, the so-called electrode–electrolyte interface. When a metal is immersed in a conductive solution, it triggers the formation of the Helmholtz layer: a double layer of charged molecules that can be approximately modeled as a resistance with a capacitor placed in parallel [18] (see Fig. 2.2).

The water molecules are attracted to the metal surface (usually negatively charged) forming the Helmholtz inner layer, while another layer of hydrated ions

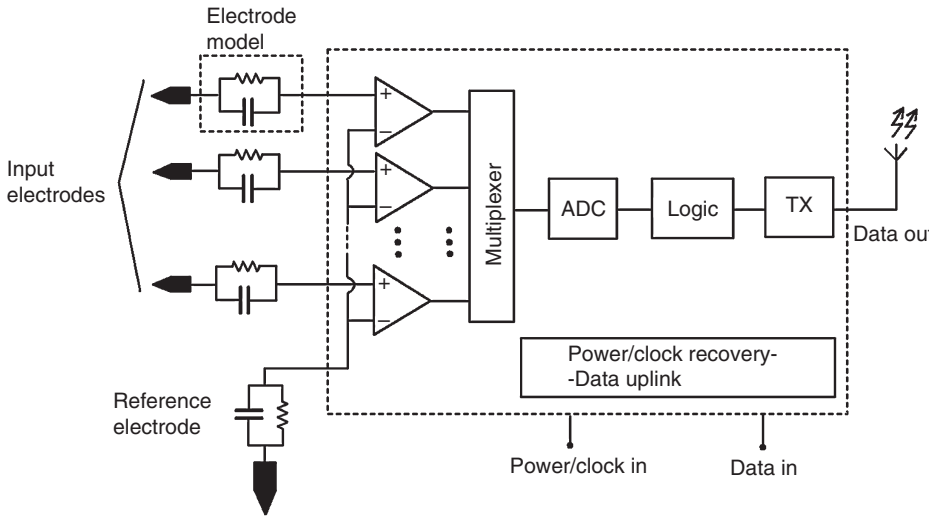

Figure 2.2. Reference architecture of a fully integrated neural recording system. The low-noise analog front end, interfaced with the electrode array, is followed by analog to digital conversion and high-throughput uplink data transmission steps. The power source is typically provided by induction link electromagnetic coupling, which also supports a low bit rate down-link for control purposes.

forms the outer Helmholtz layer. The two opposed charged layers cause a capacitive response, while the small leakage currents flowing through them are accounted by a resistive impedance, $\mathrm{Re}(Z_e(\omega))$, which is dependent on frequency. In the frequency band of interest, the electrode impedance model can be mostly approximated by a single capacitor. Commercially available electrodes for single-unit extracellular recordings feature impedance at 1 kHz ranging from several hundreds of kiloohms up to more than 1 MΩ. These values correspond to a capacitance in the 150–350-pF range. The front-end amplifier should be properly designed to interface such large capacitive impedance without signal losses. On the other hand, the real part of the electrode impedance contributes to thermal noise like an equivalent resistance [19]. It turns out that, for a typical electrode, both the thermal noise due to the electrode impedance and the background fluctuations generate 5–10 μV_{rms} over a 10-kHz bandwidth. These values suggest setting a conservative 5 μV_{rms} as the upper limit for the input-referred rms noise of the front-end amplifier within the signal band.

In addition to noise, the physiological fluctuations of charged carriers in the extracellular matrix as well as additional reactions (binding and unbinding of extracellular proteins on the tip surface) generate a DC bias on the tip that can be as high as hundreds of millivolts. Since neural extracellular signals are 1000 times smaller, a DC removal technique will be required to be able to record the APs. In summary, this introductory discussion defines four key requirements for a neural probe that have impact on the design of the complete system:

1. The parallel strategy of cortical information processing makes multichannel recording mandatory for both monitoring tools and neuroprosthetic devices. An array of electrodes is needed.
2. The neural signal spectrum and the characteristics of the background signals collected by the electrodes immersed in the extracellular matrix require band-pass filtering to remove both DC offset and high-frequency noise.
3. The noise of the electrode and the background noise due to distant firing neurons set the minimum noise requirements of the front-end electronics.
4. The thermal conduction properties of neural tissues determine the total power budget for a complete neural recording system (amplifiers, analog to digital converter [ADC], processing unit, receiver [Rx] and transmitter [Tx]).

These specs, together with additional limitations set by the available technologies, guide system partitioning and force the careful design of all the functional blocks.

2.3 SYSTEM ARCHITECTURE AND POWER BUDGET PARTITIONING

In principle, a neural recording system is a conventional signal acquisition chain comprising an array of sensors and amplifiers, analog-to-digital conversion, a processing unit, and a wireless link (Fig. 2.2). Moreover, a power management block is required to provide stable power supply to the whole system. Most of the design challenges come out from the objective to bring the system to a single chip solution, meeting tight noise specs and power budget limitations. As seen in the previous section, the noise floor of 5 μV is basically set by disturbs picked up from the environment and by the electrode impedance. The tight power budget of 10–15 mW is set by the limited system size and by brain physiology. This Section is now devoted to describe the system architecture, the partitioning of the power budget, and the target performance achievable by the different building blocks.

2.3.1 System Architecture

The small amplitude of the cellular neural signals and the peculiar capacitive impedance of the probe suggest selecting and amplifying the signals immediately, before performing sampling and analog to digital (A/D) conversion. The large potential variations due to background local field potentials (LFPs) and disturbs force the use of differential stages. A reference electrode immersed into the same extracellular environment provides the reference potential. After amplification and filtering, the signals are sampled, digitalized, and transmitted to a backbone receiver. In addition, a power source is required to provide stable power supply to the whole system. This stage may also provide a low-rate downlink for setting

and control signals. Most of the design challenges derive from the goal of squeezing a complex system on a single die with close volume constraints, tight noise specs, and power budget limitations. Increasing pressure on performance is also provided by the expectation to handle a large number of array electrodes. The number of analog channels increases following the array size, thus requiring more power budget. A similar trend is also shared by sampling, A/D conversion, and transmission stages. As the number of input channels increases the overall information throughput, the generated bit rate, will rise, thus demanding more power. It is therefore obvious that full integration of neural recording systems interfacing those large electrode arrays required for high-density neuron-by-neuron interfacing needs to minimize down to the ultimate limits the energy required by analog and digital processing and transmission. Tough challenges are also met in providing energy to the system, which can be easily characterized by current values of several milliamperes, needed under mostly continuous operation. Since use of implanted batteries has to be avoided due to their size, potential toxic composition, and finite lifetime, power transfer by inductive link is seen as the only viable solution to power the implanted circuit [20–24]. This choice makes efficient energy coupling and management an additional requirement for these systems. In summary, matching very low-power consumption while meeting low-noise performance of each analog front-end and parallel operation of a large number of channels in an implantable system is a highly challenging design task. Before going into the details of the analog front-end design, the following paragraphs offer an overview of the state of the art of the different system blocks, pointing out the most recent performance, the contribution to energy savings provided by the long-lasting downsizing of the silicon technologies, and the key points to be addressed by future advances.

2.3.2 Power and Noise Constraints in Amplifier Design

Power consumption of the analog front end is essentially set by the noise requirements and by the g_m/I transistor efficiency. A noise efficiency factor (NEF) was proposed in Steyaert et al. [25] to compare the noise performance of different designs. The starting point is the expression of the total equivalent input noise of a bipolar differential amplifier. Taking into account only thermal noise of the input transistors and neglecting base resistance contribution, the input-referred rms noise, $V_{\text{in,rms}}$, is given by

$$V_{\text{in,rms}} = \sqrt{\text{BW} \cdot \frac{\pi}{2} \cdot \frac{4kT}{g_m}}, \tag{2.1}$$

where BW is the frequency bandwidth. The NEF of the system can be therefore defined as

$$\text{NEF} = V_{\text{in,rms}} \sqrt{\frac{2I_{\text{tot}}}{\pi U_T \, 4kT\text{BW}}}, \tag{2.2}$$

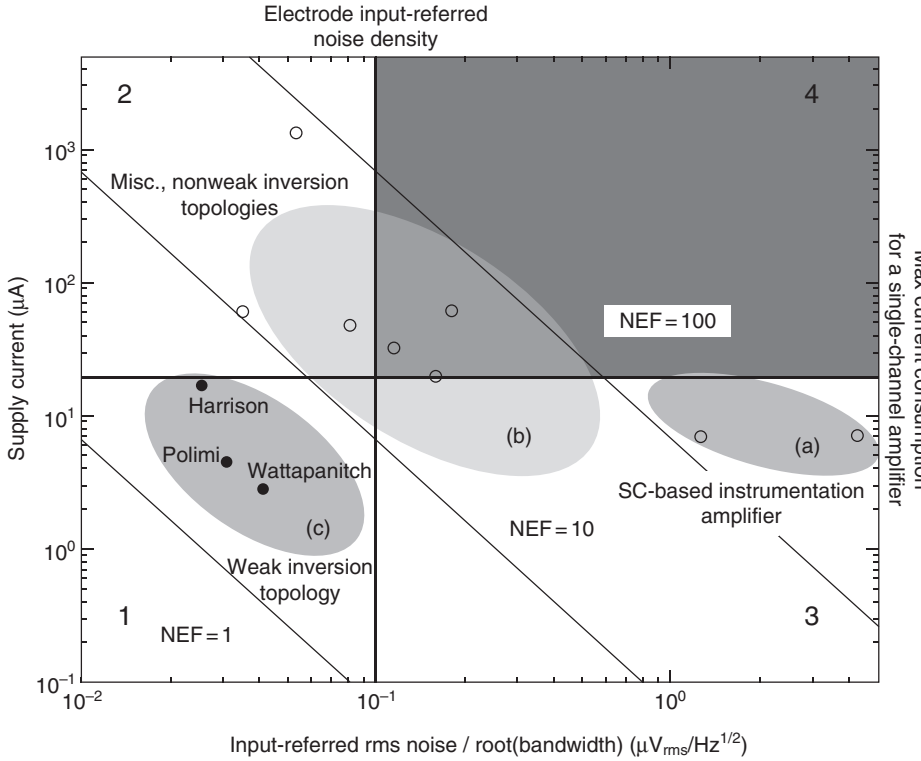

Figure 2.3. Noise performance and current consumption of different OTA topologies. Circuits within Region 1 comply with both the noise and the low-power requirements of neural recording front ends. Constant NEF lines are also drawn to highlight the power–noise trade-off. Circuits exploiting subthreshold operation of CMOS transistors guarantee the best in class performance.

where I_{tot} is the total supply current, $U_t = kT/q$ is the thermal voltage, and BW is the signal bandwidth of the amplifier. Equation 2.2 suggests a theoretical limit of $\sqrt{2}$ for a bipolar differential pair. Since CMOS transistors have a g_m/I ratio lower than the bipolars, the NEF values for CMOS amplifiers, like those considered in the following, is expected to be higher, Figure 2.3 shows the performance of state-of-the-art designs.

The horizontal axis quotes the input noise spectral density, while the current consumption of the amplifying stages runs on the vertical axis. The vertical line drawn at $0.1\,\mu V_{rms}/\sqrt{Hz}$ roughly corresponds to the input noise from the neural probes. The horizontal line at $20\,\mu A$ defines an upper limit of the current budget allowed to these stages. Constant NEF lines highlight instead the noise/consumption trade-off reached by the different solutions. The design plane results clearly divided into four different regions:

- Region 1 includes designs with both noise performance and power consumption matching the application requirements

- Region 2 includes design with good noise performance but high power consumption
- Region 3 comprises amplifiers with very low power consumption but noise floor exceeding the noise of typical neural electrodes
- Region 4 includes designs not matching neither noise nor power specs

The results taken from the literature fall into three different clusters: cluster (a) mainly includes switched capacitor-based instrumentation amplifiers; they operate at reasonable low power but their noise is dominated by the white noise sources of the differential inputs and $1/f$ noise is added when sampling frequency is lowered close to the signal bandwidth (about 8 kHz). For their specific features, they may be used in LFP/EEG recordings where electrode noise poses more relaxed constraints. Cluster (b) includes a variety of low-power design: JFET input buffers, current-feedback single-stage amplifiers, or simple folded cascode topologies where little attention was posed to minimize the white noise contribution of the input stages. Proper transistor sizing and operation in the subthreshold regime, where the g_m/I efficiency of MOS transistors peaks, are instead the key point addressed by designs in cluster (c), which are the best in class results so far reported. It is worth noting that the ultimate NEF value for a differential CMOS analog front end is 2.02. This value is already close to the published results. It follows that even if technology downsizing will proceed further, the current consumption of the analog front end is expected to remain in the order of 5 μA per channel, while power dissipation may marginally benefit from further scaling of the voltage supply.

