Chapter 2
Neural Recording Front-End Design

2.1 Introduction

Neural signal recording revolutionizes our understanding of the human brain. Since the first extracellular recording pioneered by the investigators Ward and Thomas in the 1950s [70], neural recording has revealed the fundamental structure and organization of the brain. The number of simultaneously recorded neurons doubled approximately every 7 years [71]. The exponential growth in the recording ability is to a large extent driven by innovations in CMOS technology, circuits and systems design, microelectrode fabrication, and bio-compatible packaging techniques.

The large-scale neural recording also provides a unique opportunity for the research in brain–machine interface (BMI), which builds the interface between brain and artificial devices [70, 72]. However, recent studies estimate that a simultaneous invasive recording of 100,000 neurons is needed for decoding full-body movements [69], which is beyond the recording ability of the cutting-edge BMI devices. At the same time, multi-channel recording from freely behaving animals in a natural environment is important for both neuroscience research and neuroprosthetic development. However, most of the research to-date still relies on rack-mount instrumentation with restrained cables. The requirements of recording high bandwidth neural signal from multi-channel, in multi-brain areas, via wireless miniature devices place a significant challenge on existing electronic technology and design techniques. The design optimization of a fully integrated neural recording front-end is thus highly desirable.

In the last two decades, a large number of neural recording front-end designs have been reported with improvements from many different aspects. The major innovations have come from novel circuit and system topologies [73–75], low-noise design techniques [76–79], large number of channel-count [76, 80–82], energy efficient designs [74, 76, 82], and low-power wireless interfaces including ISM band FSK [76, 83, 84], FM [85, 86], UWB [66, 81], and backscattering [82, 87, 88].

X. Liu, J. Van der Spiegel, *Brain-Machine Interface*,
https://doi.org/10.1007/978-3-319-67940-2_2

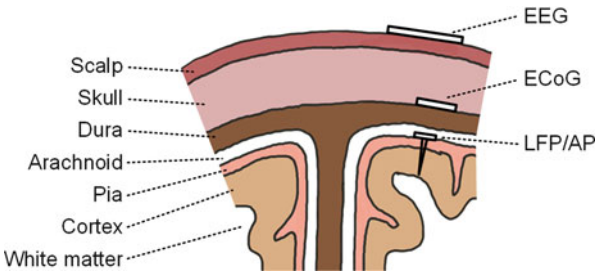


Fig. 2.1 The sources of the neural signals and their locations relative to the brain (not drawn to scale)

In addition, several systems have been used in freely behaving animal experiments [68, 69], and some of the prototype devices are fully integrated and are potentially implantable [80, 82, 89].

This chapter presents the analysis and design of neural recording front-end. Several novel circuit and system designs are proposed to improve the state-of-the-art. The chapter is organized as follows. Section 2.1 introduces the characteristics of the neural signal. The design specifications of neural recording circuit and systems are summarized. Section 2.2 reviews prior work, and analyzes the key trade-offs in the neural amplifier design, followed by a design of a general–purpose low-noise neural amplifier. Section 2.3 proposes a novel pre-whitening neural amplifier design, which exploits the frequency characteristics of the neural signal to relax the dynamic range and linearity requirement of the recording front-end. Section 2.4 presents the design of a 10-bit low-power SAR ADC for neural signal digitization. Section 2.5 presents the design of a complete neural signal acquisition front-end with compressive sensing for long-term neural signal recording in freely behaving animals.

2.1.1 Signal Characteristics

Neural signal can be recorded via invasive or non-invasive electrodes. Figure 2.1 shows the most common used neural signal sources and the corresponding electrode placements [90]. The electroencephalography (EEG) is the electrical brain activity recorded from the scalp, the electrocorticography (ECoG) is the electrical brain activity recorded beneath the skull, and local field potential (LFP) and action potential (AP) are electrical signals recorded within the parenchyma. The AP is the individual neuron activity, and the LFP is the activities from multiple nearby neurons.

Figure 2.2 illustrates the amplitude and frequency characteristics of different types of neural signals [91]. Main noise sources are also marked in this figure, including the thermal and flicker noise from the electrodes and electronic recording device, and the interferences from the environment.

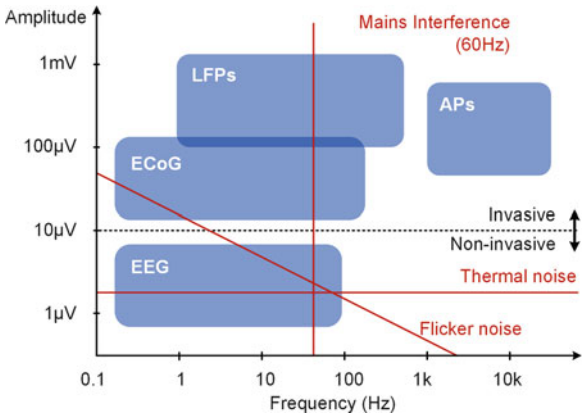


Fig. 2.2 The amplitude and frequency characteristics of different types of neural signals, in comparison with noise sources

Table 2.1 Summary of specifications for neural recording

Requirement	Range	Unit
Input dynamic range	10	mVpp
Electrode offset	±300	mV
Input impedance	>10	Mohm
Common mode rejection	60, at 50–60 Hz 30, at 100–120 Hz	dB
Gain accuracy	Error <10% and <±10	μV
Gain stability over 24 h	<3	%
Noise	50	μV
Crosstalk	<0.2 <5	mV %
Timing accuracy	<30 (over 24 h)	s
Temporal alignment	Error <20	ms

2.1.2 Design Specifications

The key specifications for a neural recording front-end include: input-referred noise, dynamic range, input impedance, linearity, common-mode rejection ratio (CMRR) and power-supply rejection ratio (PSRR), and so on. The minimum requirements are listed in Table 2.1, cited from International Electro-Technical Commission (IEC) medical electrical equipment standard 60601-2-47 [92]. It should be noticed that the requirement for a specific application will usually be higher than the general standard.

In addition to the requirements of the neural amplifier circuit, there are several key requirements of a successful chronic invasive neural recording system:

1. Longevity requirement: safe electrode interface, minimum tissue damage or infection;
2. Noise, bandwidth, and channel-count requirement for the target signal source;
3. Reliable data storage or wireless transmission;

4. Low power for minimum heat damage and long-term recording. In addition, the BMI research usually requires the front-end to be highly programmable, wireless compatible with commercial equipment and sensors, and also can be easily upgraded. All of these features together are required for a practical recording system for neuroscience research and BMI development. A balance between the requirements of each system block needs to be carefully considered.

Several figure-of-merits (FoM) are commonly used for evaluating and comparing different neural recording front-end designs. The most important FoM for noise and power performance is the noise efficiency factor (NEF). The NEF was first proposed by M. Steyaert et al. from the Katholieke Universiteit Leuven in 1987 [93], and was resurrected by R. Harrison et al. from the University of Utah in 2003 [73]. The NEF is defined as:

$$\mathrm{NEF} = \overline{V_{\mathrm{ni,rms}}} \sqrt{\frac{2I_{\mathrm{tot}}}{\pi \Phi_t \cdot 4kT \cdot \mathrm{BW}}} \tag{2.1}$$

where $\overline{V_{\mathrm{ni,rms}}}$ is the input-referred rms noise voltage of the amplifier, I_{tot} is the amplifier's total supply current. For a single bipolar transistor, the input-referred rms noise is:

$$\overline{V_{\mathrm{ni,rms}}} = \sqrt{\frac{4kT\Phi_t}{I_{\mathrm{tot}}} \cdot \frac{\pi \mathrm{BW}}{2}} \tag{2.2}$$

So, the NEF of a single bipolar transistor is 1 [73]. Paper [73, 93] predicted that all practical circuits must have a NEF greater than 1, however, later developed techniques overcame this limitation [44, 94].

It should be noticed that the NEF leaves the supply voltage out of the trade-off. As a result, two amplifier designs with different supply voltages but a same supply current and noise performance may have the same NEF. To mitigate this tissue, R. Muller et al. from the University of California, Berkeley proposed a FoM named power efficiency factor (PEF) in 2012 [95]. The PEF is defined as:

$$\begin{aligned} \mathrm{PEF} &= \mathrm{NEF}^2 V_{DD} \\ &= \frac{V_{\mathrm{ni,rms}}^2 \cdot P_{\mathrm{tot}}}{\pi \cdot kT/q \cdot 4kT \cdot \mathrm{BW}} \end{aligned} \tag{2.3}$$

The PEF gives a direct trade-off between power and noise, and two amplifiers with the same input rms noise and power consumption should have the same PEF. Reducing supply voltage significantly reduces the power consumption resulting a better PEF. However, reducing supply voltage usually comes at a cost of lowering the dynamic range. In order to compare the overall system's efficiency, D. Han et al.

from the Nanyang Technological University further proposed a FoM named system efficiency factor (SEF) in 2013 [96]. The SEF is defined as:

$$\mathrm{SEF} = \frac{\mathrm{PEF}}{\mathrm{DR}_{\mathrm{out}}} \tag{2.4}$$

where

$$\mathrm{DR}_{\mathrm{out}} = 10\log\frac{V_{\mathrm{amp,max}}^2}{2\cdot G_{\mathrm{AFE}}^2 V_{\mathrm{ni,rms}}^2} \tag{2.5}$$

where $V_{\mathrm{amp,max}}$ is the maximum voltage swing of the amplifier, and G_{AFE} is the voltage gain of the amplifier. SEF takes noise, power, and dynamic range performance into account, thus is more suitable for a system level performance comparison.

2.2 Design of a Low-Noise Neural Amplifier

2.2.1 Review of Prior Work

Numerous designs of neural amplifier have been reported in literature. The motivation of this section is not to give a comprehensive survey of prior work, but to analyze the key design trade-offs with featured examples. Review, tutorial, and comprehensive surveys for neural amplifier designs can be found in [70, 97–99].

2.2.1.1 System Topology

Many electrical engineers are familiar with the classical instrumentation amplifier using a 3-opamp topology. The 3-opamp instrumentation amplifier has a high input impedance, a good CMRR, but at a low power efficiency. Commonly used low-power CMOS neural amplifiers use capacitor and resistor elements to set the closed-loop gain. Typical block diagrams are shown in Fig. 2.3. Using capacitive gain elements, the design is inherently AC coupled [73]. Thus, the input common-mode range is from ground to VDD, limited by the ESD protection circuits. The only active component is the operational transconductance amplifier (OTA). The CMRR is mainly limited by the mismatch of the capacitors. The input impedance is limited by the size of the input capacitor and is usually frequency dependent. Using resistive gain elements, the design is inherently DC coupled [100]. The input common-mode range is less than VDD. The DC headroom is VDD/gain, which is much smaller than the capacitive counterpart. The input impedance is limited by the

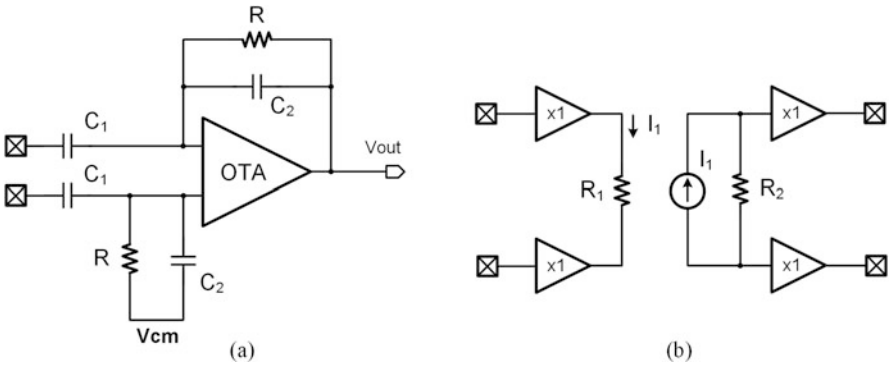


Fig. 2.3 The block diagram of the typical neural amplifiers with (**a**) capacitive gain element and (**b**) resistive gain element. The x1 symbol is for a unity gain buffer

Table 2.2 Comparison of neural amplifiers with capacitive and resistive gain elements

Gain element	Capacitive	Resistive
Gain	C_1/C_2	R_2/R_1
Noise PSD	$v^2_{\text{Thermal}}, v^2_{\text{Flicker}}$	$v^2_{\text{Thermal}}, v^2_{\text{Flicker}}, v^2_R$
Input impedance	$1/j\omega C_1$	$1/j\omega C_p$
DC headroom	VDD	VDD/gain
Input CM range	VDD	<VDD

parasitic capacitance, which is usually much higher than the capacitive counterpart. The CMRR is mainly limited by the mismatch of the analog buffers. Unlike the capacitive gain elements, the resistors also contribute to the overall noise. Several key features of the capacitive and resistive amplifiers are listed in Table 2.2 [101]. In summary, the topology using capacitive gain elements enjoys several inherent advantages over the resistive counterpart. Although a lot of techniques have been reported to address these problems [74, 100, 101], the capacitive gain element topologies are the mainstream designs for neural amplifiers.