2.3.3 Analog-to-Digital Conversion

Analog-to-digital conversion follows low-noise amplification. As already pointed out in the previous section, the maximum amplitude for an extracellular AP is about 1 mV, while the minimum signal is approximately 20 μV. Therefore, the ratio between the full-scale range (FSR) and the ADC least significant bit should be better than 1 mV/20 μV ≅ 50, leading to a minimum number of 6 bits. In addition, the ADC resolution should be large enough to prevent quantization noise to impair sensitivity. In other terms, the quantization noise level should be kept well below the electrode noise. This requirement translates into

$$\frac{\mathrm{FSR}}{2^n G\sqrt{12}} << 5\mu V_{\mathrm{rms}}, \tag{2.3}$$

where G is the amplifier gain and n is the bit number. The obvious result is that the higher the gain, the lower the noise added by the ADC quantization. On the other hand, denoting as $A_{\max}$ the maximum amplitude of the extracellular AP, the maximum gain is given by

$$G_{\max} = \frac{\mathrm{FSR}}{A_{\max}}, \tag{2.4}$$

leading Equation 2.3 to be written as

$$\frac{A_{\max}}{2^n\sqrt{12}} = \frac{1\mathrm{mV}}{2^n\sqrt{12}} << 5\mu\mathrm{V}_{\mathrm{rms}} \tag{2.5}$$

$$n >> \log_2 \frac{A_{\max}}{\sqrt{300}\mu\mathrm{V}_{\mathrm{rms}}} \tag{2.6}$$

that gives again a minimum number of 6 bits. In practice, taking into account that the effective number of bit, ENOB, is always lower than the nominal figure, the adoption of an 8-bit ADC will provide a safe margin to the design.

Let us now consider the power consumption (P), which is definitely affected by the requested throughput. The ADC sampling rate is set by the channel parallelism and the signal bandwidth; serving 64 electrodes sampled at 30 kHz requires a sampling rate, F_S, of about 2 Msps. In this frame, an 8-bit resolution translates into a total throughput of about 20 Mbps. Both these values for resolution and sampling frequency are quite moderate and suggest for minimum power, to rely on the effective charge redistribution process of a capacitive successive approximation register (SAR) architecture.

In more general terms, the ability of an ADC to fulfill throughput requirements with minimum power consumption is summarized by a Figure of merit (FOM) defined as

$$\mathrm{FOM} = \frac{P}{2^{\mathrm{ENOB}} \cdot F_S}. \tag{2.7}$$

Such an FOM has empirical ground and takes into account that the ADC power consumption scales with the dynamic range and throughput requirements. It basically gives the energy dissipated per quantization level. Figure 2.4 shows its historical trend for Nyquist rate and oversampled topologies [26], the dashed line corresponding to a twofold efficiency improvement every 18 months. It is clear that technology downsizing is making possible steady improvements of ADC efficiency, with the Nyquist rate architectures benefiting the most [27]. This result may appear somehow surprising since it is well known that, contrary to digital circuits, analog circuits may require larger power dissipation as the voltage supply scales with the technology. In fact, to keep constant the signal-to-noise ratio, the noise level should be improved at the same pace of the voltage headroom reduction. Since noise trades with power dissipation, larger power consumption should be expected. The reason for the opposite trend is essentially that performance of moderate resolution ADCs are not limited by thermal noise but by process constraints (i.e., minimum capacitance that can be integrated) and component

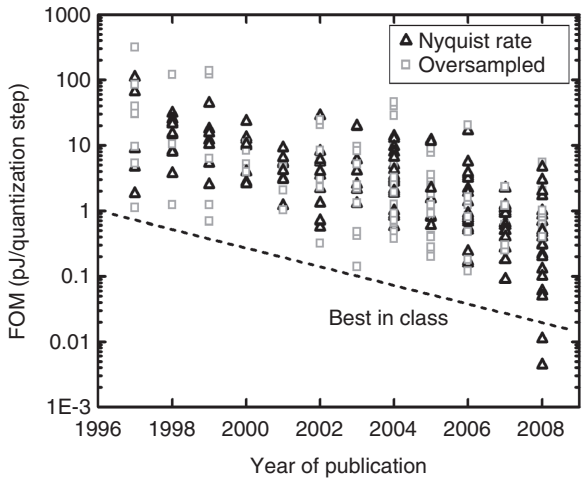

Figure 2.4. A/D Figure of merit: historical trend of Nyquist rate and oversampled A/D. The dashed line corresponds to a twofold efficiency improvement every 18 months.

mismatches [27]. Both these limitations improve with technology scaling, thus making it possible to attain better efficiency. In addition, as the energy cost for digital operations improves with scaling, a new analog design style is emerging. Power-hungry, high-performance amplifiers tend to be replaced by simplified analog stages and more power-efficient digital circuits are introduced to compensate for nonidealities and mismatches. The relentless technology downsizing, the emerging digitally assisted analog design style, the accrual of design experience, and the judicial use of all process options are pushing up ADC efficiency. In this frame, two recent results have clearly step out of the trend line in Figure 2.4 as examples of highly efficient design.

The work of Craninckx and Van der Plas [28] presents a SAR architecture in a 90-nm technology with 65 fJ/quantization step FOM, while the best result in the literature has been achieved using a 65-nm technology, demonstrating a 10-bit, 10-Msps SAR ADC with 4.4 fJ/quantization step [29]. On the other hand, the adoption of less scaled technologies is the reason for the less aggressive results so far reported in neural recording. The systems in Wise et al. [14], Harrison et al. [30], and Gosselin et al. [31] show much higher FOM values, ranging from 2 to 10 pJ/quantization step, the best performance being 2.72 pJ/quantization in a 0.18-μm technology. However, a lot of room is left for energy savings. Even taking a conversion energy of 0.1 pJ/quantization step, well above the latest results in Figure 2.4, the complete digitization of a channel requiring 8-bit resolution and 30-ksps sampling rate translates into a power budget less than 1 μW per channel, in agreement with the target proposed by Zumsteng et al. [32].

It should be also noted that throughput requirements could be relaxed using spike detection and sorting compression algorithms. Just a simple threshold

detector can be indeed used to identify neural spikes [30], lowering to a mere 1–2 kbit/s the data rate needed for each electrode (i.e., less than 150 kbit/s for 64 channels). The drawback of this simple scheme is that no information is given on which neuron is firing, which is a serious limit in neural prosthesis systems where the activity of each neuron has to be isolated to better predict the intended movement [33]. In this perspective, researchers in Olsson and Wise [34] demonstrated a digital circuit that is able to catch the spike features needed for spike sorting; namely, the amplitudes of the peak and the trough as well as the width of the spike [35]. A similar approach, but implemented with very low-power analog circuits, is reported in Bonfanti et al. [36], demonstrating a throughput reduction to 5 kbit/channel and a capability to isolate single neuron activity. However, although featuring interesting performance, spike-sorting approaches have never been used so far in real neurophysiology experiments where careful spike clustering is still mandatory. Spike sorting therefore remains an area where major breakthroughs will be essential to improve performance and reliability of fully implantable neural systems.

2.3.4 Data Link

Data transfer from or to implantable devices is one of the most crucial functional blocks of a brain–machine interface (BMI). The inductive link adopted to supply power may be also used to download signals to control the implantable device. In some cases, this channel has been used for the uplink, even if at a low data rate [37]. Referring to the reference system in Figure 2.2, the most natural choice is therefore to use a high-throughput, low-power TX to transfer the large amount of data generated by the array electrodes, while the downlink, with much lower capacity, is provided through the inductive link. This section is devoted to give an overview of the viable solutions for the high-capacity TX. The next section will instead shortly deal with the inductive links.

The design of high-capacity, low-power data links is a very challenging and hot design issue. For short-range applications, two options are being considered: wireless ultrawide band (UWB) and infrared (IR) transmission. Due to the tight power budget, the key figure is the energy spent per transmitted bit. Figure 2.5 shows the performance of different wireless and IR TX systems taken from the literature and makes possible to appreciate the state of the art in this rapidly evolving field. Note that transmitters for WLAN (IEEE 802.11a/b) and short-range communication protocols (Zigbee and Bluetooth) are very power hungry and do not match the high-throughput requirements. As an example, Bluetooth transmitters need about 100 nJ/bit, most of the energy being spent in baseband processing and in the coherent radio frequency (RF) circuitry. Picoradio solutions are therefore implemented reducing to the essentials the transmitter structure, avoiding coherent transmissions, and using simple communication protocols. This is the case of UWB TXs that works by transmitting nanosecond pulses with frequency components in the 3.1–10.6-GHz band. The transmitter structure has digital pulse generators, while no phase-locked loop (PLL) or tuned filters are

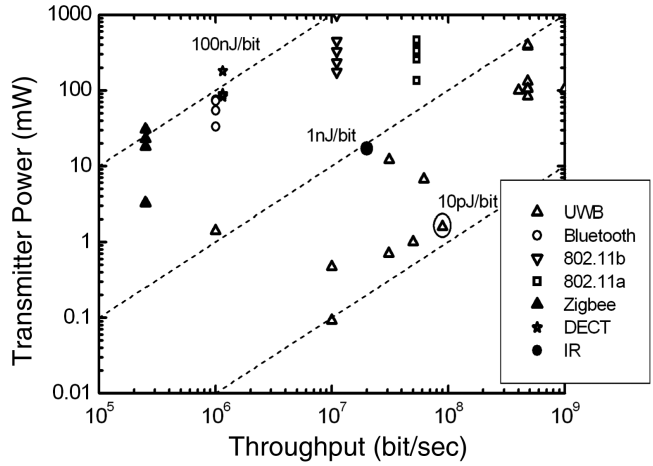

Figure 2.5. Power consumption versus throughput dependence of wireless and IR fully integrated transmitters. The dashed lines highlight the energy needed per bit transmission. The circled triangle refers to the only integrated transmitter for neural recording applications [40].

involved. The power amplifier (PA) has an intermittent operation, thus reducing the power dissipation. Throughput up to 10 Mbit/s has been demonstrated for an RF identification (RFID) tag with less than 10 pJ/bit energy consumption and a communication range of 13.9 m [38].

Note that taking as a reference the 240 kbps generated by each neural recording channel (30-ksps sampling rate, 8-bit resolution), a 10 pJ/bit transmission energy corresponds to an impressive target of just 2.5 μW per channel. It is true that possible shortfalls of UWB transmissions may result from the tissue absorption at such high frequencies, but preliminary simulation studies suggest that such an issue could be fixed in the future [39]. Moreover, encouraging results have been already demonstrated. The UWB transmitter in Chae et al. [40], designed for neural recording systems in a 0.35-μm technology, reaches a 90 Mbit/s throughput with a power consumption of only 1.6 mW (i.e., 17 pJ/bit) even if using an off-chip antenna, which may pose issues in implanted systems. Unfortunately, no details have been given on the antenna size and on the communication range to elaborate further.

In addition to wireless approaches, IR signals are being considered as an alternative. In this solution, the implanted chip is interfaced with a laser positioned under the head skin to exploit the low skin absorption in the IR light spectrum. Figure 2.5 shows the energy efficiency achieved with the IR systems reported in the literature [41]. A power consumption of 20 mW is needed to reach a throughput of 20 Mbps, corresponding to 1 nJ/bit, two orders of magnitude higher than wireless TXs. Moreover, IR transmission requires precise alignment between the laser source and the receiver, and the range is limited to less than 1 cm. At the present stage, we may therefore conclude that, even if further devel-

opments may improve this performance, wireless solutions seem to be more promising.

2.3.5 Power Source

Intensive research is also ongoing on inductive coupling systems for powering implanted circuits. These systems are basically composed of an external transmitter (typically a class E amplifier) driving a coil and an on-chip receiver that picks up the signal reaching a secondary magnetically coupled coil. This coil is tuned to the transmitter frequency with a proper shunt capacitor. The transmitter generates an RF magnetic field in the megahertz range, thus inducing a harmonic voltage in the receiver tank. An RF limiter is adopted to clamp the amplitude of the induced voltage and to protect the implanted circuit from overvoltages. A wideband rectifier converts the AC signal into an unregulated DC supply voltage, and finally, a voltage regulator provides a stable and regulated power supply to the implanted device.

One of the most critical parameters of those systems is the carrier frequency. High frequencies have the advantage of transferring more energy per unit of time and to allow a smaller tank inductor and capacitor. Unfortunately, increasing the carrier frequency determines more current consumption in the receiver and more power absorption in the biological tissue, degrading the efficiency. Typically, the frequency is chosen in the 1–10-MHz range, considering also the need of a particular clock frequency on the chip (e.g., A/D converter clock).

The inductive link is also used for back telemetry, that is, to transmit data to the implantable system. These data are exploited to set and adjust some parameters of the recording system: for example, to select a bank of amplifiers, to control the gain of the amplification chain, or to set the A/D converter resolution. A low-rate control channel is enough for these purposes and can be easily obtained by modulating either the amplitude [20] or the frequency [23] of the power carrier. A transfer rate of 16 kbit/s has been demonstrated in Harrison et al. [24] using an amplitude shift keying modulation of a 2.56-MHz carrier. It is interesting to note that the overall power management system, in a 0.6-μm technology, needed 1 mW to deliver 8 mW of power, while an additional 0.5 mW was required by clock and data recovery circuits. These numbers highlight that the impact of supply coupling and regulation on the overall power budget is very significant. High-efficiency, low-power voltage converters without off-chip components are essential enabling blocks for future implanted biomedical devices. Research breakthroughs in this field are still needed.

2.3.6 System Perspectives

In summary, technologies and circuit designs are opening the way to high-density neural recording systems capable of handling high-throughput transmission of sampled signals derived from hundreds of simultaneously recording sites. The technology trends discussed above show that the 240-kbps throughput generated

by each channel of a neural recording system (30-ksps sampling rate, 8-bit resolution) may require just 5 μW for low-noise amplification and filtering, 1 μW for digitalization, and about 2.5 μW for wireless transmission, leading to an estimate of overall power allocation of about 10 μW per channel. Spike detection and sorting may add further room for system flexibility. In the future, technology scaling is expected to further improve the energy performance of A/D converters and RF transmitters. Advances are certainly needed to improve the overall efficiency of the induction link power source, which otherwise may need up to 10% of the overall power budget. Packaging technologies may be asked to allow the heterointegration of some off-chip components (inductors, antenna, and filters).

Neural recording systems share with other systems on chip most of all these blocks. Therefore, it can be expected that design advances made in the frame of other applications will be also beneficial to these systems. On the other hand, the analog front end is quite peculiar and needs a custom approach. This is due to the large capacitive input impedance of the electrode, the tight input noise floor, and the need to decouple the large DC input offset. The next sections will be therefore devoted to cover the most relevant issues met in the design of these stages.

2.4 PREAMPLIFIER AND FILTER

2.4.1 Preamplifier and Filter Requirements

The small amplitude of extracellular neural signals, the high impedance of the electrode–tissue interface, and the need to reduce noise and disturbs caused by clock artifacts and aliasing impose the performance of amplification and filtering before sampling and converting the signals into the digital domain. In these systems, an integrated preamplifier for neural signals must

1. Have much higher input impedance than the electrode–tissue interface (i.e., much larger than 1 MΩ at 1 kHz of frequency) and negligible DC input current
2. Block the DC offset generated at the electrode–tissue interface (up to 300 mV) to prevent the saturation of the amplifier
3. Select the band where most of the signal power lies, that is 300 Hz–10 kHz [34]; such a band-pass filtering removes both low-frequency signals such as LFP that lays in the 1–100-Hz band and can be as large as several millivolts, and high-frequency noise that can be aliased back into the signal band due to sampling performed by analog multiplexing
4. Have sufficient dynamic range to convey spikes having a peak-to-peak amplitude of up to 500 μV
5. Have a high common-mode rejection to minimize interference from 50–60-Hz power line noise and other common-mode disturbs such as human breath, brain pulsing, heartbeat, and movement artifacts, and a high-power

supply rejection if the power supply noise is significant, for example, for an AC inductive power link

6. Feature an input-referred noise lower than the electrode/background noise that can be as high as $5\,\mu V_{rms}$
7. Have minimum power consumption so as not to impact the overall power budget

This Section is devoted to discussing the design options and the most relevant choices made to comply with these requirements. The result is the best in class low-noise amplifier for neural recording so far reported, which is an improved version of the neural amplifier described in Borghi et al. [42]. Numerical values are given referring to the actual design that was implemented in a 0.35-μm AMS CMOS process with a 3-V power supply.

In order to comply with the common-mode rejection requirements, there is no other choice than using a differential input stage. Its gain has to be a significant fraction of the overall amplification required from the electrode to the A/D converter. Considering a power supply and an FSR for the A/D converter of 3 V, the whole FSR is fully exploited, adopting a total gain of 5000–6000, the best partition being to assign a 1000–2000 gain to the preamplifier and leaving the remaining to the variable-gain amplifier (VGA) placed before the ADC. A higher VGA gain has to be avoided for two reasons: (1) the large bandwidth, in the megahertz range, of this amplifier, which processes signals coming from the different multiplexed channels; (2) so as not to amplify too much clock signal spurs and artifacts as well as power supply noise.

2.4.2 Preamplifier and Filter Architecture

In principle, three possible strategies can be followed in designing the continuous-time high-gain architecture of the preamplifier/filter stage. The stage has to provide enough gain (60–70 dB), and select the signal band removing the DC electrode offset (Fig. 2.6).

The solution depicted in Figure 2.6a shows a high-pass filter, with a cut-off frequency f_{HP}, set to the lower side of the signal band (i.e. $f_{HP} \approx 300$ Hz), followed by a high-gain preamplifier that fixes the upper side of the amplifier band, cutting off noise at frequencies higher than f_{LP}. Figure 2.6b shows an alternative approach where a first amplifier provides moderate gain to prevent saturation of the overall chain. Offset and low-frequency signals at frequencies below f_{HP} are removed at the amplifier output. A second stage is used to boost the gain to the desired level. The third solution (see Fig. 2.6c) features a first amplifier with AC-coupled inputs that implements a high-pass filter with a cutoff frequency f_1 well below f_{HP}, a second high-pass filter that correctly sets the low-frequency cutoff band at f_{HP}, followed by a second gain stage, which also acts as low-pass filter. The architecture in Figure 2.6a is interesting since the limited number of stages may save power. Unfortunately, as it will be explained in Section 2.4.3.3, the noise introduced by

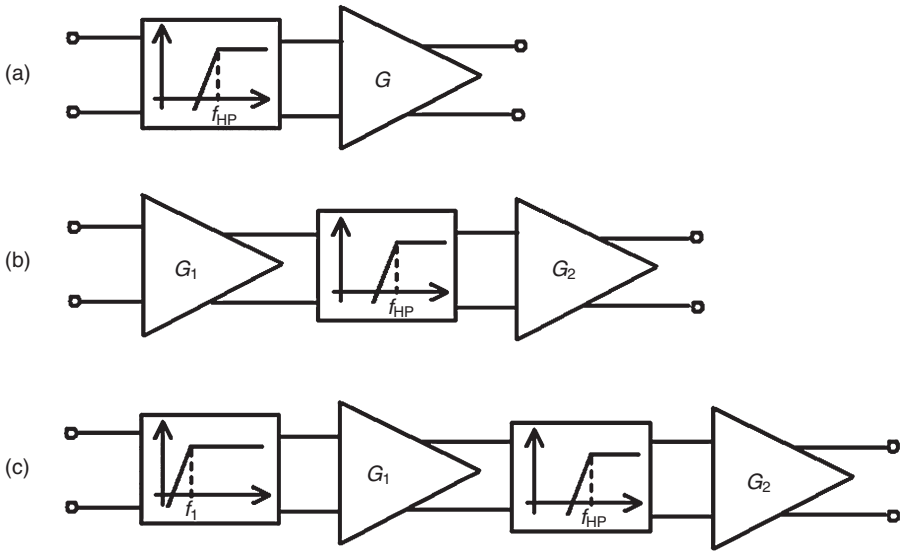

Figure 2.6. Different architectures of the low-noise preamplifier performing signal amplification, offset removal, and cutoff of the high-frequency components exceeding the band of interest.

the high-pass filter impairs its performance. Also, the solution in Figure 2.6b is critical since the possible level of electrode offset limits the first-stage gain to values as low as 2–3, thus preventing the suppression of the noise contributions arising from the following stages. The best solution for the input stage of a neural recording system is therefore the architecture in Figure 2.6c. The very low-frequency high-pass filter does not significantly contribute to the equivalent input noise that is mainly set by the first amplifier. Figure 2.7 shows the schematic of the selected three-stage circuit.

The first stage is an AC-coupled high-pass filter, using two MOS-bipolar pseudo-resistors as feedback elements [43]. This solution enables the synthesis of high-value resistance without using large area components. Figure 2.8 shows the experimental resistance of the pseudobipolar elements used in the first stage.

A peak value close to 0.5 TΩ is reached. It is true that the resistance dependence on the voltage swing may causes distortion; however, in the following, some solutions will be described to circumvent this potential issue. The midband amplifier gain is given by $G_1 \cong -C_1/C_2$. Its value is set to about –67 by taking $C_1 = 10\,\text{pF}$ and $C_2 = 150\,\text{fF}$. The high-pass pole frequency is instead placed below 10 Hz just to reject the offset and the slow voltage drift of the electrode. A $G_{M\text{-}C}$ high-pass filter with a cutoff frequency of about 300 Hz is introduced after the first stage [30]: it properly sets the low-frequency cutoff of the signal band. In addition to the beneficial impact on noise, which will be discussed in the following section, the stage filters out the signals from the background LFPs that can prevent a correct detection of the potential spikes or even saturate the amplifier. A G_M cell

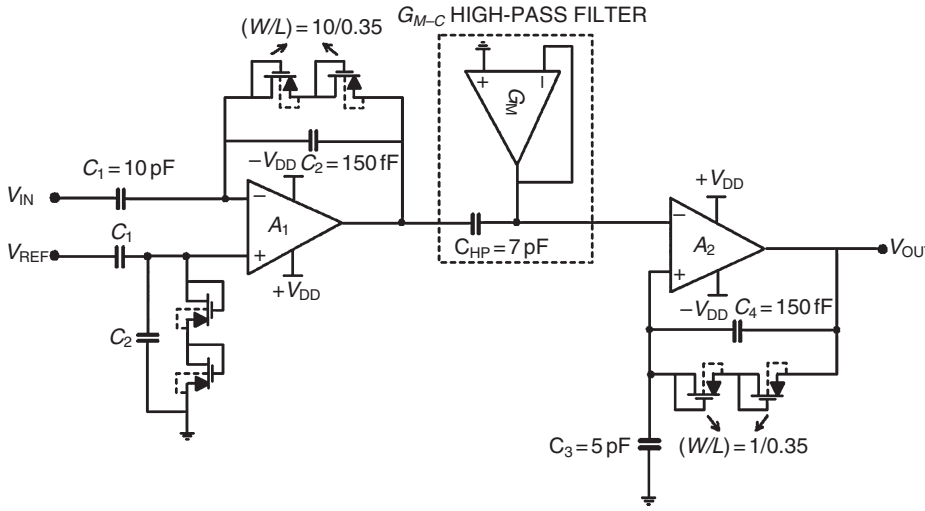

Figure 2.7. Schematic structure of the low-noise neural amplifier, implementing the architecture in Figure 2.6c.