In addition to passive gain elements, active feedback topologies have also been used to shape the frequency response [95, 102]. With an active feedback, the large input capacitor can be replaced by a small integrating capacitor, and a high input impedance can be achieved. However, the active feedback adds to the power consumption, and also contributes to the overall noise of the system.

Moreover, open-loop amplifiers have also been reported in literature [103, 104]. Compared with the closed-loop topologies, open-loop amplifiers can achieve a higher power efficiency, but usually suffer from a poor linearity. But since the neural signal has a small amplitude, the linear input range may be good enough in certain applications.

2.2.1.2 Low-Noise OTA

Operational transconductance amplifier (OTA) usually serves as the core of a low-noise neural amplifier. Commonly used OTA structures include: (a) current mirror OTA, (b) two-stage OTA with Miller compensation, (c) folded cascode OTA, and (d) telescopic OTA. Current mirror OTA usually has two stages with the dominant pole located at the second stage. No compensation capacitor is required to maintain stability. A detailed noise analysis is presented in [73]. However, the current mirror OTA has a limited gain, and there exists a trade-off between the noise and the phase margin. A gain boosting circuit can be used to enhance the gain of this amplifier, which is important in low voltage design in advanced CMOS technology. Two-stage OTA with a Miller compensation capacitor is also widely used in neural amplifiers. A detailed design analysis can be found in [105, 106]. The folded cascode OTA can achieve a high gain in a single stage, at the price of a higher power consumption. Paper [77] describes the strategy in choosing the parameters in a folded cascode OTA design for the optimal power-noise efficiency. Telescopic OTA can achieve the highest gain in a single-stage. But telescopic OTAs have very limited input range and voltage swing, which make the design very challenging especially in a low supply voltage.

In summary, all circuit topologies have pros and cons. Modified and improved versions have been widely reported. The voltage gain and input-referred noise of different OTA topologies are derived and summarized in [98, 99]. For thermal noise, increasing the transconductance of the input devices is usually critical. Thus, maximizing the transconductance for a given supply current is important for achieving an optimal power-noise efficiency. Besides, the supply current can be programmed to optimize the power efficiency in different noise conditions [44].

2.2.1.3 Other Noise Reducing Techniques

Many circuit techniques have been proposed in the literature to reduce the noise in the amplifier circuits. Commonly used low-noise techniques include chopping [78, 79, 107], auto-zeroing [108], digital assisted trimming [44, 109], analog and digital filtering, and so on.

For example, chopping is a very popular technique among neural amplifier designs, especially for EEG recordings. Figure 2.4 illustrates the concept of chopping. Before amplification, the input signal is modulated by a chopping frequency f_{chop}, which is much higher than the signal frequency. The modulated signal is then located to a frequency higher than the filter noise. After the amplification, the signal is converted back to the baseband frequency, at the same time, the flicker noise will be up-converted to the chopping frequency, which can be removed by a lowpass filter. Chopping technique reduces both flicker noise and DC offset, and the circuits after the chopping switches can achieve an excellent CMRR. However, it should be noticed that chopping also causes extra non-idealities, including offset, ripple, charge injection, clock feed-through, switch noise, and so on. Many techniques have

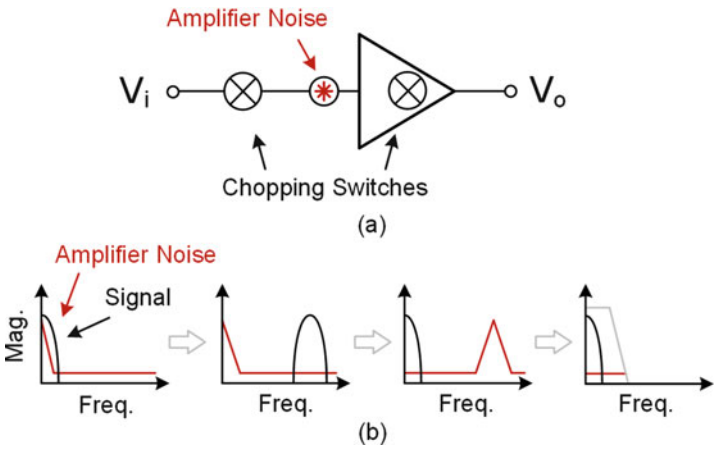


Fig. 2.4 (**a**) The block diagram of a chopping amplifier. (**b**) Illustration of the signal and noise spectrum before and after chopping

been proposed to suppress these problems, including: chopping within the feedback loop [78], chopping at the virtual ground [107], ripple reduction techniques [79], input impedance compensation [110], offset cancellation [82, 111], and so on.

2.2.2 Circuit Implementation

This section presents the analysis and design of a general–purpose neural recording front-end. Figure 2.5 shows the high-level block diagram of the proposed design. The building blocks of neural recording front-end consist of: a low-noise neural amplifier, a programmable gain amplifier (PGA), a multiplexer, an ADC, and a control module. This section mainly discusses the design of the neural amplifier.

The neural amplifier uses a fully differential, capacitor feedback topology. The input capacitors block the electrode offset and the half-cell potential from the electrode–tissue interface. The closed-loop differential gain is set to be 40 dB by $C_{\mathrm{IN}}/C_{\mathrm{FB}}$. A relative high gain is used to relieve the noise requirement of the following stages. A large MOS pseudo-resistor is used in the feedback loop. The highpass time constant is determined by $R_{\mathrm{pseudo}} \cdot C_{\mathrm{FB}}$. The circuit schematic of the pseudo-resistor is shown in Fig. 2.6. Compared with the MOS-bipolar resistor implemented in [73], this resistor has a larger linear range. Besides, setting the gate voltage to ground can short the feedback loop and force the input gate to mid-supply. This is a useful feature to implement a fast recovery from motion or stimulation artifact, which will be discussed in Chap. 5. A simulation of the MOS resistor in IBM 180 nm CMOS technology is shown in Fig. 2.7. The W/L of the MOS used in this simulation is 2 μm/2 μm. The simulated impedance is in the order of 100 GΩ.

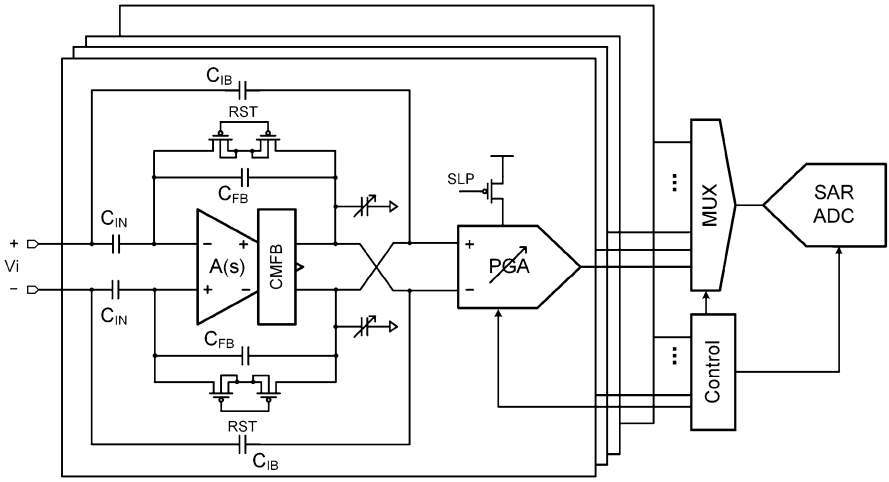
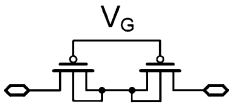


Fig. 2.5 The block diagram of the proposed low-noise neural recording front-end

Fig. 2.6 The circuit schematic of the MOS pseudo-resistor



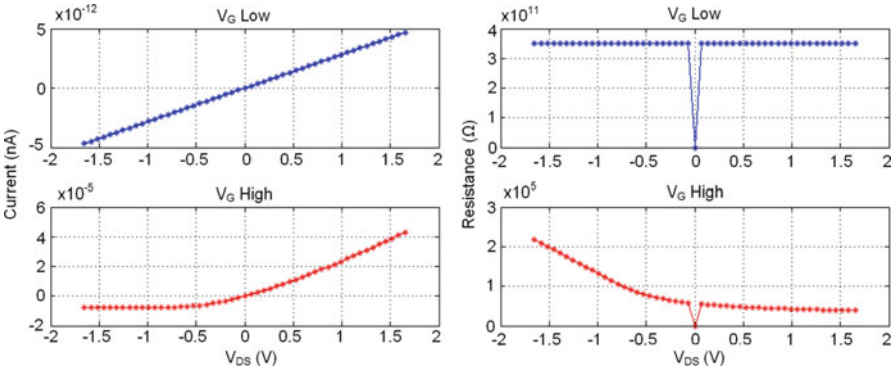


Fig. 2.7 The simulated resistance of the MOS pseudo-resistor. The *left column* shows the current versus the voltage applied, and *right column* shows the derived resistance

A very accurate resistance usually cannot be derived from the simulation. In practice, the highpass frequency is usually set to be much lower than the required signal frequency band to prevent the resistor's noise from rolling into the signal. Additional highpass or bandpass filter can be implemented in the following stage, if a better frequency shaping is necessary.

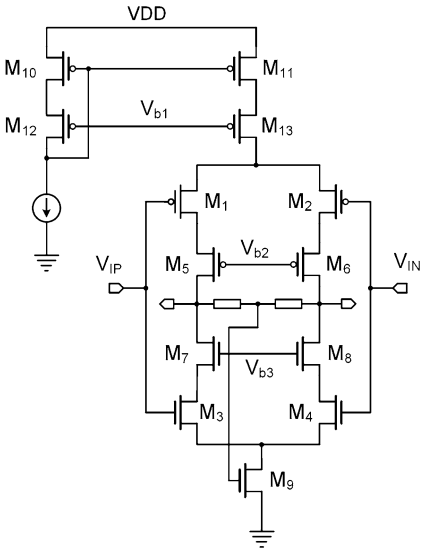


Fig. 2.8 The circuit schematic of the fully differential low-noise OTA with a complementary input stage

The circuit schematic of the designed OTA is shown in Fig. 2.8. The OTA has been designed to maximize the noise and power efficiency. A single-stage amplifier with a high gain is used to avoid the stability compensation in two-stage structures. The overall gain of the amplifier is given by:

$$A_v = (g_{m1} + g_{m3})(g_{m5} r_{o5} r_{o1} || g_{m7} r_{o7} r_{o3}) \tag{2.6}$$

where g_{mX} is the transconductance of the transistor M_X, and r_{oX} is the output resistance of the transistor M_X. The output thermal noise is:

$$\overline{i_{no}^2} = 4kT\gamma(g_{m1} + g_{m2} + g_{m3} + g_{m4})\Delta f \tag{2.7}$$

where $k = 1.38 \times 10^{-23}$ J/K is the Boltzmann constant. The input-referred thermal noise is:

$$\overline{v_{\text{ni}}^2} = \frac{4kT\gamma(2g_{m1} + 2g_{m4})}{(g_{m1} + g_{m3})^2}\Delta f \tag{2.8}$$

Taking the flicker noise into account, the total input-referred noise power of the OTA can be expressed as:

$$\overline{v_{\text{ni,tot}}^2} = \frac{1}{(g_{m1}+g_{m3})^2}\left[8kT\gamma(g_{m1}+g_{m3}) + 2\left(\frac{K_N g_{m3}}{C_{\text{ox},N} f W_N L_N} + \frac{K_P g_{m1}}{C_{\text{ox},P} f W_P L_P}\right)\right]\Delta f \tag{2.9}$$

The flicker noise can be reduced by increasing the size of the input transistors or using techniques like chopping. If only thermal noise is considered in the following design optimization, the input-referred noise voltage equals to

$$\overline{V_{\text{ni,rms}}} = \sqrt{\frac{8kT\gamma}{g_{m1}+g_{m3}}\frac{\pi}{2}\text{BW}} \tag{2.10}$$

The noise efficiency factor (NEF) [93] for this amplifier can be derived as:

$$\begin{aligned}\text{NEF} &= \overline{V_{\text{ni,rms}}}\sqrt{\frac{2I_{\text{tot}}}{\pi\Phi_t \cdot 4kT \cdot \text{BW}}} \\ &= \sqrt{\frac{8kT\gamma}{g_{m1}+g_{m3}}\frac{\pi}{2}\text{BW}\frac{2I_{\text{tot}}}{\pi\Phi_t \cdot 4kT \cdot \text{BW}}} \\ &= \sqrt{\frac{2\gamma I_{\text{tot}}}{(g_{m1}+g_{m3})\Phi_t}}\end{aligned} \tag{2.11}$$

Thus, a lower NEF (the lower the better) can be expected if a higher power efficiency (g_m/I_{tot}) is achieved.

In this work, complementary input devices are used. The overall transconductance can be approximately doubled without increasing the quiescent current. Besides, all input transistors are biased in the sub-threshold region to achieve a high power efficiency [112]. In the conventional operation (above threshold):

$$g_m = \sqrt{2\mu C_{\text{ox}}\frac{W}{L}I_D} \quad \text{and} \quad g_m \propto \sqrt{I_D} \tag{2.12}$$

In the sub-threshold operation:

$$g_m = \frac{\kappa I_D}{\Phi_t} \quad \text{and} \quad g_m \propto I_D \tag{2.13}$$

where Φ_t is the thermal voltage. Thus, sub-threshold operation gives a higher transconductance than conventional above threshold operation in a same drain current I_D. It should be noticed that sub-threshold operation has a limited bandwidth

due to the low biasing current and the relative large device parasitic capacitance. But neural signal has a low bandwidth in nature, so it is usually not a limiting factor in neural amplifier design. A simulation result shows that 98% of the noise is from the four input transistors, and flicker noise contributes more than thermal noise in the frequency range from 1 to 10 kHz.

Cascode transistors (M5–M8) are used to increase the voltage gain. However, this is usually at the cost of limiting the voltage headroom, and can be a challenge in a low-supply voltage using advanced CMOS technology. The simulation shows an open-loop gain of 90 dB is achieved in this OTA in a biasing current of 1 μA. The common mode feedback (CMFB) loop is merged in the main current path to avoid additional biasing current. Pseudo-resistors are used to get the common mode voltage without loading the amplifier. A drawback of this design is the threshold dependence of the common mode voltage.

A high input impedance reduces the signal attenuation. In a practical neural recording using a multi-channel electrode array, the recording electrode and the reference electrodes are usually not the same type of electrode, and may have a large impedance difference. Thus, even if the neural amplifier achieves a perfect common mode rejection, it cannot reject the conversion of common-mode signal to differential-mode due to the electrode mismatch. This problem can be relieved by increasing the input impedance of the neural amplifier. Positive current feedback [79] can be used to boost the input impedance by providing the driving current required at the input stage. A post-layout simulation of the input impedance boosting circuit is shown in Fig. 2.9.

A programmable capacitor array (C_L) is put at the output of the low-noise OTA. The bandwidth of the closed-loop amplifier is given by:

$$\mathrm{BW} = \frac{g_m}{C_L}\frac{C_2}{C_1} \tag{2.14}$$

The C_L can be programmed. The bandwidth of the OTA can also be tuned by changing the biasing current.

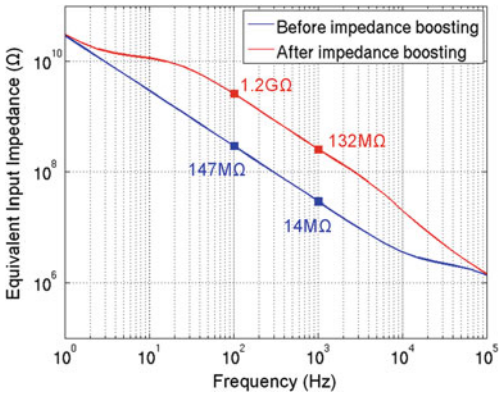


Fig. 2.9 A post-layout simulation of the effects of the input impedance boosting

The programmable gain amplifier (PGA) implemented in this work (2.5) is a classical 3-opamp amplifier. The gain is set by the resistors' ratio and can be chosen from 7, 10, and 19. Thus, the maximum gain of a recording channel is 1900. Additional analog buffers are added to the debugging points to drive the IO pads directly.

2.2.3 Measurement Results

The design has been fabricated in IBM 180 nm CMOS technology. The micrograph of the chip is shown in Fig. 2.10. The occupied silicon area of the full chip is $4.5 \times 1.5\,\text{mm}^2$, including IO pads. One recording channel has a dimension of $400\,\mu\text{m} \times 320\,\mu\text{m}$.

Bench testing was conducted to verify the function and performance of the fabricated chip. Figure 2.11 shows a measurement of the neural amplifier's output with a 1 kHz sinusoidal input signal. A resistor divider consists of a 2 kΩ and 1 Ω was applied at the output of the function generator to scale the signal amplitude, resulting a gain of 1/2001. The neural amplifier was configured to have the maximum gain of 1900. The measured gain was 1892.94, which corresponds to an absolute gain error of 0.37%.

The measured differential-mode and common-mode frequency responses of the low-noise amplifier are shown in Fig. 2.12. The closed-loop gain is set to be 60 dB. The highpass frequency corner is approximately 0.5 Hz. The measurement shows a CMRR above 110 dB.

The input-referred noise spectrum is shown in Fig. 2.13. An integration under this curve from 1 Hz to 7 kHz yields an rms noise voltage of 2.55 μV. This noise

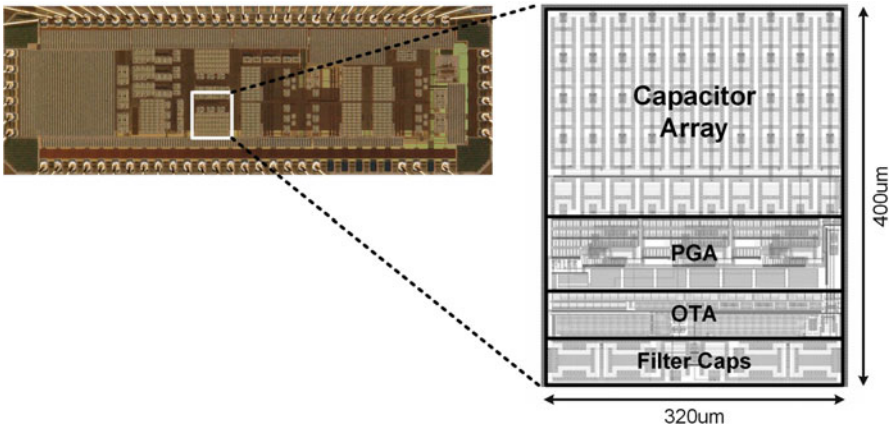


Fig. 2.10 The microphotograph and layout of one channel of the neural recording front-end. Major building blocks are highlighted in the layout

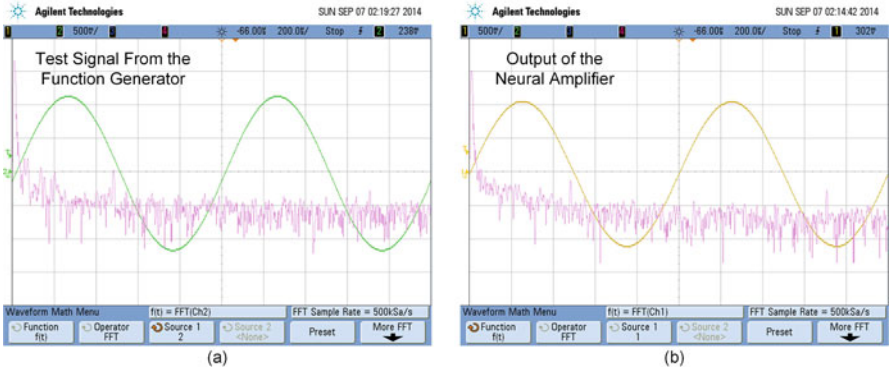


Fig. 2.11 The measured response of the neural amplifier with a 1 kHz sinusoidal input signal. The amplifier is configured with the maximum gain of 1900. The midband gain error is 0.37%

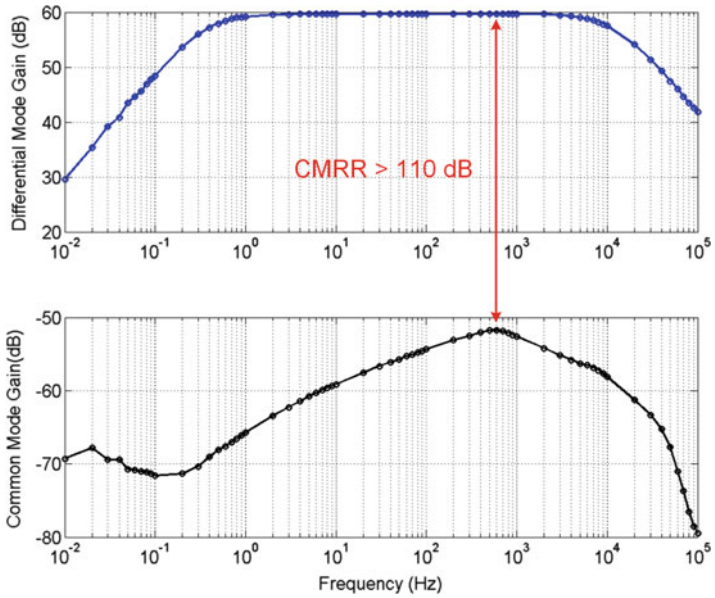


Fig. 2.12 The measured differential-mode and common-mode frequency responses of the low-noise neural amplifier

level was measured with a closed-loop gain of 60 dB, and the inputs were shorted using an internal switch. The noise density was calculated as:

$$\text{Noise Density} = \frac{V_{\text{rms}}}{\sqrt{\text{BW}\pi/2}} \tag{2.15}$$

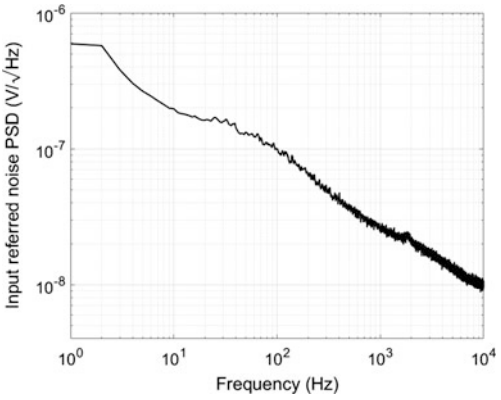


Fig. 2.13 The measured input-referred voltage noise spectrum. An integration under this curve from 1 Hz to 7 kHz yields an rms noise of 2.55 μV

Table 2.3 The neural front-end specifications summary

Parameter	Value
Process	180 nm CMOS
Supply voltage	3.3 V
LNA current	2 μA (biasing current not included)
Closed-loop gain	40 dB
Gain error	0.37%
Bandwidth	1–7 kHz
Integrated noise	2.55 μV
Noise density	24.3 nV/rtHz
NEF (Eq. (2.1))	1.68
PEF (Eq. (2.2))	9.38
Input range	4 mV
CMRR	>110 dB

The noise density in the 7 kHz bandwidth is 24.3 nV/rtHz. The calculated NEF is 1.68, and the PEF is 9.38.