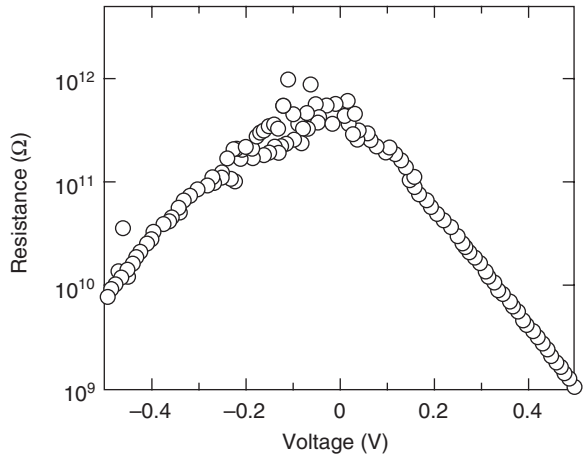

Figure 2.8. Experimental incremental resistance of the pseudoresistor elements implemented in the first stage of the preamplifier in Figure 2.7. Reprinted with permission from Borghi et al. [44]. (Copyright © 2007 IEEE)

was preferred to a tunable subthreshold MOS resistor as in Wattanapanitch et al. [45] because these latter components feature a large spread of the resistance value, even up to one decade over the same die [46]. After the selective high-pass filter, a second noninverting gain stage is added to provide further signal amplification and to define the high-frequency cutoff, f_{LP}. The stage is a single-ended capacitive-coupled voltage amplifier with a gain G_2 of about 33, achieved by using

$C_3 = 5\,\text{pF}$ and $C_4 = 150\,\text{fF}$. A capacitive-coupled structure was preferred to a purely resistive feedback amplifier to minimize the current drawn by the operational transconductance amplifier (OTA) output stage. The DC voltage at the amplifier input is set by the pseudoresistor element in the feedback path. The low-pass cutoff frequency is determined by the gain-bandwidth product of the second operational amplifier ($GBWP_2$) and is set to about $GBWP_2/G_2 = 15\,\text{kHz}$. Note that at very low frequencies, this noninverting stage has a unity gain in order not to amplify the offset of the operational amplifier.

2.4.3 Low-Power Low-Noise Amplifier Design

The noise performance of the whole front-end amplifier depends on the design of the first-stage OTA, its input noise power density being mainly determined by a peculiar compromise between the thermal and the flicker noises of the first stage. 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. (C_p is not drawn in Fig. 2.7. 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, beyond the band of interest.) To simplify the noise analysis, let us now neglect the noise contribution due to the two pseudoresistors. This term will be addressed in a following step. Under this assumption, the input-referred noise of the overall amplifier can be written as

$$\overline{E_{n_{\text{eq}}}^2} = \left(\frac{C_p + C_1 + C_2}{C_1}\right)^2 \cdot \overline{E_{n_{\text{eq OTA}}}^2} = \left(1 + \frac{1}{|G_1|} + \frac{C_p}{C_1}\right)^2 \cdot \overline{E_{n_{\text{eq OTA}}}^2}, \tag{2.8}$$

where $\overline{E_{n_{\text{eq OTA}}}^2}$ is the input-referred voltage noise of the first-stage OTA.

Equation 2.8 highlights that the input parasitic C_p, which is mainly determined by the input transistor dimensions, cannot be increased too much without impairing 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, a judicious trade-off has to be reached to get the best overall performance. Furthermore, Equation 2.8 suggests increasing the gain of the first stage, G_1, maximizing C_1. The limit to its value is set by the source impedance that is in the 150–350-pF range. A C_1 value of 10 pF therefore guarantees high-enough input impedance. To get the target gain of $G_1 = -C_1/C_2 \approx -67$, C_2 is taken equal to 150 fF. For the parasitic C_p, an upper limit of 0.5 pF guarantees enough margins to keep under control the $1/f$ noise at a cost of a marginal 10% increase of the overall amplifier with respect to the OTA input-referred noise (see Eq. 2.8). Due to these design choices $\overline{E_{n_{\text{eq}}}^2}$ and $\overline{E_{n_{\text{eq OTA}}}^2}$ become almost the same. From here, both terms will be therefore denoted with the same symbol, $\overline{E_{n_{\text{eq}}}^2}$.

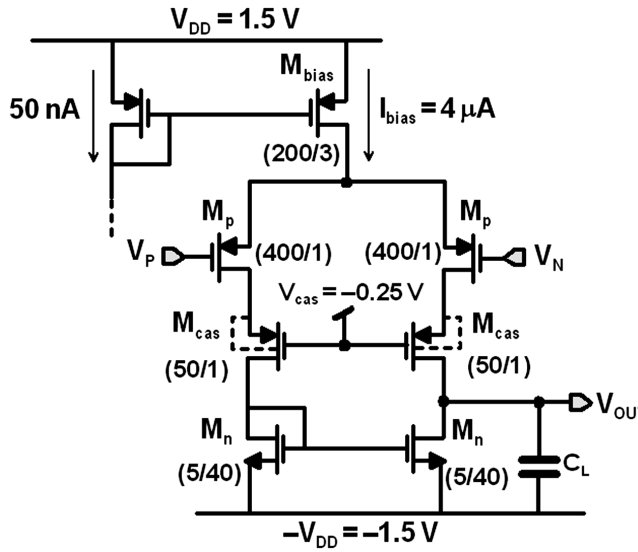

Figure 2.9. Schematic of the telescopic cascode op-amp used in the first stage of the presented amplifier. Reprinted with permission from Borghi et al. [42]. (Copyright © 2008 IEEE)

2.4.3.1 OTA Noise Analysis and Optimization. In order to better exploit the noise–power trade-off, the first OTA stage has been synthesized following a telescopic cascode approach (Fig. 2.9).

This well-known configuration guarantees excellent noise performance, thanks to its very few transistors, while cascoding and proper transistor sizing ensure enough gain.

The input-referred noise power spectral density of the circuit in Figure 2.9 is given by

$$\overline{E_{n_{\text{eq}}}^2} \cong \frac{8kT\gamma}{g_{mp}} + \frac{8kT\gamma}{g_{mp}}\left(\frac{g_{mn}}{g_{mp}}\right) + \frac{2K_p^{(1/f)}}{C'_{ox}W_pL_p}\frac{1}{f} + \frac{2K_n^{(1/f)}}{C'_{ox}W_nL_n}\left(\frac{g_{mn}}{g_{mp}}\right)^2\frac{1}{f}. \tag{2.9}$$

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

$$\overline{E_{n_{\text{eq}}}^2} \cong \frac{8kT\gamma}{g_{mp}} + \frac{2K_p^{(1/f)}}{C'_{ox}W_pL_p}\frac{1}{f}, \tag{2.10}$$

γ being equal to 2/3 for transistors working in strong inversion or $1/(2\kappa)$ in weak inversion ($\kappa \approx 0.7$ [47]). In order to minimize the thermal noise without increasing the current consumption, the M_p transistors have to work in weak inversion, where the transconductance reaches the maximum value. We can estimate the transconductance using the EKV model [48], valid in all regions of inversion, as

$$g_m \cong \frac{\kappa I_D}{U_T} \cdot \frac{2}{1+\sqrt{1+4\times \mathrm{IC}}}, \tag{2.11}$$

where IC is the inversion coefficient [47], defined as

$$\mathrm{IC} = \frac{\kappa I_D}{2\mu C'_{ox}\left(\frac{W}{L}\right)U_T^2}. \tag{2.12}$$

In the subthreshold or weak inversion region, this coefficient is much less than 1. Using Equation 2.11, the overall-amplifier input-referred thermal noise power spectral density results in

$$\overline{E^2_{n\mathrm{th}}} = \frac{8kTU_T}{\kappa^2 I_{\mathrm{bias}}}, \tag{2.13}$$

where I_{bias} is the bias current of the telescopic cascode amplifier. Assuming a first-order roll-off of the frequency response, the noise bandwidth is $(\pi/2)\cdot 15\,\mathrm{kHz} \approx 23.56\,\mathrm{kHz}$. Setting an upper limit of $3\,\mu V_{\mathrm{rms}}$ for the thermal noise contribution in this noise band, a minimal I_{bias} value of about $4\,\mu A$ is required. In order to assure that input pair transistors work in the weak inversion region (IC < 0.1), their form factor has to be less than about 300; we set $(W_p/L_p) = 400$ to be conservative.

The flicker noise contribution can be reduced by increasing the PMOS transistor area ($W_p \cdot L_p$). Setting a noise corner frequency lower than 100 Hz and considering that for the adopted technology $K_p^{(1/f)} \approx 2\times 10^{-26}\ \mathrm{V^2F}$ and $C'_{ox} \approx 5\,\mathrm{fF/\mu m^2}$, the input PMOS transistors were sized with $W_p \cdot L_p = 400\,\mu\mathrm{m}^2$ (thus $W_p = 400\,\mu\mathrm{m}$ and $L_p = 1\,\mu\mathrm{m}$). This sizing leads to a stray OTA input capacitance C_p of approximately 500 fF that does not excessively impair the equivalent input noise.

Moreover, in order to fulfill the requirement $(W/L)_p >> (W/L)_n$, it was set that $W_n = 5\,\mu\mathrm{m}$ and $L_n = 40\,\mu\mathrm{m}$, thus forcing M_n transistors to work in strong inversion (see Table 2.1).

This choice makes it possible to avoid the adoption of a cascoded mirror not to degrade the amplifier gain, while the introduction of M_{cas} guarantees an output resistance given $R_{\mathrm{out}} = r_{0n} \| g_{m\mathrm{cas}} r_{0\mathrm{cas}} r_{0p} \approx r_{0n} \approx 108\,\mathrm{M\Omega}$ and an amplifier gain $g_{mp} r_{0n} \approx 75\,\mathrm{dB}$. In summary, the noise within the amplifier band may be reduced by careful transistor sizing to the sole thermal noise. In this limit, the minimum

TABLE 2.1. Transistor Operating Points

Transistor	$V_{GS} - V_T$ (mV)	IC	g_m (A/V)	r_0 (MΩ)
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

NEF achievable with a single differential stage can be estimated from Equations 2.2 and 2.13 leading to a value of $\sqrt{2}/\kappa = 2.02$. Actually, the contribution to the thermal noise of the current mirror transistors cannot be completely neglected. In fact, even if M_p and M_n transistors work in weak and strong inversion regions, respectively, the transconductance ratio is only about 7 (see Table 2.1) in the present design. Considering the general expression of the transconductance given by Equation 2.11 and the total input referred thermal noise given by the first two terms of Equation 2.9, the theoretical NEF for this preamplifier becomes

$$\mathrm{NEF_{th}} = \frac{\sqrt{2}}{\kappa} \cdot \sqrt{1 + \frac{2\gamma\kappa}{\sqrt{\mathrm{IC}_n}}} \cong 2.15, \tag{2.14}$$

being $\gamma \approx 2/3$, $\kappa \approx 0.7$, and IC_n the inversion coefficient of current mirror transistors. As a reference, consider that the folded cascode amplifier described in Wattanapanitch et al. [45] targets a minimum theoretical NEF = 2.47, while a limit of 2.9 was found in Harrison and Charles [43] for a mirrored cascode op-amp. Note also that in Holleman and Otis [49], the authors propose an amplifier with an NEF = 1.8, but the topology is not differential and thus it has an intrinsic very low power supply rejection ratio (PSRR).

However, the achievable NEF of the proposed amplifier is slightly larger compared with the previously valuated theoretical limit. Three factors combine to raise the *NEF*:

1. 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 2.1), that means an increase of the rms input noise, and also of the NEF, of about 1.034. Note that the technology scaling is beneficial from this point of view since, for a given bias current, the inversion coefficient tends to be lower due to a larger specific oxide capacitance (see Eq. 2.12).
2. The input referred noise of the overall amplifier is a factor $(1 + 1/|G_1| + C_p/C_1)^2$ larger than the input-referred noise of the first operational amplifier, as stated by Equation 2.8. This factor determines an increase of the NEF of about 1.065, considering a parasitic input capacitance of about 0.5 pF, mainly due to the OTA input transistors. A small contribution derives from the strays associated to the input capacitor plates. This contribution can be drastically reduced connecting the capacitor bottom

plate (which has the largest parasitism) to the amplifier input and the top plate to the OTA terminal. In this way, a parasitic capacitance larger than 1.5 pF is avoided.

3. The current drawn in the second amplifying stage contributes to the total current in Equation 2.2 but does not contribute to lowering the input-referred thermal noise. NEF increases by a factor of $\sqrt{1+(I_2/I_1)}$, I_1 and I_2 being the current drawn by the first and the second operational amplifier, respectively. This corresponds to an increase of about 1.037 of the NEF. Considering these three factors, the estimated NEF of the overall amplifier is about 2.45. The adoption of a telescopic cascode stage has to be reconsidered in the frame of very low-voltage technologies. In these cases, a folded cascode topology is more appropriate even if with some penalty on power dissipation and NEF value. Moreover, telescopic cascode is known to have a small output swing. However, most of the time, the values are large enough to accommodate the dynamics of the first amplifying stage. In the 0.35-μm design in Figure 2.9, the positive voltage swing is ≈450 mV ($|V_{T,p}| + V_{cas}$) while the negative swing is about 1V (1.5V – $V_{ov,n}$), posing no issue to the amplifier operation.

2.4.3.2 Second Amplifying Stage. A second gain stage is needed to increase the dynamic range of the input signal before multiplexing. Since this stage requires a relative large output swing, two-stage OTA (not shown in the figures) can be used. Provided that the first stage has enough gain, the impact of this second amplifier on the input-referred noise is negligible; therefore, its power dissipation can be reduced without affecting the noise performance. In fact, the noise of the second OTA is dominated by input transistors, designed to operate in the weak inversion region. Imposing a contribution to the equivalent input noise added by the second-stage OTA 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 a bias current of 200 nA in the first stage, while a current of 100 nA is drawn by the second stage.

Finally, nonlinear distortion may instead be of some concern. Distortion arises from the nonlinear high-resistance pseudoresistor placed in the feedback path, which is driven by a large output voltage swing. For this reason, the subthreshold MOS transistors have been 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 orders of magnitudes lower than the signal through the capacitors. This way, even if the resistance value is modulated under a 1.5-V peak-to-peak output swing, the total harmonic distortion (THD) generated driving the stage with a 1-kHz sinusoidal input was verified to remain less than 5%. Moreover, the application to neural recording is intrinsically more robust to minimal distortion of the impulse response.

2.4.3.3 High-Pass Filter Design and Optimization. Let us now consider the noise due to the MOS-bipolar pseudoresistors that was neglected in Equation

2.9. Denoting as R the small signal resistance of the two PMOS subthreshold transistors, it was experimentally verified that their current noise spectral density complies with the usual equation

$$\overline{S_R^2} = \frac{4kT}{R}, \tag{2.15}$$

taking for R the peak incremental resistance (Fig. 2.8). Their contribution to the OTA input-referred noise power spectral density follows a $1/f^2$ law and is given by

$$\overline{E_{n_R}^2} = 2\frac{4kT}{R}\frac{1}{(2\pi C_1 f)^2} = \frac{8kTf_1}{2\pi G_1^2 C_2 f^2}, \tag{2.16}$$

f_1 being the cutoff frequency of the high-pass filter set by the first amplifier stage. The factor 2 in Equation 2.16 accounts for the two pseudoresistors, one in the feedback path and one connected to the positive input terminal of the operational amplifier. For the sake of simplicity, let us now neglect the presence of the selective high-pass filter following the first stage. By integrating Equation 2.16 from f_1 to the low-pass cutoff frequency of the overall amplifier, f_{LP}, we get

$$\int_{f_1}^{f_{LP}} \overline{E_{n_R}^2} df \cong \frac{8kT}{2\pi G_1^2 C_2}, \tag{2.17}$$

which is independent of the cutoff frequency f_1. From Equation 2.17, the contribution of the pseudoresistors to the OTA input-referred noise results about $2.75\,\mu V_{rms}$. A similar result can be obtained considering the total noise due to the two pseudoresistors at the output of the first stage, $2kT/C_2$, and dividing it by the squared midband gain, $(G_1)^2$. This means an input-referred noise of $3.44\,\mu V_{rms}$. This value is not negligible and explains why it is not convenient to implement the selective high-pass filter (i.e., the 300-Hz cutoff frequency high-pass filter) in the first stage. This solution would add the noise of the pseudoresistors to the other contributions, thus degrading the noise performance. The high-pass filter in the topology of Figure 2.7 is instead able to cut off most of the noise from the pseudoresistors of the first stage. However, a careful choice of the filter capacitor is mandatory to reduce the noise contribution of the $G_{M\text{-}C}$ filter. The point will be discussed in the following. Let us now suppose that $f_1 << f_{HP}$, f_{HP} being the cutoff frequency of the $G_{M\text{-}C}$ high-pass filter. The pseudoresistors of the first stage and the $G_{M\text{-}C}$ filter determine the $1/f^2$ component of the input-referred noise, which can be denoted as $\overline{E_{eq}^2}^{(1/f^2)}$. For frequency larger than f_1, it results

$$\overline{E_{eq}^2}^{(1/f^2)} = \frac{8kT}{R}\frac{1}{(2\pi C_1 f)^2} + \overline{S_{nI}^2}\frac{1}{(2\pi G_1 C_{HP} f)^2}, \tag{2.18}$$

$\overline{S_{nI}^2}$ being the output current noise of the high-pass filter. Let us now assume, for the sake of simplicity, that the current noise generated by the G_M cell ($G_M = 1/R_{HP}$) can be written as $4kT/R_{HP}$. Equation 2.18 becomes

$$\overline{E_{eq}^2}^{(1/f^2)} = \frac{8kTf_1}{2\pi G_I^2 C_2 f^2} + \frac{4kTf_{HP}}{2\pi G_I^2 C_{HP} f^2}. \tag{2.19}$$

Integrating Equation 2.19 in the overall amplifier band, that is from f_{HP} to f_{LP}, it turns out:

$$\int_{f_{HP}}^{f_{LP}} \overline{E_{eq}^2}^{(1/f^2)} df \cong \frac{8kT_1}{2\pi G_I^2 C_2} \frac{f_1}{f_{HP}} + \frac{4kT}{2\pi G_I^2 C_{HP}}. \tag{2.20}$$

The result suggests reducing the first term by minimizing the f_1/f_{HP} ratio. A large capacitor value C_{HP} is also needed to reduce the second term. In practice, the G_M cell is a simple differential stage in which the bias current can be externally tuned, as depicted in Figure 2.10.

Due to the small bias current of 1 nA needed to synthesize a high-value resistance, all the transistors work in the weak inversion region and their transconductance is $g_m = G_M \approx \kappa I/U_T$. The current noise of this configuration is therefore

$$\overline{S_{nI}^2} = 4\frac{2kT}{\kappa} g_m = \frac{8kT}{\kappa R_{HP}}. \tag{2.21}$$

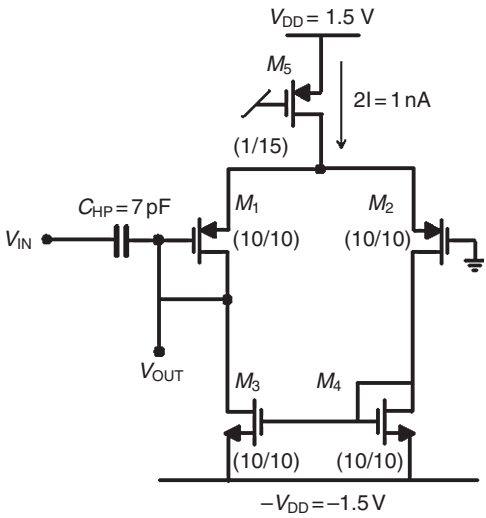

Figure 2.10. Schematic structure of the $G_{M\text{-}C}$ high-pass filter in Figure 2.7. Note the very low bias current needed to synthesize a high-value equivalent resistance.

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 2.20. In summary, in order to keep the input-referred noise due to the first-stage pseudoresistors and the G_M cell in the amplifier band below $1\,\mu V_{rms}$, f_1 has to be at least a factor of 10 smaller than f_{HP} and, at the same time, C_{HP} must be greater than 2 pF. The two pseudoresistors in the circuit of Figure 2.7 were designed to have $f_1 < 10$ Hz (with $f_{HP} = 300$ Hz) and $C_{HP} = 7$ pF.

2.5 SYSTEM INTEGRATION AND RESULTS

The front end was integrated in a wireless neural recording system. Figure 2.11 shows the block diagram of the system, implemented in a 0.35-μm AMS (austriamicrosystems) CMOS process with a 3-V power supply delivered by two coin batteries.

The system was designed to sense the signals provided by an array of 16 neural probes. After amplification, sampling, and conversion, the data are transmitted using an IR laser transmitter. The adoption of an IR transmitter was decided as first step (1) to avoid the broadband external antenna needed in UWB systems and (2) so as not to face skull and skin absorption at high frequencies [39]. The circuit front end features an array of 16 preamplifiers, connected to an analog multiplexer operating at a 40-kHz sampling rate. Multiplexing is performed connecting the output of each preamplifier to a pass-gate switch. A shift register, controlled by a 320-kHz clock (ck_{mux} in Fig. 2.11), sequentially enables each amplifier to access the common data lead; each pass gate is switched on both the rising and the falling edge of multiplexer clock, avoiding clock transitions in the middle of the sampling window. In this topology, each amplifier has to drive a long routing line connecting all the amplifier outputs to the multiplexer. The resulting load capacitance is large (about 1 pF) and has to be driven in a 1.56-μsec time frame. An AB-class buffer is therefore needed to boost each line. Its design was based on a differential flipped voltage follower [50]. To save power, each buffer is switched off when the corresponding amplifier is not selected and turned on one clock edge before, that is, when the previous amplifier is selected. This way, at any time, only two of the 16 buffers are on, draining about 10 μA each from the power supply. Further amplification is provided before sampling by a VGA added to the amplification chain at the multiplexer output. This stage is a noninverting amplifier whose feedback resistor is digitally controlled by 3 bits in order to change the VGA gain from 1 to 8. The amplified signal is then digitized by an 8-bit SAR ADC operating at a clock frequency (ck_{adc} in Fig. 2.11) of 10.24 MHz (640 kS/s) for a 5.12 Mbit/s output stream. The conversion cycle is composed by a sample phase, lasting seven clock cycles (683.6 nsec), followed by an evaluation phase, eight-clock-cycles long. A single bit is evaluated within each clock cycle starting from the most significant bit (MSB). The 16th clock period is devoted to the end of conversion operations and to reset the ADC. The parallel outputs of the converter are then processed by a logic block before to be sent to the laser driver.

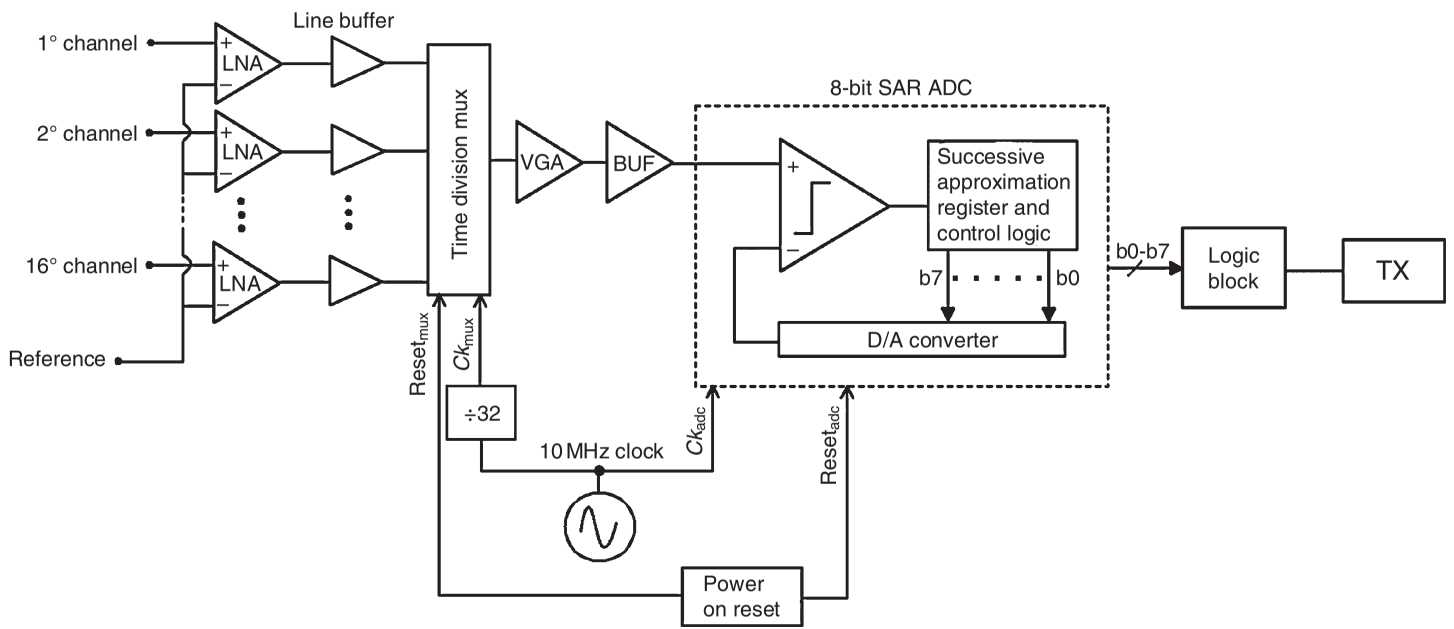

Figure 2.11. Block diagram of the implemented system.