The summarized measured specifications of the design are listed in Table 2.3.

In summary, this section presents the design of a general–purpose low-noise neural amplifier. The design achieves a low noise floor, an accurate gain, a good CMRR in a good power efficiency. The design was later used in in vivo neural signal acquisition.

Table 2.4 compares the measured performance of this work with prior published neural recording front-end design. This work achieves a comparable performance among the state-of-the-art designs.

Table 2.4 Comparison with prior works

Work	'03 [73]	'07 [78]	'07 [74]	'10 [107]	'13 [111]	'14 [44]	'17
Publication	JSSC	JSSC	JSSC	JSSC	JSSC	JSSC	This work
Technology	1.5 µm	0.8 µm	0.5 µm	180 nm	180 nm	180 nm	180 nm
Noise (µV)	2.2	0.95	2.26	1.3	0.91	5.23	2.55
BW (Hz)	0.025–7.2k	0.05–100	0.5–1k	100	100	7k	1–7k
Current (µA)	16	1	11.1	3.5	NA	0.97	2
Supply (V)	5	1.8–3.3	3	1	1	1.8	3.3
NEF	4.03	4.6	9.2	9.4	5.1	1.77	1.68
PEF[a]	81.2	38.1	253.9	88.4	26.2	5.6	9.3

[a]Not provided by the author, but calculated using Eq. (2.2)

2.3 A Pre-whitening Neural Amplifier

2.3.1 Introduction

The power spectrum of electrocorticography (ECoG) and local field potential (LFP) have a characteristic $(1/f)^n$ drop with frequency [31]. This phenomenon has been observed in multiple species including humans [113]. At frequencies around 1Hz, the signal amplitude can be as large as a few millivolts, and attenuates at $1/f^2$ until 80 Hz, then attenuates at $1/f^4$ [114]. At the same time, the noise power density of the CMOS front-end is usually inversely proportional to the frequency [115].

$$\overline{V_n^2} = \frac{K}{C_{\text{ox}}WL} \cdot \frac{1}{f} \tag{2.16}$$

where K is a process-dependent parameter on the order of 10^{-25} V^2F. This suggests that the SNR of the recording front-end improves as the frequency decreases, as illustrated in Fig. 2.14a. Intuitively, if a wideband recording front-end is designed to achieve the voltage swing requirement for the low-frequency signal, at the same time preserves the SNR for the high-frequency signal, it needs to be designed with an ultra-high dynamic range. An ultra-high dynamic range wideband low-noise amplifier and a high-resolution ADC design are challenging and will cost a high power consumption.

In this work, a pre-whitening amplifier is proposed to address this problem. The basic idea of the pre-whitening processing is illustrated in Fig. 2.14b. If we reduce the gain for the low-frequency content, a sufficient SNR may still be preserved for the recording purpose, and the dynamic range requirement of the system can be significantly relaxed. Since the frequency shaping processing is similar to a whitening filter, which turns the signal into a near white signal, the amplifier is named pre-whitening amplifier in this work. The simplest way to implement this pre-whitening amplifier is via a highpass filter. If the cut-off frequency of the highpass filter is known, the amplitude and phase of the original signal can be recovered in the post-recording processing.

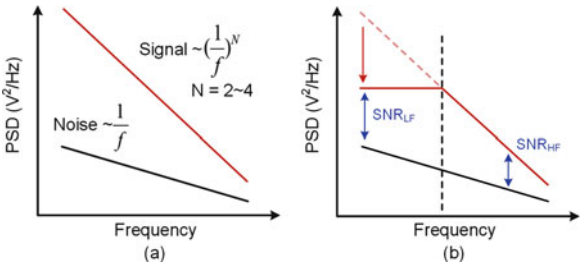


Fig. 2.14 Illustration of the pre-whitening filter. (**a**) The neural signal displays a $1/f^n$ power characteristic, while the recording front-end has a $1/f$ noise power characteristic. (**b**) The pre-whitening filter shapes the frequency response of the recording front-end to reduce the overall dynamic range requirement, while still preserves a sufficient SNR

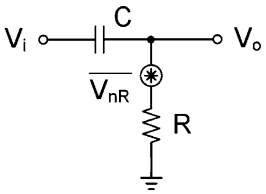


Fig. 2.15 A first order RC highpass filter with noise source

In summary, a pre-whitening neural recording front-end is proposed. In the pre-whitening amplifier, the frequency response of the neural amplifier is shaped according to the characteristic of the neural signal. The design significantly reduces the dynamic range requirement of the neural amplifier and the ADC resolution without sacrificing the signal quality. In the following sections, possible circuit implementations of the pre-whitening neural amplifier are analyzed. The key design trade-offs are described, and the simulation and experimental results of the proposed design are presented.

2.3.2 Analysis of Pre-whitening Neural Amplifier Design

The thermal noise power spectral density of a resistor is given here for convenience:

$$\overline{V_{nR}} = \sqrt{4kTR} \tag{2.17}$$

where $k = 1.38 \times 10^{-23}$ is the Boltzmann's constant, T is the absolute temperature in Kelvin. If a recording electrode has an impedance of 100 kΩ, it should have a noise density of 40.7 nV/$\sqrt{\text{Hz}}$. Assume the neural signal's frequency band of interest is from 1 Hz to 10 kHz, the electrode gives an integral thermal noise of 4.07 μV in this frequency range.

Consider the simplest case of a first order RC highpass filter. Figure 2.15 shows the circuit and the noise source.

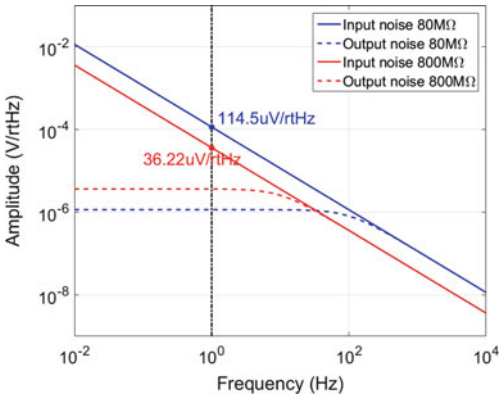


Fig. 2.16 Noise simulation of RC highpass filters with frequency corners at 10 and 100 Hz. The capacitor value is set to be 20 pF

The equivalent output noise of the RC highpass filter is given by:

$$\begin{aligned} \overline{V_{n,o}} &= \frac{1}{1+s\mathrm{RC}} \cdot \overline{V_{nR}} \\ &= \frac{\sqrt{4kTR}}{1+s\mathrm{RC}} \end{aligned} \tag{2.18}$$

While the input-referred noise of the RC highpass filter is given by:

$$\begin{aligned} \overline{V_{n,i}} &= \frac{1}{s\mathrm{RC}} \cdot \overline{V_{nR}} \\ &= \frac{\sqrt{4kTR}}{s\mathrm{RC}} \end{aligned} \tag{2.19}$$

The input-referred noise increases with a decreasing frequency. This is an important observation and provides some intuition for the following analysis. Figure 2.16 shows the simulation of a first order RC highpass filter. Both the output and the input-referred noise are plotted. The capacitor value is set to be 20 pF, and the resistor values are set to be 800 MΩ and 80 MΩ, and the cut-off frequency is 10 Hz and 100 Hz, respectively. The input-referred noise densities at the 1 Hz are marked in the figure. It should be noticed that using a larger capacitor value with the same cut-off frequency can achieve a lower noise density. However, large capacitors take a lot of silicon area, thus is not suitable for multiple channel recording front-end integration.

In summary, a simple RC filter is not suitable for implementing the proposed pre-whitening amplifier. In the following section, two methods of implementation are discussed: (1) pre-whitening after a wideband LNA, and (2) pre-whitening at the direct neural interface.

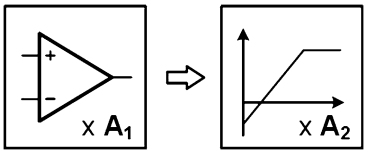


Fig. 2.17 The block diagram of a pre-whitening filter after a wideband low-noise amplifier

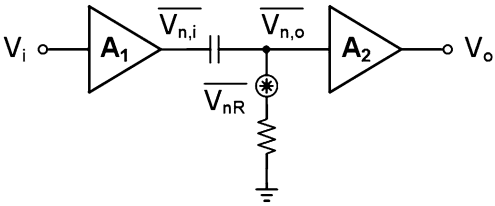


Fig. 2.18 The circuit schematic of a pre-whitening filter after a wideband low-noise amplifier. Noise source from the resistor is shown

2.3.2.1 Pre-whitening After a Wideband Low-Noise Amplifier

The block diagram of a pre-whitening filter after a wideband low-noise amplifier is shown in Fig. 2.17. Since the filtering is implemented after the wideband amplifier, the input-referred noise from the filter will be attenuated by the gain of the wideband amplifier. Again, assume using a simple RC filter, Fig. 2.18 shows the circuit diagram and the noise source.

The input-referred noise of the recording front-end from the filter is then given by:

$$\begin{aligned} \overline{V_{i,\text{rms}}} &= \frac{1}{A_1}\sqrt{\int_{f_L}^{f_H} \overline{V_{n,i}^2} \cdot df} \\ &= \frac{1}{A_1 \pi C}\sqrt{\frac{kT}{R}\left(\frac{1}{f_L} - \frac{1}{f_H}\right)} \end{aligned} \tag{2.20}$$

Assume the frequency band of interest is from 1 Hz to 10 kHz, and the first stage wideband neural amplifier has a gain A_1 of 100. If we set the highpass frequency of the second stage to be 10 Hz, the integral noise is 0.11 μV, and if we set the highpass frequency to be 100 Hz, the integral noise is 1.14 μV. In both cases, the integral noise is lower than the thermal noise of an electrode with an impedance of 100 kΩ (Sect. 2.3.1).

Several active highpass filters can achieve a lower input-referred noise than the simple RC filter. Consider the circuit with a capacitive feedback as shown in Fig. 2.19. The signal's transfer function can be expressed as:

$$\begin{aligned} H_{\text{sig}}(s) &= \frac{s\text{RC}_1}{s\text{RC}_2 + 1} \\ &= \frac{C_1}{C_2} \cdot \frac{s}{s + \frac{1}{\text{RC}_2}} \end{aligned} \tag{2.21}$$

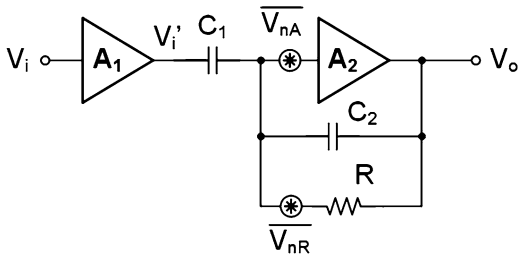


Fig. 2.19 An implementation of an active highpass filter. Noise sources from the resistor and the second stage amplifier are shown

The midband gain of the amplifier A_{CL} is $\frac{C_1}{C_2}$, and the highpass frequency is set by $\frac{1}{\mathrm{RC}_2}$. The amplifier A_2's noise transfer function can be expressed by:

$$\begin{aligned} H_{nA}(s) &= 1 + \frac{s\mathrm{RC}_1}{s\mathrm{RC}_2 + 1} \\ &= 1 + \frac{C_1}{C_2} \cdot \frac{s}{s + \frac{1}{\mathrm{RC}_2}} \end{aligned} \tag{2.22}$$

Since the amplifier's noise transfer function and the signal's transfer function have the same highpass frequency $\frac{1}{\mathrm{RC}_2}$, the amplifier's noise is shaped in the same way as the signal. Thus, the filter won't cause frequency dependent SNR degradation.