This digital stage provides the serialization of the binary data and the inclusion of service bits (e.g., to indicate the start and stop of each digital word and to recover the information of which channel a byte belongs to). A low-power on-chip relaxation oscillator generates the ck_{adc} signal. After a frequency divider by 32 it also provides the multiplexer clock, ck_{mux}. At the power-up, the negative pulse from a power on reset (POR) stage resets the multiplexer serial register and starts the digital conversion. This reset signal is mandatory to assure proper synchronization between the multiplexer and the sampling phase of the A/D converter since each channel has to be sampled at the end of the corresponding multiplexing period. The system was successfully tested for data transmission at 5.12 Mbit/s through pigskin, about 4-mm thick. The power consumption was 7.7 mW, dominated by the laser driver. A 4-mA current was needed to drive an off-chip 850-nm vertical cavity surface emitting laser (VCSEL) packaged in a TO-18. Transmission was performed using an on-off keying (OOK) modulation. The 6-mW power needed by the transmitter, leading about 1.2 nJ/bit, is in line with the performance of similar transmitters (Fig. 2.5). Before discussing in more detail the performance of the low-noise front end, it is worth noting that the ADC required 570 μW for a sampling rate of 640 kS/s, a differential non-linearity (DNL) < 0.4 least significant bit (LSB), an integral non-linearity (INL) < 0.8 LSB, and an ENOB of 7.3. These values correspond to a power budget of about 35 μW per channel and to an FOM of 5.86 pJ/quantization step. It favorably compares with the best value of 2.72 pJ/quantization step, reported so far in neural recording systems [51].

2.5.1 Amplifier and Filter Characterization

Figure 2.12 shows a die microphotograph of the implemented system with the 16-channel amplifier array, the A/D converter, the logic block, and the laser driver. Each amplifier occupies an area of 0.066 mm^2, line buffer included. Figure 2.13 shows the transfer function of a typical amplifier. The midband gain is 67.1 dB, while the signal band extends from 290 Hz to 15 kHz, close to the expected values. The gain deviation measured on different amplifiers of different dies is less than 0.2 dB, while the high-pass and low-pass frequencies have a standard deviation less than 9 and 4%, respectively. Figure 2.14 shows the measured input-referred noise power spectrum compared with the result obtained simulating the extracted view. It was obtained by dividing the output noise power spectral density by the square of the overall transfer function.

The input-referred thermal noise is about $25\,nV/\sqrt{Hz}$ close to the value expected from the bias current value. Following Wattanapanitch et al. [45], the equivalent input noise is evaluated by dividing the total output noise for the midband gain. Using this procedure for the implemented amplifier and considering the total output noise in the 10 Hz–50 kHz band, the input-referred noise results 3.53 μV, while the measured current consumption is 4 μA for the first-stage OTA and 0.3 μA for the operational amplifier of the second gain stage (the consumption of the $G_{M\text{-}C}$ high-pass filter is negligible). The NEF of the entire amplifier is about 2.41, which is the best result reported to date (see Table 2.2) and close to the value estimated in the previous section for the present design.

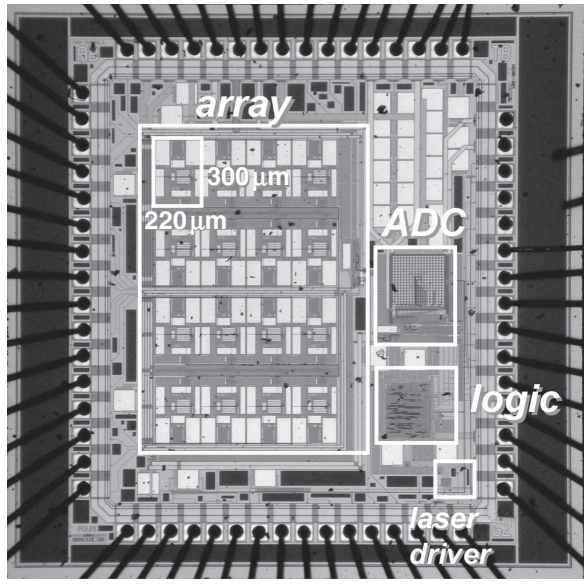

Figure 2.12. Die photo of the integrated system with 16-channel amplifier array, A/D converter, logic block, and laser driver.

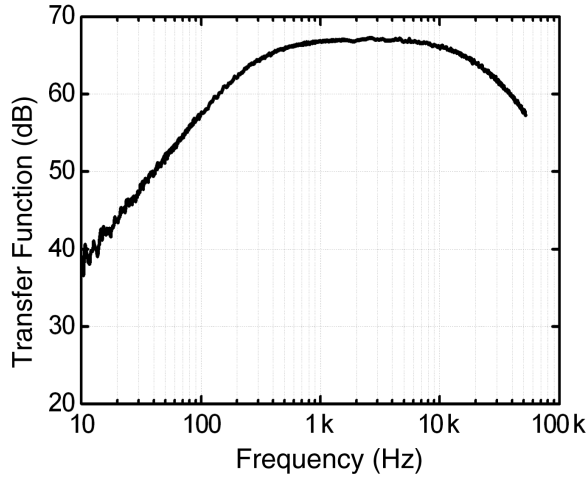

Figure 2.13. Transfer function of the implemented low-noise amplifier. The band ranges from approximately 290 Hz to 15 kHz.

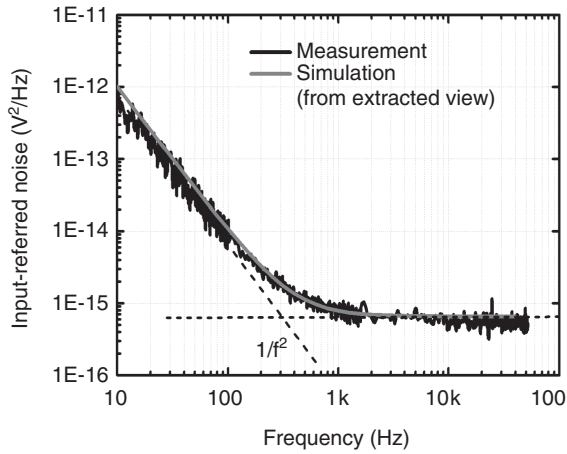

Figure 2.14. Measured and simulated input-referred noise spectra of the low-noise amplifier. The spectrum is obtained dividing the amplifier output noise by the overall transfer function.

TABLE 2.2. Comparison of Low-Power Low-Noise Neural Amplifier Characteristics

Amplifier	Technology (μm)	Bandwidth	Input Noise (μV_{rms})	Supply Current (μA)	Gain (dB)	NEF
Reference [45]	0.5	45 Hz–5.32 kHz	3.06	2.7	40.85	2.67
Reference [43]	1.5	0.025–7.2 kHz	2.2	16	39.5	4.0
Reference [31]	0.18	102–9.1 kHz	5.6	4.67	49.52	4.9
Reference [52]	0.5	94–8.2 kHz	1.94	8	39.6	2.9
Reference [53]	1.5	0.015–4 kHz	3.6	8	39.3	4.9
This work	0.35	290–15 kHz	3.53	4.3	67.1	2.41

To validate the operation of the amplifier under real conditions, the PSRR (defined as the ratio of the differential-mode gain to the gain from the power supply to the output) was measured to be larger than 50 dB in the amplifier band, sufficient to enable proper working even in the most challenging conditions. Finally, the common-mode rejection ratio (CMRR, defined as the ratio of the differential-mode gain to the common-mode gain) was measured to be greater than 65 dB in the signal band (see Table 2.3 for the amplifier performance summary). The amplifier was also tested in *in vivo* experiments on an anesthetized rat. Recordings were performed using an array of electrodes (Tucker Davis) implanted in the primary motor cortex (M1) and with an impedance of about 1 MΩ at 1 kHz. Figure 2.15 shows a trace recorded with the amplifier, while in Figure 2.16 a collection of APs from the same electrode is presented.

TABLE 2.3. Performance Summary of the Low-Power Low-Noise Amplifier

Technology	0.35-μm AMS
Midband gain	67.1 dB
High-pass cutoff frequency	290 Hz
Bandwidth	15 kHz
Input-referred noise	3.53 μV_{rms}
Noise efficiency factor	2.41
THD	5%
CMRR (10 Hz–15 kHz)	65 dB
$PSRR^+$ (30 Hz–15 kHz)	55 dB
$PSRR^-$ (30 Hz–15 kHz)	50 dB
Supply voltage	3 V
Supply current	4.3 μA
Area	0.066 mm^2

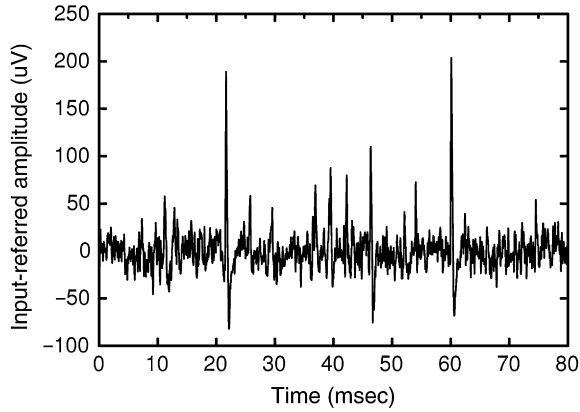

Figure 2.15. Extracellular recording from the primary motor cortex of a rat obtained with the low-noise amplifier.

2.6 REFINEMENTS OF AMPLIFIER DESIGN

The low-noise front end presented here (Fig. 2.7) and those in Harrison and Charles [43] and Wattanapanitch et al. [45] follow the same three-stage scheme of Figure 2.6c: the main difference being the topology of the first OTA stage. All these amplifiers can be classified as passive low-frequency suppression schemes. They use a capacitive feedback network with an AC-coupled input and high-impedance pseudoresistors. For this reason, they need very large integrated capacitors to achieve satisfactory gains. Moreover, nonlinearities arising from

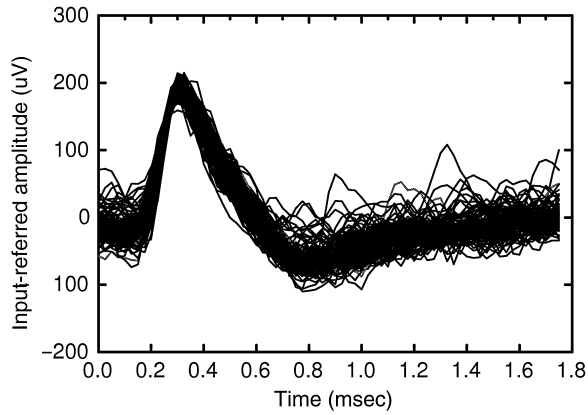

Figure 2.16. Action potential pulses from the trace in Figure 2.15 plotted starting form the same time origin.