Let's look at the resistor's noise transfer function:

$$\frac{V_o - V_{nR}}{R} = sC_2V_o = 0 \tag{2.23}$$

$$H_{nR}(s) = \frac{1}{s\mathrm{RC}_1} = \frac{1}{A_{\mathrm{CL}}} \frac{1}{s\mathrm{RC}_2} \tag{2.24}$$

where $\frac{1}{\mathrm{RC}_2}$ is the signal's highpass frequency, and A_{CL} is the closed-loop gain of the second stage. So compared with the implementation in Fig. 2.18, the noise is further suppressed by the gain of the second stage. The overall input-referred noise density from the resistor is given by:

$$\overline{V_{nR,i}} = \frac{1}{A_1 A_{\mathrm{CL}}} \frac{\sqrt{4kTR}}{s\mathrm{RC}_2} \tag{2.25}$$

In a practical design, the sum of $\overline{V_{nR,i}}$ and the input-referred noise of the first stage should be lower than the noise and the SNR requirement of the recording system. Again, assume the cut-off frequency of the pre-whitening amplifier is 10 Hz, the capacitor C_2 is 2 pF, the closed-loop gain of the first and second stage is 100 and 40, respectively. The $\overline{V_{nR,i}}$ at 1 Hz is 28.6 nV$\sqrt{\mathrm{Hz}}$, which is lower than the thermal noise density of a 100 $k\Omega$ electrode (Sect. 2.3.2). If the cut-off frequency of the

pre-whitening amplifier is 100 Hz, the noise density $\overline{V_{nR,i}}$ at 1 Hz is 90.5 nV/$\sqrt{\text{Hz}}$, which is still lower than most low-noise neural amplifier designs at 1 Hz, and is sufficient for the SNR requirements in most intracortical neural recordings.

2.3.2.2 Low-Noise Neural Amplifier with Integrated Pre-whitening Filter

This section discusses the possible methods of integrating the pre-whitening filter into the first stage low-noise amplifier. It is more challenging to design the pre-whitening filter at the first stage because of the noise increase with decreasing frequency due to the filter's response. But there are also some advantages. The electrode interface usually has an offset caused by half-cell potential up to several hundred millivolts, as reviewed in Sect. 2.1. The recording amplifier will need to reject this large offset, typically accomplished by using a highpass filter with a cut-off frequency below 1 Hz. However, it is difficult to implement such a large time-constant on-chip. One solution is to use MOS pseudo-resistor, as described in Sect. 2.2.2. But the pseudo-resistors have reliability problem for the use in implanted medical devices, and they are susceptible to electromagnetic interface and degradation over time [31]. If the pre-whitening filter can be integrated into the first stage amplifier, the sub-Hertz filter can be avoided.

Consider the capacitor-coupled neural amplifier in Fig. 2.20. The signal's transfer function is:

$$H_{\text{sig}}(s) = \frac{C_1}{C_2} \cdot \frac{s}{s + 1/RC_2} \tag{2.26}$$

Thus, the highpass corner frequency is determined by $1/\text{RC}_2$. The transfer function of the amplifier's noise is:

$$H_{nA}(s) = 1 + \frac{C_1}{C_2} \cdot \frac{s}{s + 1/\text{RC}_2} \tag{2.27}$$

And the resistor noise's transfer function is:

$$H_{nR}(s) = \frac{1}{s\text{RC}_1} \tag{2.28}$$

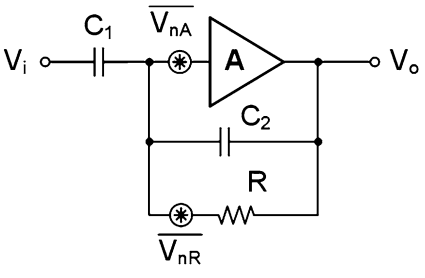


Fig. 2.20 A typical capacitor-coupled neural amplifier. Noise sources are marked in the figure

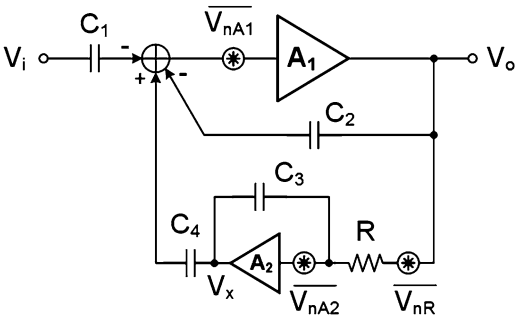


Fig. 2.21 A capacitor-coupled instrumentation amplifier with a DC servo loop

The input-referred noise density from the resistor is:

$$\overline{V_{nR,i}} = \frac{1}{A_{\mathrm{CL}}} \frac{\sqrt{4kTR}}{s\mathrm{RC}_2} \tag{2.29}$$

Compared with Eq. (2.25), the only difference is that this input-referred noise is no longer attenuated by preamplifier. If A_{CL} is designed to be the product of the gain of the two stages in previous section, it can achieve the same noise performance. But it is difficult in practical designs.

There are other circuit topologies to implement a highpass frequency response. One method is to use a DC servo loop. A typical example is shown in Fig. 2.21.

$$sC_1V_i(s) + sC_2V_o(s) - sC_4V_x(s) = 0 \tag{2.30}$$

$$Vx(s) = -\frac{1}{s\mathrm{RC}_3}V_o(s) \tag{2.31}$$

The signal's transfer function can be expressed as:

$$\begin{aligned} H_{\mathrm{sig}}(s) &= \frac{s\mathrm{RC}_1C_3}{s\mathrm{RC}_2C_3 + C_4} \\ &= \frac{C_1}{C_2} \cdot \frac{s}{s + C_4/\mathrm{RC}_2C_3} \end{aligned} \tag{2.32}$$

The mid-band gain of this circuit is $\dfrac{C_1}{C_2}$. Compared with Eq. (2.26), the high-pass frequency corner is $\dfrac{C_4}{C_2}\dfrac{1}{\mathrm{RC}_3}$, where $\dfrac{1}{\mathrm{RC}_3}$ is the frequency corner of the integrator in the feedback loop.

The noise transfer function of the amplifier A_1 is:

$$C_1V_{nA1} = C_2(V_o - V_{nA1}) + C_4(V_x - V_{nA1}) \tag{2.33}$$

$$H_{nA1}(s) = \frac{C_1 + C_2 + C_4}{C_2} \cdot \frac{1}{1 + C_4/sRC_2C_3} \tag{2.34}$$

The noise transfer function of the amplifier A_2 is:

$$\frac{V_o - V_{na2}}{R} = sC_3(V_{na2}) = sC_4Vx = sC_2V_o \tag{2.35}$$

$$H_{nA2}(s) = \frac{s\mathrm{RC}_3 + 1}{s\mathrm{RC}_3(C_2/C_4) + 1} \tag{2.36}$$

The noise transfer function of the resistor is:

$$\frac{V_o - V_{nR}}{R} = sC_3V_x \tag{2.37}$$

$$sC_4V_x = sC_2V_o \tag{2.38}$$

$$H_{nR}(s) = \frac{1}{1 + s\mathrm{RC}_3\dfrac{C_2}{C_4}} \tag{2.39}$$

The input-referred noise density from the amplifier A_1 is:

$$\overline{V_{nA1,i}} = \frac{C_1 + C_2 + C_4}{C_1}V_{nA1} \tag{2.40}$$

The input-referred noise density from the amplifier A_2 is:

$$\overline{V_{nA2,i}} = \frac{C_4}{C_1}\left(1 + \frac{1}{s\mathrm{RC}_3}\right)V_{nA2} \tag{2.41}$$

The input-referred noise density from the resistor is:

$$\overline{V_{nR,i}} = \frac{C_4}{s\mathrm{RC}_1C_3}\sqrt{4kTR} \tag{2.42}$$

Compared with Eq. (2.29), the noise contribution from the resistor also depends on the ratio of C_4/C_3. However, reducing the ratio of C_4/C_3 decreases the input voltage headroom.

The resistor can be further replaced by a switched capacitor circuit. A simplified circuit schematic is shown in Fig. 2.22. In the switched capacitor circuit, the time-constant can be better controlled by the ratio of the capacitors and the switching

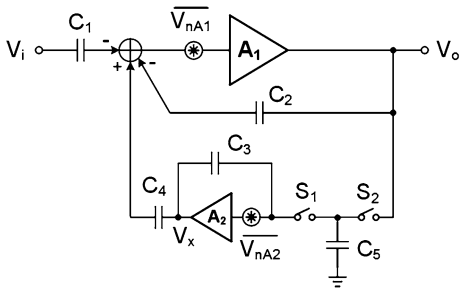


Fig. 2.22 A capacitor-coupled instrumentation amplifier with a DC servo loop implemented by switched capacitor circuits

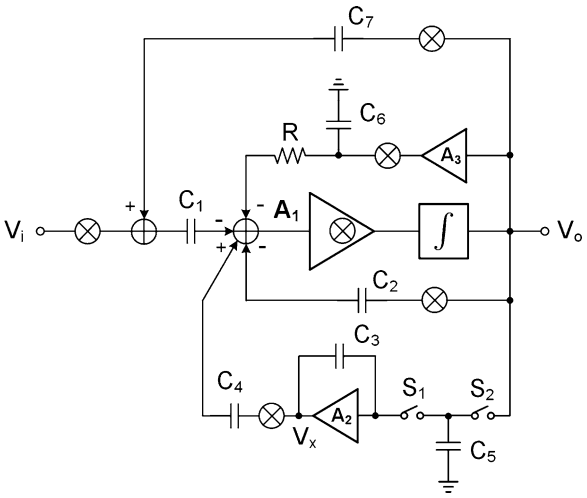


Fig. 2.23 A capacitor-coupled chopping amplifier with a DC servo loop and an input impedance boosting loop

frequency. However, in order to achieve the required large time-constant, a large capacitor tank is to be implemented. Several techniques have been proposed to reduce the required capacitor size [78, 116].

If an ultra low-noise is required for the low-frequency signal component, chopping technique can be combined in the pre-whitening amplifier design. An example of a chopping pre-whitening amplifier is shown in Fig. 2.23. With chopping, the flicker noise can be removed, and the amplifier can guarantee a good SNR for the signal even with the lower gain at the low frequency. However, there are also many trade-offs involved with the chopping amplifier design [79, 107, 110]. The effects it takes may counteract the benefits from the pre-whitening.

2.3.3 Circuit Implementation

A pre-whitening amplifier using the architecture presented in Sect. 2.3.2 is designed to demonstrate the idea. The circuit schematic of the designed pre-whitening amplifier is shown in Fig. 2.24. A single-ended architecture is used in this work. A conventional low-noise current mirror OTA is used in both stages [73]. A T-connected pseudo-resistor (TPR) proposed in [44] is used as the feedback resistor in the first stage. If the equivalent resistance of the transistor X is R_X, the total equivalent resistance of the TPR is $R_1+R_2+R_1 \cdot R_2/R_3$. The pseudo-resistor in the second stage is the same as the one used in Sect. 2.2. The gate voltage can be used to tune the resistance R_B over a large range, which is used to tune the cut-off frequency of the pre-whitening filter. The first stage has a closed-loop gain of 100, and the second stage has a closed-loop gain of 40, which are the same as the assumptions in the previous analysis.

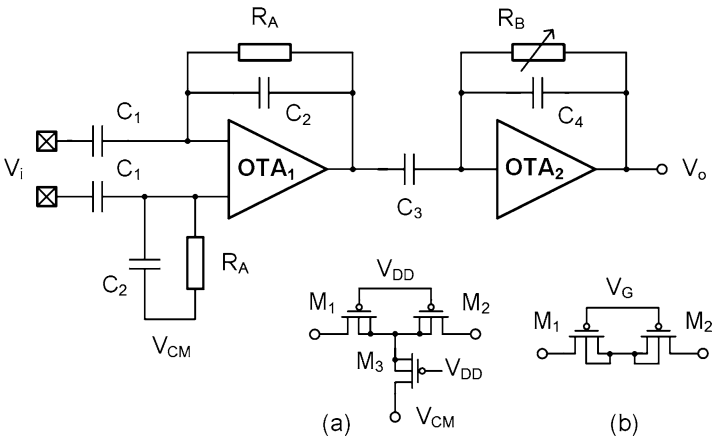


Fig. 2.24 The circuit schematic of the designed pre-whitening amplifier. The pseudo-resistors R_A and R_B used in the first and second stage are shown in subplot (**a**) and (**b**), respectively

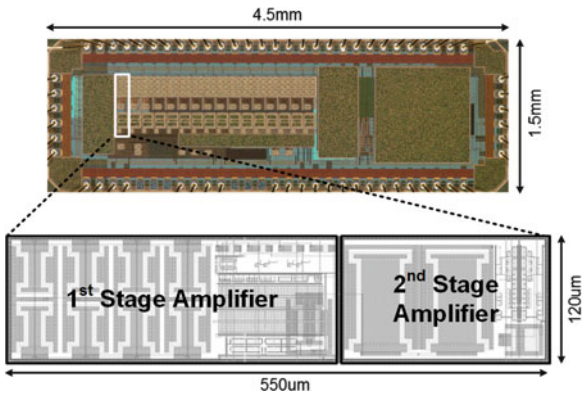


Fig. 2.25 The microphotography and layout of one channel of the pre-whitening amplifier

2.3.4 Measurement Results

The design has been fabricated in IBM 180 nm CMOS technology. The micrograph of the chip and the layout of one recording channel are shown in Fig. 2.25. The full chip occupies a silicon area of $4.5 \times 1.5\ \text{mm}^2$, including IO pads. One recording channel has a dimension of $550\ \mu\text{m} \times 120\ \mu\text{m}$.

Bench testing was conducted to evaluate the performance of the fabricated chip. Figure 2.26 shows both the amplitude and phase frequency response of the pre-whitening amplifier. The simulated frequency response is also plotted in dashed lines for comparison. With the information of the frequency response, the original signal can be recovered from the pre-whitened recording.

Synthetic neural signal was generated using an arbitrary function generator 33521A from Agilent to the test the pre-whitening amplifier. A 1-min neural signal

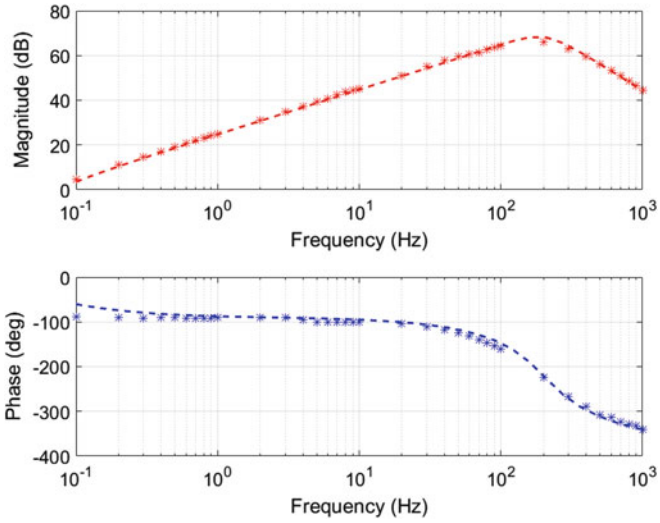


Fig. 2.26 The measured frequency response of the pre-whitening amplifier in comparison with the simulation result

containing local field potentials was used for testing. The signal was recorded using RZ2 workstation from Tucker-Davis Technologies. The signal was sampled at 24.4 kSps in a resolution of 24-bit. A resistor divider consisting of 2 kΩ and 1 Ω was applied at the output of the function generator, gives a gain of 1/2001. The neural amplifier was configured to have a maximum gain of 4000.

The designed amplifier can be configured to do both conventional wideband recording and frequency shaping pre-whitening recording. The power spectral density (PSD) was calculated for 24 channels of the LFP recordings. Figure 2.27 shows a comparison of the PSD of the conventional wideband recording and the pre-whitening recording. The result clearly shows that the spectrum of the pre-whitening recording was flattened in the low-frequency range, which saves the voltage headroom by more than an order of magnitude. The reduction in the dynamic range relaxes the requirement of the linear range of the low-noise amplifier and the ADC design.

The reconstruction of the signal was performed in Matlab. Figure 2.28 shows a comparison of a 10-s segment of the conventional recording, the pre-whitening recording, and the reconstruction from the pre-whitening recording. Pearson correlation coefficient is used here to evaluate the accuracy of the reconstruction [117]. The correlation coefficient is defined as:

$$\rho(x,y) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\overline{x_i-\mu_x}}{\sigma_x}\right)\left(\frac{\overline{y_i-\mu_y}}{\sigma_y}\right) \tag{2.43}$$

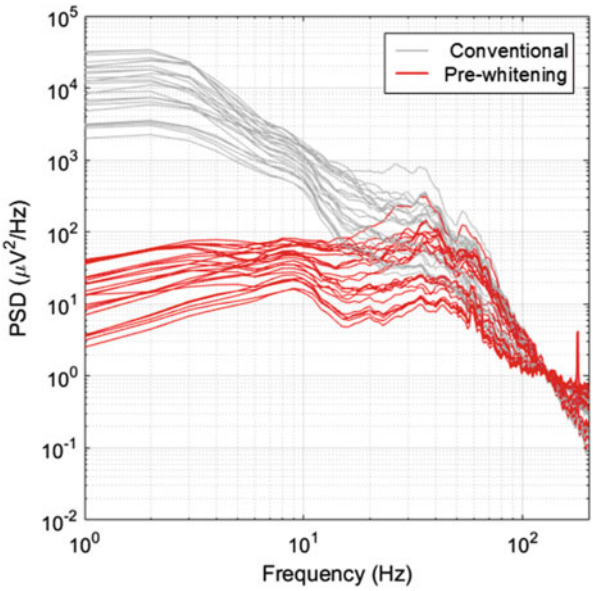


Fig. 2.27 Comparison of the PSD of the pre-whitening amplifier and the original signal

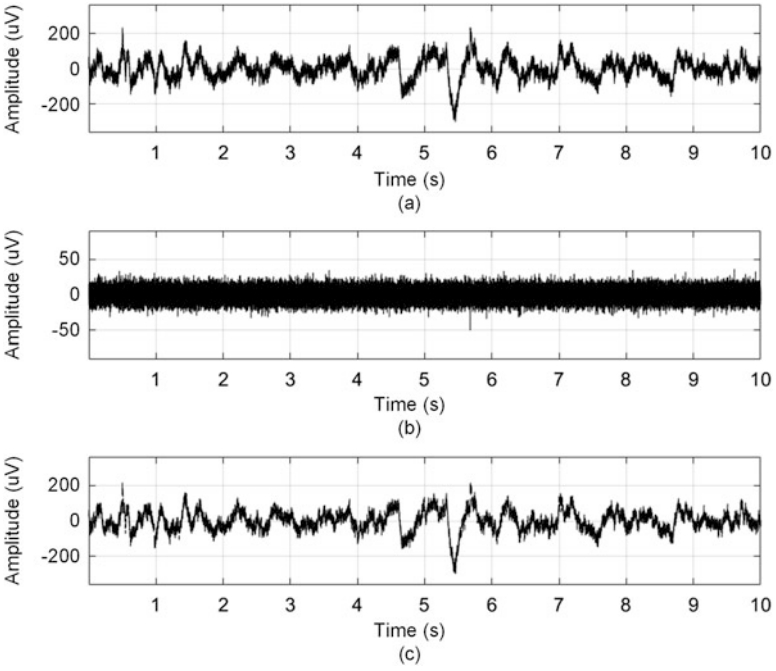


Fig. 2.28 Comparison of (**a**) the wideband signal, (**b**) the measured output of the pre-whitening amplifier, and (**c**) the reconstructed signal from the pre-whitening amplifier's recording

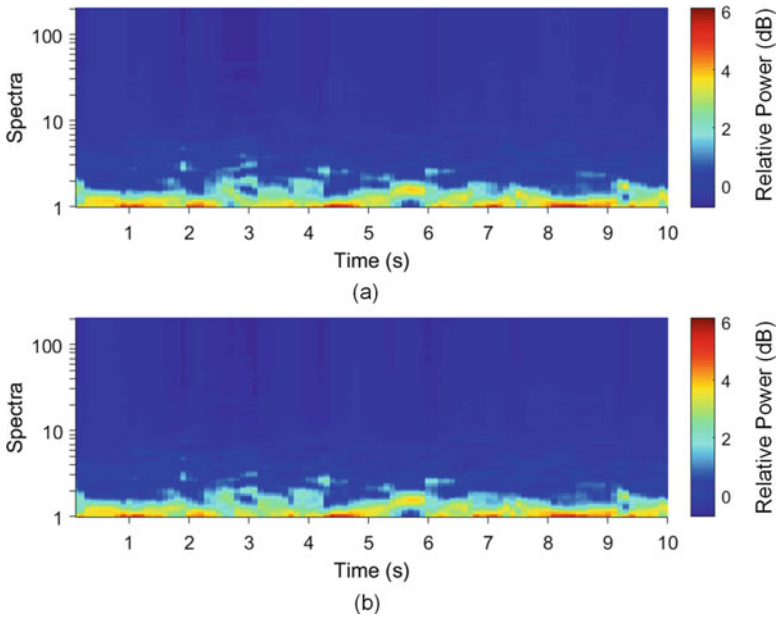


Fig. 2.29 Comparison of the spectrum of (**a**) the original signal and (**b**) the reconstructed signal from the pre-whitening amplifier's recording

where μ_x and σ_x are the mean and standard deviation of the signal x, and μ_y and σ_y are the mean and standard deviation of the signal y. The result shows a correlation coefficient of 97.6%, which indicates a faithful recovering of both phase and amplitude. A high-order zero-phase digital filter of 1–200 Hz was applied before the comparison. Even in this case, both phase and amplitude will need to be recovered at the same time to completely reconstruct the original signal.

Power spectral density estimation was calculated by periodgram for both the original signal and the reconstructed signal from the pre-whitening recording, as shown in Fig. 2.29. The result clearly shows that the pre-whitening processing can provide a faithful reconstruction of the spectrum content.

In summary, the pre-whitening amplifier design takes advantage of the characteristics of the neural signal. Since the power density of the neural signal including ECoG and LFP drops faster with frequency than the filter noise of the CMOS recording front-end, there is an opportunity to design a recording front-end with less gain at low frequency while preserving a sufficient SNR for the wideband signal. The design significantly reduces the dynamic range and linearity requirement of the low-noise amplifier and the ADC. The circuit implementation of the pre-whitening front-end is analyzed in this section with a detailed noise analysis. A prototype was designed and fabricated in CMOS technology. Experimental results are presented in comparison with simulation and theoretical computation. The proposed pre-whitening amplifier provides an opportunity to improve the

performance of neural recording-end without a power penalty, and it can be advantageous to integrate it into a high channel-count neural recording front-end system.