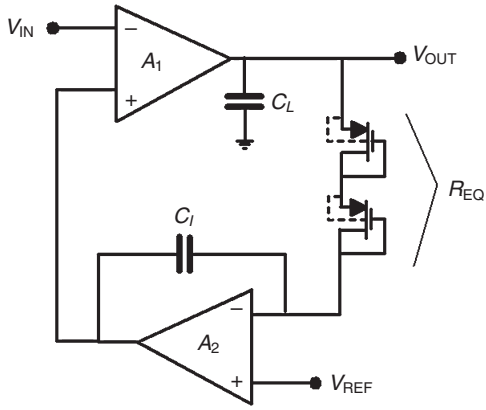

Figure 2.17. Schematic structure of an amplifier performing active low-frequency suppression for minimum area consumption [31].

pseudoresistors must be properly minimized by sizing or additional circuit solutions. This Section is therefore devoted to cover alternative solutions proposed in the literature to reduce the area occupation and to manage pseudoresistor nonlinearities.

2.6.1 Active Low-Frequency Suppression Amplifier

Reduced area occupations can be achieved by adopting the active scheme based on closed-loop control of the DC level proposed by Gosselin et al. [31] and depicted in Figure 2.17.

This system consists of a low-noise OTA (A_1) featuring an inverting Miller integrator within its feedback path. The active integrator is made by a second OTA (A_2), a capacitor (C_I), and high-value resistor (R_{EQ}) implemented with two subthreshold PMOS transistors. Its time constant (τ) sets the –3-dB high-pass cutoff frequency (f_{HP}) of the amplifier, being $\tau = R_{EQ}C_I$. The midband gain of the amplifier equals the low-frequency open-loop gain of the first OTA, A_1, while the –3-dB low-pass cutoff frequency f_{LP} is set by the A_1 dominant pole. The Miller integrator integrates the amplifier output voltage and applies it to the positive input terminal of the first OTA. As a result, this node tracks the DC level at the V_{IN} node, which effectively cancels out any DC offset, drift, and other low-frequency input voltages. The active integrator yields a loop gain greater than unity from the output node to the positive input terminal of the first OTA at low frequency while it provides a strong attenuation at higher frequency, that is, in the neural signal band. This means that in the frequency band of interest the gain of the amplifier is determined by the first OTA that sets the in-band gain and the low-pass cutoff frequency. Considering A_2 as an ideal OTA, the transfer function of the overall amplifier can be written as

$$T(s) = \frac{-s\tau A_{01}}{s\tau + A_{01}} \cdot \frac{1}{1 + s\tau_L}, \tag{2.22}$$

A_{01} being the gain of the first OTA and $\tau_L = 1/(2\pi f_{LP})$ the time constant determined by its output resistance and capacitor C_L. This transfer function yield a –3-dB high-pass cutoff frequency that is set by the gain A_{01} and τ:

$$f_{HP} = \frac{1}{2\pi} \frac{A_{01}}{R_{EQ}C_I}. \tag{2.23}$$

Clearly, the DC and the low-frequency suppression performance of the amplifier are determined by the second OTA gain, A_{02}, which was chosen by the authors to be 20 kV/V. Other parameters of their implementation are: $A_{01} = 400$, $C_L = 400$ fF, $C_I = 1$ pF, $f_{HP} = 100$ Hz, and $f_{LP} = 9$ kHz. The main advantages of this amplifier topology are the high input impedance that is not determined by a large capacitor as in traditional AC-coupled amplifier and the small area (0.050 mm^2, while the amplifier described in the previous Section is 0.066 mm^2); only two relatively small capacitors are required just to set the high-pass and the low-pass poles of the band-pass filter. The amplifier gain is determined by the open-loop gain of an OTA, which is the major drawback of this topology since it cannot be well controlled. Gosselin et al. [31] measured a gain of 50.1 dB with a standard deviation of 0.6 dB in a same die and about 2 dB considering different batches. Another drawback of this topology is related to the noise-power efficiency of the overall amplifier. In fact, both operational amplifiers determine the equivalent input noise of the scheme in Figure 2.17. Since in the neural signal band the second OTA can be considered closed in a buffer configuration, the input-referred noise power spectral density can be approximated as

$$\overline{E_{n_{eq}}^2} = \overline{E_{n_{A1}}^2} + \overline{E_{n_{A2}}^2}. \tag{2.24}$$

Considering γ as the ratio between the A_1 and A_2 input stage bias current, the achievable NEF with this active low-frequency suppression scheme is a factor of $(1+\gamma)/\sqrt{\gamma}$ larger than the one evaluated in Section 2.4.3. For $\gamma = 1$, that is, with OTAs biased with the same current, the minimum NEF is about 4.04, while Gosselin et al. [31] get an NEF of 4.6 biasing the second OTA at a current two times larger than the first OTA for loop stability reasons. In conclusion, an active suppression of the low-frequency components makes it possible to reduce the silicon real estate, avoiding the adoption of large capacitors, but at the cost of doubling the target NEF.

2.6.2 Low Signal Distortion Amplifier

The amplifier presented in Section 2.4.2 uses a MOS-bipolar pseudoresistor in the feedback path to provide a very low cutoff frequency. However, as seen in Figure 2.8, the pseudo-MOS resistance is highly dependent on the output signal level, which may result in signal distortion and high-pass filter cutoff frequency variation. The same problems arise in the solution adopted by Olsson et al. [54] where an NMOS pseudoresistor with controlled gate voltage is used to achieve a tunable high-pass cutoff frequency, or in Aziz et al. [55] where a PMOS is used instead of an NMOS transistor. A solution for synthesizing a high-value resistance independent of the signal level is presented in Yin and Ghovanloo [53] and described in Figure 2.18.

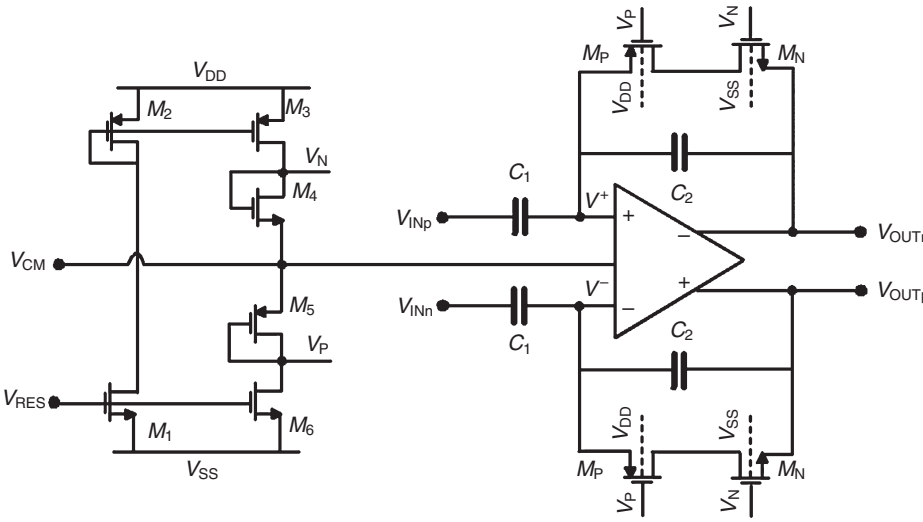

Figure 2.18. Schematic structure of an amplifier with minimal signal distortion [53].

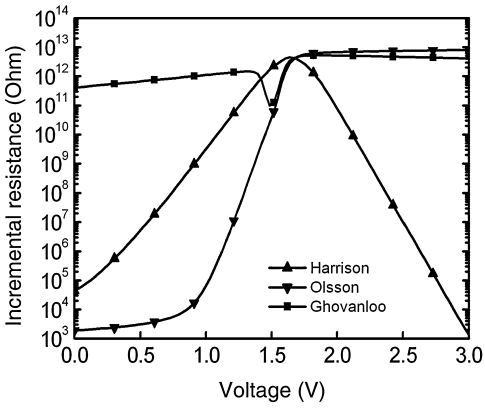

Figure 2.19. Comparison between the incremental resistance of different pseudoresistor elements. Harrison refers to the pseudoresistor used in the presented design and in Harrison and Charles [43], Olsson to a gate-controlled NMOS [54], and Ghovanloo to the scheme presented in Yin and Ghovanloo [53].

Yin and Ghovanloo [53] used a voltage-controlled PMOS and NMOS pseudoresistor to create a stable low cutoff frequency in a fully differential amplifier. In this configuration, an external control voltage, V_{RES}, biases M_1 in the deep subthreshold region; the same current passes through M_4 and M_5, which are diode-connected and sized with large W/L. Noting that the input terminal of the amplifier are kept around V_{CM} by the common feedback network, it can be shown that M_4 and M_5 keep the current passing through the PMOS-NMOS pseudoresistors (M_P and M_N) around k times $I_{D4} = I_{D5}$, k being the ratio $(W/L)_N/(W/L)_4 = (W/L)_P/(W/L)_5$, regardless of the output voltage variations. In fact, when the voltage at the positive output node, $V_{\mathrm{OUT}p}$, is lower than the voltage at the negative input terminal of the OTA, V_n, a current kI_{D5} passes from V^- to $V_{\mathrm{OUT}p}$ because M_P and M_5 have the same gate to source voltage; when $V_{\mathrm{OUT}p} > V^-$, the current passing from $V_{\mathrm{OUT}p}$ to V^- keeps the voltage between M_P and M_N, V_X, slightly higher than V^-, and so $V_{\mathrm{GS}n} = V_N - V_X \approx V_N - V^- = V_{\mathrm{GS4}}$. Therefore, the current passing through the pseudoresistor $M_4 - M_5$ is k times the current in M_4 and M_5. This explains how the resistance of the NMOS-PMOS pseudoresistor is almost constant even for the large voltage signal applied to M_P and M_N transistors. Figure 2.19 shows a comparison between simulated incremental resistances of different MOS pseudoresistors used in Yin and Ghovanloo [53], Olsson et al. [54], and Harrison et al. [30], the latter being the same pseudoresistor adopted in the first stage of the amplifier described in Section 2.4.2. The comparison is made setting transistor sizes and gate voltages in order to have the same resistance at small voltage. The NMOS-PMOS pseudoresistor in Yin and Ghovanloo [53] shows the minimum resistance deviation for the whole range of applied signals, while the other topologies present a variation of different decades. This solution can be therefore used to lower signal distortion in high-gain amplifiers with capacitive-coupled topologies such as the one in Figure 2.7, instead of using a second stage with pure linear resistors.

2.7 CONCLUSIONS

Since the dawn of microelectronic industry, integrated technologies have been fueling tremendous advances in science, engineering, and applications, leading to an increasing inclusion of intelligence in infrastructures, equipment, and products. This trend, leveraging on silicon device miniaturization, is still ongoing and is having a profound impact in all fields, medical science and therapeutics included. In the forthcoming years, the availability of decananometer silicon technologies, and the advances in micromechanical and packaging manufacturing, energy-conversion techniques, and material engineering are expected to provide the solutions needed to develop fully miniaturized, low-power, energy-autonomous smart systems. These systems will promote a more intimate smart link between humans, from a high-level interaction down to the cellular level, "things," and environment. Implantable recording systems are a challenging test field for deeply scaled technologies since demanding performance is required for its application and by the tight constraints imposed by the surrounding environment, that is, the body. But big challenges translate in big opportunities: The potentials of this trend are already clearly visible, neurotechnology being one of the leading examples. Technological advances are enabling innovative interfaces between neurons and electronics, opening the way to new therapeutic devices for neurological diseases as well as to detailed investigation tools of the cognitive processes. The chapter reviewed the performance requirements and the perspectives of fully integrated neural recording systems, pointing out the issues faced in the definition of optimal architectures and function partitioning. In this frame, energy efficiency and low-noise design are key ingredients. The fundamental metrics to quantitatively judge the trade-off between noise, power consumption, and processing speed have been introduced and adopted to compare the most recent system implementations. It has been shown that a 10-μW power budget target per sensing channel is attainable by using cutting-edge technologies and careful design. Finally, we focused on the particular issue of neural amplifier design: Leveraging on a detailed breakdown of the noise sources and by means of an insightful design strategy, we addressed the problem of noise–power trade-off and we presented a neural amplifier that achieved the best performance so far reported.