2.4 Design of a Low-Power Analog-to-Digital Converter

2.4.1 Introduction

A low-power analog-to-digital converter (ADC) is an essential component in a neural interface system. In a typical bidirectional neural interface system, ADCs can be used to digitize neural signal, sensory signal, extracted neural features, and stimulation compliance voltage. Among all ADC topologies, successive approximation register (SAR) ADCs have advantages in accuracy and power efficiency at a moderate sampling rate. Firstly, a SAR ADC does not require a high gain and high bandwidth opamp for high accuracy and linearity. Secondly, SAR logic mainly consists of digital circuits, so the speed and power scales down with deep submicron CMOS technologies. Thirdly, if a capacitive DAC is used, no static power is consumed, thus the power scales with the sampling rate. Comprehensive reviews and tutorials of SAR ADC design can be found in [118–120].

Recently, a lot of techniques for power-efficient SAR ADC designs have been reported. Among these techniques are split capacitor array [121, 122], monotonic capacitor switching [123], partial floating capacitor switching [124], step-charging design [125], reference free design [126], asynchronous timing [127], and so on. In addition to techniques for general–purpose SAR ADC designs, several techniques have been reported to optimize the design particularly for neural or sensory signal. Among them are:

- Adaptive resolution or dynamic range: including programming the number of bits [128], or adding additional programmable gain amplifier before the ADC [129]
- Data dependent or data-driven sampling: for example, combine action potential detection and digitization together [130]. Besides, the sampling rate can also be adapted to the activity using a continuous time level-crossing sampling [131].
- Delta difference sampling: since the neural signal has both slow and fast oscillations over time, normal sampling during a slow activity period is not energy efficient. So digitize only the difference [44, 132], using a bypass window [133], or using LSB-first approach [134] can achieve a better power efficiency.

In this section, the design of a voltage-mode 10-bit SAR ADC is presented. Specifications are analyzed, circuit design details are described, and measurement results are presented.

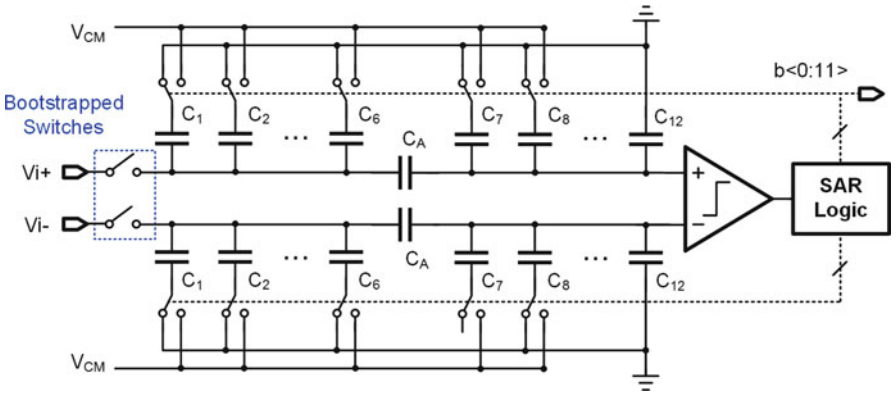


Fig. 2.30 The circuit diagram of the 10-bit voltage-mode SAR ADC

2.4.2 Circuit Implementation

The circuit diagram of the 10-bit voltage-mode SAR ADC is shown in Fig. 2.30. The major building blocks are: (1) comparator, (2) SAR logic, (3) DAC, and (4) sample and hold switch. Fourteen clock cycles are used to complete one conversion, allowing four clock cycles for sampling.

A commonly used capacitive DAC is employed in this SAR ADC. Since the required capacitor size in a conventional binary capacitor array can be very small without compromising the ENOB, custom designed capacitors are often used to achieve a minimum total input capacitance with an ultra low-power consumption [123, 135]. However, these designs usually require a custom characterization for a specific fabrication process. In this work, a split capacitor array is adopted to reduce the total capacitance, lowering the power consumption and area. The capacitors are realized as a standard metal-insulator-metal (MIM) structure available in the standard PDK.

A monotonic switching procedure is applied to minimize the power consumption from unnecessarily charging and discharging of the capacitor array [123]. In the monotonic switching procedure, the first comparison is performed without switching, and the total capacitance is half of the conventional capacitive SAR ADC's DAC array [123].

Figure 2.31 shows the circuit schematic of the SAR timing generation module. A global reset signal is used to synchronize the start of the conversion, and the control logic generation is cyclic. clk is the input clock, clks is the signal for the sampling switch, clkc is the clock for the comparator, and clk[x] is for the bit[x] of the DAC.

The sample and Hold (S/H) circuit is critical in achieving good SFDR for an ADC design. The bootstrapped switch is commonly used since it provides a constant small on-resistance [136]. The circuit schematic of the bootstrapped switch implemented

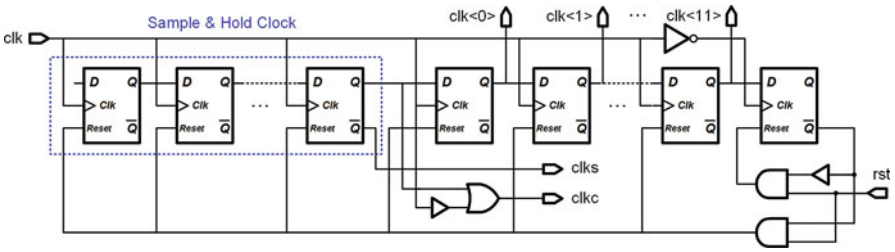
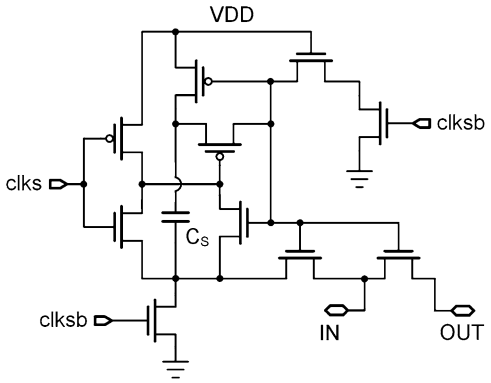


Fig. 2.31 The circuit schematic of the SAR timing generation module

Fig. 2.32 The circuit schematic of the bootstrapped switch



in this work is shown in Fig. 2.32. The gate to source voltage of the switch transistor is fixed at the supply voltage by the capacitor C_S.

Figure 2.33 shows the circuit schematic of the comparator. The comparator consists of a pre-amplifier and a dynamic latch. Since the input voltage has a range from the ground to V_{cm}, the comparator uses a PMOS input stage. The current source NMOS are used in parallel with the diode-connected NMOS for increasing the gain [115]. The pre-amplifier provides moderate gain to reduce the input referred mismatch from the latch. The latch consumes no static current. When clkc (as shown in Fig. 2.31) is high, the outputs are reset to high. When clkc goes to low, the regeneration latch forces one output to high and the other to low. The SAR logic only takes the V_{OP} and generates an inverted signal V'_{ON} to avoid the metastability problem.

A Class-AB output stage has been designed to drive the sample-hold circuits of the following ADC stage. To digitize a single-ended signal, a single-to-differential converter (S2D) can be integrated. An example of the S2D circuit is shown in Fig. 2.34. The resistor values are designed to be R_1=R_3=R_4, and the voltage gain is $2(1+R_2/R_1)$. R_2, which can be programmed by a shift register.

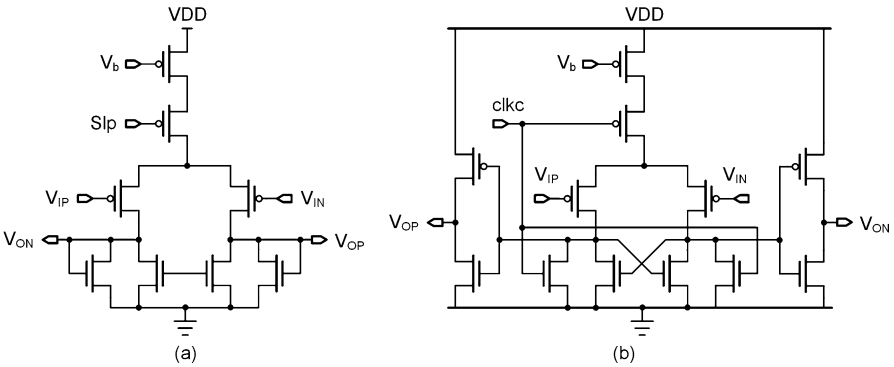


Fig. 2.33 The circuit schematic of the comparator. (**a**) A pre-amplifier, and (**b**) a dynamic latch

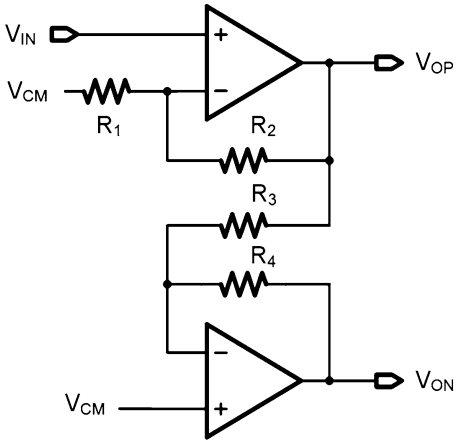


Fig. 2.34 The circuit schematic of a single-to-differential converter

2.4.3 Measurement Results

The ADC has been fabricated in IBM 180nm CMOS technology. The layout of the 10-bit SAR ADC is shown in Fig. 2.35 with major building blocks highlighted. The total occupied silicon area is $220\,\mu\text{m} \times 190\,\mu\text{m}$. The measurement results of the design are presented below.

The differential nonlinearity (DNL) and integral nonlinearity (INL) of the ADC were measured using slow ramps. The result is shown in Fig. 2.36. The peak DNL and INL are −0.49/+0.56 LSB and −0.82/+0.77 LSB, respectively.

The SAR ADC's dynamic performance was measured with a low-frequency input tone and a near Nyquist frequency input tone. The output spectrums are shown in Figs. 2.37 and 2.38, respectively.

The spurious-free dynamic range (SFDR) achieved in these tests was 76.54 dB and 71.6 dB, respectively. The signal-to-noise and distortion ratio (SNDR) was measured to be 56 dB and 54.6 dB, respectively. The effective number of bit (ENOB) is defined as:

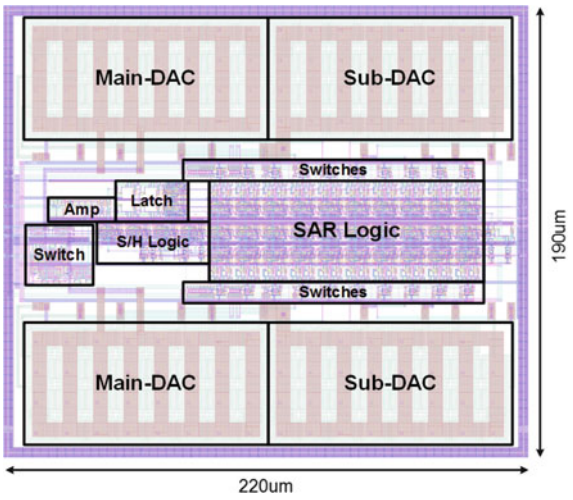


Fig. 2.35 The layout of the 10-bit SAR ADC with major building blocks highlighted

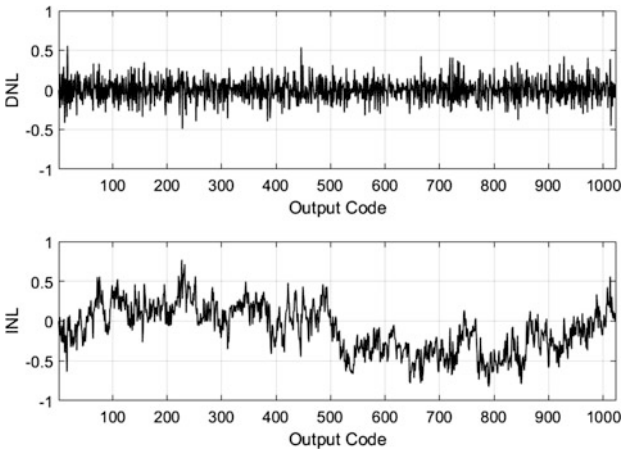


Fig. 2.36 The measured DNL and INL of the 10-bit SAR ADC. The worst DNL is −0.49/+0.56 LSB, and the worst INL is −0.82/+0.77 LSB

$$\text{ENOB} = \frac{\text{SNDR} - 1.76}{6.02} \tag{2.44}$$

The ENOB of the designed ADC was measured to be 9.01 and 8.77, respectively. The figure-of-merit (FoM) is calculated using:

$$\text{FoM} = \frac{\text{Power}}{2^{\text{ENOB}} \times f_s} \tag{2.45}$$

The FoM of the ADC is 98 fJ/conv-step at 1 MSps with a supply of 1.8 V. The measured specifications of the ADC were summarized in Table 2.5.