REFERENCES

[1] P. Andlin-Sobocki, B. Jonsson, H. Wittchen, and J. Olesen, "Costs of disorders of the brain in Europe," *Eur. J. Neurol.*, 12, pp. 1–27, 2005.

[2] WHO, *Neurological Disorders: Public Health Challenges*. Geneva: WHO Press, 2007.

[3] S. Tong and N. Thakor, *Quantitative EEG Analysis Methods and Clinical Applications*. Norwood, MA: Artech House, 2009.

[4] S. Breit, J. Schulz, and A. Benabid, "Deep brain stimulation," *Cell Tissue Res.*, 318, pp. 275–288, 2004.

[5] R. Briggs, H. Eder, P. Seligman, R. Cowan, K. Plant, J. Dalton, D. Money, and J. Patrick, "Deep brain stimulation," *Otol. Neurotol.*, 29(2), pp. 114–119, 2008.

[6] D. Di Lorenzo and J. Bronzino, *Neuroengeneering*. Boca Raton, FL: CRC, 2007.[56]

[7] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neuroscience*. 4th edition, chapter 9. New York: McGraw-Hill, 2000.

[8] G. Buzsaki, *Rhythms of the Brain*. New York: Oxford University Press, 2006.

[9] A. Geva and D. Kerem, "Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervisioned fuzzy clustering," *IEEE Trans. Biomed. Eng.*, 45(10), pp. 1205–1216, 1998.

[10] G. Buzsaki, "Large-scale recording of neuronal ensembles," *Nat. Neurosci.*, 7, pp. 446–451, 2004.

[11] D. Henze, Z. Borhegyi, J. Csicsvari, A. Mamiya, K. Harris, and G. Buzsaki, "Intracellular features predicted by extracellular recordings in the hippocampus *in vivo*," *J. Neurophysiol.*, 84, pp. 390–400, 2000.

[12] K. Pettersen and G. Einevoll, "Amplitude variability and extracellular low-pass filtering of neuronal spikes," *Biophys. J.*, 94, pp. 784–802, 2008.

[13] K. Harris, D. Henze, J. Csicsvari, H. Hirase, and G. Buzsaki, "Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements," *J. Neurophysiol.*, 84, pp. 401–414, 2000.

[14] K. Wise, D. Anderson, J. Hetke, D. Kipke, and K. Najafi, "Wireless implantable microsystem: High density electronic interfaces to the nervous system," *Proceedings of the IEEE*, pp. 76–97, 2004.

[15] P. Campbell, K. Jones, R. Hubert, K. Horch, and R. Normann, "A silicon-based, three dimensional neural interface: Manufacturing processes for an intracortical electrode array," *IEEE Trans. Biomed. Eng.*, 38(12), pp. 758–768, 1991.

[16] T. Seese, H. Harasaki, G. Saidel, and C. Davies, "Characterization of tissue morphology, angiogenesis, and temperature in the adaptive response of muscle tissue to chronic heating," *Lab. Invest.*, 78, pp. 1553–1562, 1998.

[17] K. Silay, C. Dehollain, and M. Declercq, "Numerical analysis of temperature elevation in the head due to power dissipation in a cortical implant," *Proceedings of the 2008 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 951–956, 2008.

[18] T. Jochum, T. Denison, and P. Wolf, "Integrated circuit amplifiers for multi-electrode intracortical recording," *J. Neural Eng.*, 6(1), p. 012001 (26pp), 2009.

[19] R. Gesteland, B. Howlandt, J. Y. Lettvin, and W. H. Pittsk, "Comments on microelectrodes," *Proceedings of the IRE*, pp. 1856–1862, 1959.

[20] N. Neihart and R. Harrison, "Micropower circuits for bidirectional wireless telemetry in neural recording applications," *IEEE Trans. Biomed. Eng.*, 52(11), pp. 1950–1959, 2005.

[21] M. Ghovanloo and S. Atluri, "A wide-band power-efficient inductive wireless link for implantable microelectronic devices using multiple carriers," *IEEE Trans. Circuits Syst.*, 10, pp. 2211–2221, 2007.

[22] A. Sodagar, K. Najafi, K. Wise, and M. Ghvanloo, "Fully-integrated CMOS power regulator for telemetry-powered implantable biomedical microsystems," *Proceedings of the CICC*, 2006, pp. 659–662.

[23] M. Ghovanloo and K. Najafi, "A wideband frequency-shift keying wireless link for inductively powered biomedical implants," *IEEE Trans. Circuits Syst.*, 51(12), pp. 2374–2383, 2004.

[24] R. Harrison, R. Kier, B. Greger, F. Solzbacher, C. Chestek, V. Gilja, P. Nuyujukian, S. Ryu, and K. V. Shenoy, "Wireless neural signal acquisition with single low-power integrated circuit," *Proceedings of the ISCAS*, pp. 1748–1751, 2008.

[25] J. Steyaert and W. Sansen, "A micropower low-noise monolithic instrumentation amplifier for medical purposes," *IEEE J. Solid-State Circuits*, 22, pp. 1163–1168, 1987.

[26] B. Murmann, *ADC performance survey 1997–2008*." Available: http://www.stanford.edu/~murmann/adcsurvey.html.

[27] B. Murmann, "A/D converter trends: Power dissipation, scaling and digitally assisted architectures," *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC)*, pp.105–112, 2008.

[28] J. Craninckx and G. Van der Plas, "A 65fJ/conversion-step 0-to-50 Ms/s 0-to-0.7 mW 9b charge-sharing SAR ADC in 90 nm digital CMOS," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 246–247, 2007.

[29] M. Elzakker, E. Tuijl, P. Geraedts, D. Schinkel, E. Klumperink, and B. Nauta, "A 1.9 nW 4.4fJ/conversion-step 10b 1 Ms/s charge-redistribution ADC," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 244–245, 2008.

[30] R. R. Harrison, P. T. Watkins, R. Kier, R. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE J. Solid-State Circuits*, 42(1), pp. 123–133, 2007.

[31] B. Gosselin, M. Sawan, and C. A. Chapman, "A low-power integrated bioamplifier with active low-frequency suppression," *IEEE Trans. Biomed. Circuits Syst.*, 1(3), pp. 184–192, 2007.

[32] Z. Zumsteg, C. Kemere, S. O'Driscoll, G. Santhanam, R. Ahmed, K. Shenoy, and T. Meng, "Power feasibility of implantable digital spike sorting circuits for neural prosthetic systems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 13, pp. 272–279, 2005.

[33] L. Hochberg, M. Serruya, G. Friehs, J. Mukand, M. Saleh, A. Caplan, A. Branner, D. Chen, R. Penn, and J. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, 442, pp. 164–171, 2006.

[34] R. Olsson III and K. Wise, "A three-dimensional neural recording microsystem with implantable data compression circuitry," *IEEE J. Solid-State Circuits*, 40, pp. 2796–2804, 2005.

[35] J. Vibert and J. Costa, "Spike separation in multiunit records: A multivariate analysis of spike descriptive parameters," *Electroencephalogr. Clin. Neurophysiol.*, 47, pp. 172–182, 1979.

[36] A. Bonfanti, T. Borghi, R. Gusmeroli, G. Zambra, A. Oliyink, L. Fadiga, A. S. Spinelli, and G. Baranauskas, "A low-power integrated circuit for analog spike detection and sorting in neural prosthesis systems," *Proceedings of the BioCAS*, pp. 257–260, 2008.

[37] Z. Tang, B. Smith, J. Schild, and P. Peckham, "Data transmission from an implantable biotelemeter by load-shift keying using circuit configuration modulator," *IEEE Trans. Biomed. Eng.*, 42, pp. 524–528, 1995.

[38] M. Baghaei-Nejad, D. Mendoza, Z. Zou, S. Radiom, G. Gielen, L.-R. Zheng, and H. Tenhunen, "A remote-powered RFID tag with 10 Mb/s UWB uplink and –18.5 dBm

sensitivity UHF downlink in 0.18 μm," *IEEE Int. Solid-State Circ. Conf.*, pp. 198–199, 2009.

[39] M. Klemm and G. Troester, "EM energy absorption in the human body tissues due to UWB antennas," *Prog. Electromagnetics Res.*, 62, pp. 261–280, 2006.

[40] M. Chae, W. Liu, Z. Yang, T. Chen, J. Kim, M. Sivaprakasam, and M. Yuce, "A 128-channel 6 mW wireless neural recording IC with on-the-fly spike sorting and UWB transmitter," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 146–148, 2008.

[41] K. Guillory, A. Misener, and A. Pungor, "Hybrid RF/IR transcutaneous telemetry for power and high-bandwidth data," *Proceedings of the IEEE EMBS Conference*, pp. 4338–4340, 2004.

[42] T. Borghi, A. Bonfanti, R. Gusmeroli, G. Zambra, and A. S. Spinelli, "A power-efficient analog integrated circuit for amplification and detection of neural signals," *Proceedings of the EMBC 2008*, pp.4911–4914, Vancouver, 2008.

[43] R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE J. Solid-State Circuits*, 38(6), pp. 958–965, 2003.

[44] T. Borghi, A. Bonfanti, G. Zambra, R. Gusmeroli, A. S. Spinelli, and G. Baranausks, "An integrated low-noise multichannel system for neural signals amplification," *Proceedings of the IEEE ESSCIRC*, Munich (Germany), pp.456–459, 2007.

[45] W. Wattanapanitch, M. Fee, and R. Sarpeshkar, "An energy-efficient micropower neural recording amplifier," *IEEE Trans. Biomed. Eng.*, 1(2), pp. 136–147, 2007.

[46] Y. Perelman and R. Ginosar, "An integrated system for multichannel neuronal recording with spike/LFP separation, integrated A/D conversion and threshold detection," *IEEE Trans. Biomed. Eng.*, 54(1), pp. 130–137, 2007.

[47] E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operations," *IEEE J. Solid-State Circuits*, 12(3), pp. 224–231, 1977.

[48] C. Enz, F. Krummenacher and E. Vittoz, "An analytical MOS transistor model valid in all region of operation and dedicated to low-voltage and low-current applications," *Analog Integr. Circuits Process.*, 8, pp. 83–114, 1995.

[49] J. Holleman and B. Otis, "A sub-microwatt low-noise amplifier for neural recording," *Proceedings of the 2007 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3930–3933, 2005.

[50] R. Carvajal, J. Ramirez-Angulo, A. Lopez-Martin, A. Torralba, J. Galan, A. Carlosena, and F. Chavero, "The flipped voltage follower: A useful cell for low-voltage low-power circuit design," *IEEE Trans. Circ. Syst.*, 52(7), pp. 1276–1291, 2005.

[51] B. Gosselin, A. Ayoub, J.-F. Ry, M. Sawan, F. Lepore, A. Chaudhuri, and D. Guitton, "A mixed-signal multichip neural recording interface with bandwidth reduction," *Proceedings of the BioCAS*, pp. 129–141, 2009.

[52] M. Mollazadeh, K. Murari, G. Cauwenberghs, and N. Thakor, "Micropower MOS integrated low-noise amplification, filtering and digitization of multimodal neuropotentials," *IEEE Trans. Biomed. Circuits Syst.*, 3(1), pp. 1–10, 2009.

[53] M. Yin and M. Ghovanloo, "A low-noise preamplifier with adjustable gain and bandwidth for biopotential recording applications," *Proceedings of the ISCAS*, pp. 321–324, 2007.

[54] R. H. Olsson, D. L. Buhl, A. M. Sirota, G. Buzsaki, and K. D. Wise, "Band-tunable and multiplexed integrated circuits for simultaneous recording and stimulation with microelectrode arrays," *IEEE Trans. Biomed. Eng.*, 52, pp. 1303–1311, 2005.

[55] J. N. Y. Aziz, K. Abdelhalim, R. Shulyzki, R. Genov, B. Bardakjian, M. Derchansky, D. Serletis, and P. L. Carlen, "256-channel neural recording and delta compression microsystem with 3D electrodes," *IEEE J. Solid-State Circuits*, 44(3), pp. 995–1005, 2009.