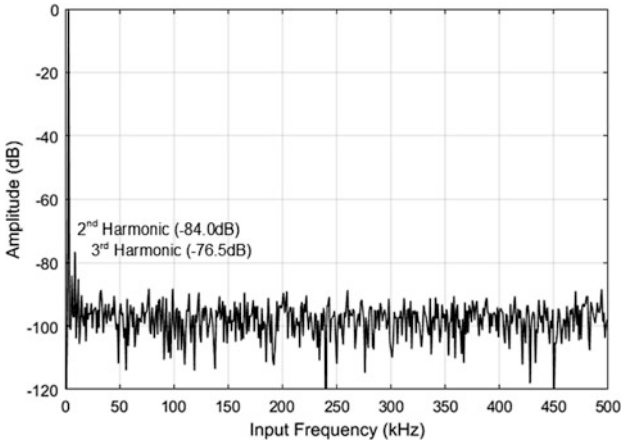


Fig. 2.37 The measured FFT spectrum at 1 MS/s with an input tone of 3 kHz. The SFDR is 76.54 dB and the SNDR is 56 dB. The ENOB at 3 kHz is 9.01

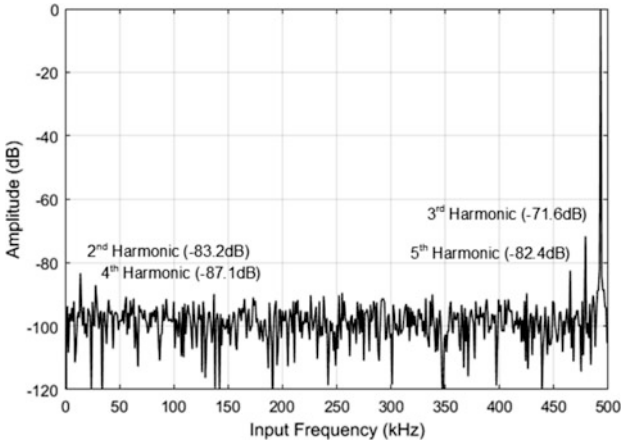


Fig. 2.38 The measured FFT spectrum at 1 MS/s with an input tone of 493 kHz. The SFDR is 71.6 dB and the SNDR is 54.6 dB. The ENOB at 493 kHz is 8.77

In summary, a 10-bit SAR ADC was presented in this section. A prototype was fabricated in 180 nm CMOS technology. The design uses an energy efficient switching procedure and a split-capacitor array. The measurement results successfully meet the design specifications, with a comparable performance among the state-of-the-art ADC designs for the neural recording purposes. As a part of the neural interface system, the power consumption of the ADC was usually not the bottleneck. So this work didn't seek to aggressively minimize the ADC's power using techniques like charge recycling [124], asynchronous timing [127], or step charging [125]. The supply voltage was kept at 1.8V to be compatible with the

Table 2.5 The measured specifications of the 10-bit SAR ADC

Specification	Measurement result
Technology	180 nm
Supply voltage	1.8 V
Input range	3 Vp-p
Sampling rate	1 MSps
Active area	0.042 mm^2
INL	−0.82/+0.77 LSB
DNL	−0.49/+0.56 LSB
SNDR	54.6 dB
SFDR	71.6 dB
ENOB	8.77
FoM	98 fJ/conv-step

neural recording front-end. The designed ADC was later integrated in a bidirectional neural interface system-on-chip, and used in long-term recording experiments in freely behaving animals.

2.5 A Compressed Sensing Neural Signal Acquisition System

2.5.1 Introduction

The wireless telemetry is the power bottleneck of most wireless neural recording systems [137]. On-chip data compression is an effective solution to reduce the power consumption by reducing the data rate. Various on-chip data compression techniques for neural signal acquisition have been proposed. For single or multi-units recording, action potential detection [138] or classification [139] is the most effective way to reduce the wireless data rate, and can also be used to drive prosthetics directly. The hardware implementation of an action potential detection unit can be as simple as a comparator with a pre-defined threshold. A compression ratio higher than 100× can be achieved with a minimum power consumption [51]. However, the action potential detection based compression drops most of the raw waveform, and is vulnerable in long-time recording since the spike waveform may change due to the electrode impedance drifting or electrode displacement. For EEG, ECoG, or LFP, wavelet transformation is an effective solution, given its high compression ratio and good reconstruction quality [140, 141]. However, the hardware implementation of a wavelet transformation is non-trivial and usually takes a considerable amount of area and power consumption. Moreover, the custom design for a specific signal type and sampling frequency significantly limit the applications of these recording systems.

Compressed sensing is an emerging signal processing technique that enables sub-Nyquist sampling and near lossless reconstruction of a signal [142, 143].

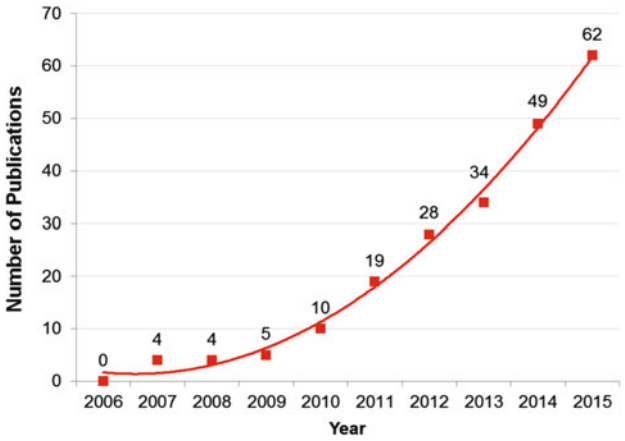


Fig. 2.39 Historical trend for publications using compressed sensing technique in biomedical signal acquisition in the past decade. Data retrieved from Web of Science

Since its introduction in 2006 [144], the compressed sensing technique has been successfully applied to rapid MRI [144], computational image sensors [145], biomedical sensors [137, 146], high frequency receivers [147], and many other applications. Compressed sensing is especially attractive to neural signal recording given its minimum hardware cost in the front-end favoring the power constraint of implanted devices.

Prior research shows the sparsity of neural signal in different frequency bands [146, 148–150]. Since an on-chip transformation using a random matrix usually achieves a sufficient incoherence with a restricted isometry property (RIP) [151], a general–purpose recording device can be designed without the knowledge of the target signal. In addition, the compressed sensing measurements can be used in signal processing (e.g., machine learning classifiers) [152]. Without a full reconstruction of the raw signal, the processing in the compressed domain can be easily implemented in a low-power embedded system.

Figure 2.39 shows a survey of publications related to compressed sensing's applications in neural recording. It clearly shows that compressed sensing has a fast growth trend, and plays an increasingly important role in the neural signal acquisition system design. This section presents the design and analysis of a fully integrated wireless compressed sensing neural signal acquisition system for chronic recording and brain–machine interface. All experimental procedures used in this study were approved by the institutional animal care and use committee (IACUC) of the University of Pennsylvania. Some of the figures and tables presented in this section were originally published in [153] ©IEEE. Reused, with permission.

2.5.2 A Brief Background of Compressed Sensing

Compressive sensing (CS) is a signal processing technique that enables sub-Nyquist sampling and near lossless reconstruction of a signal with sparsity in a certain domain. The technique is particularly appealing for low-power high channel-count neural signal recording. This section gives a brief introduction to the compressive sensing theory. Detailed explanation and rigid mathematical proof can be found in [143, 144, 151].

2.5.2.1 Compression Process

Assume the digitized signal $\mathbf{x}$ has a dimension of N, denoted by $\mathbf{x} \in \mathbb{R}^{N\times 1}$. Consider a general linear measurement process that computes $\mathbf{y}$ with a full row-rank matrix denoted by $\Phi \in \mathbb{R}^{M\times N}$, and $M \ll N$

$$\mathbf{y} = \Phi\mathbf{x} \tag{2.46}$$

where $\mathbf{y}$ is the compressive sensing data, and Φ is the sensing matrix. It should be noticed that the sensing matrix is known to the reconstruction algorithm. The signal x can be expressed as:

$$\mathbf{x} = \sum_{i=1}^{N} s_i \Psi_i \tag{2.47}$$

where $\mathbf{s}$ is the representation of the signal in the Ψ domain. The signal $\mathbf{x}$ is $K-$ *Sparse* if only K of the $\mathbf{s}$ coefficients are non-zero. The signal is compressible if it is $K-$ *Sparse*. The $\mathbf{y}$ can be written as:

$$\mathbf{y} = \Phi\Psi\mathbf{s} \tag{2.48}$$

2.5.2.2 Reconstruction Process

The signal reconstruction process is to use the $\mathbf{M}$ measurements in $\mathbf{y}$, the measurement matrix Φ, and the basis Ψ to reconstruct the signal $\mathbf{x}$, or equivalently, its sparse representation $\mathbf{s}$. Since $M \ll N$, the equation is underdetermined, which means there are infinite $\mathbf{x}$ (or $\mathbf{s}$) that satisfy the condition. Therefore, the signal reconstruction process is to find out the signal's sparse coefficient vector.

The classical approach is to find the vector in the translated null space with the smallest ℓ_2 norm by solving:

$$\hat{\mathbf{s}} = \text{argmin}||\mathbf{s}'||_2 \quad \text{such that} \quad \Phi\Psi\mathbf{s}' = \mathbf{y} \tag{2.49}$$

However, the ℓ_2 minimization usually has difficulty in finding a K-Sparse solution. ℓ_0 norm can recover a K-Sparse signal exactly with a high probability.

$$\hat{\mathbf{s}} = \text{argmin}||\mathbf{s}'||_0 \quad \text{such} \quad \text{that} \quad \Phi\Psi\mathbf{s}' = \mathbf{y} \tag{2.50}$$

Unfortunately, solving Eq. (2.50) is both numerically unstable and NP-complete. While ℓ_1 norm can exactly recover K-Sparse signal and closely approximate the signal with a high probability.

$$\hat{\mathbf{s}} = \text{argmin}||\mathbf{s}'||_1 \quad \text{such} \quad \text{that} \quad \Phi\Psi\mathbf{s}' = \mathbf{y} \tag{2.51}$$

This is a convex optimization problem and can be conveniently reduced to a basis pursuit problem, with a computational complexity about $O(N_3)$.

2.5.2.3 Reconstruction Evaluation Criteria

Several numerical derivations are used to evaluate the performance of individual reconstruction algorithms and dictionaries. The commonly used criteria include compression ratio and signal-to-noise and distortion ratio.

The Compression Ratio (CR) is defined as:

$$\text{CR} = \frac{N}{M} \tag{2.52}$$

The signal-to-noise and distortion ratio (SNDR) is defined as [137]:

$$\text{SNDR} = 20 \times \log \frac{||\mathbf{x}||_2}{||\mathbf{x} - \hat{\mathbf{x}}||_2}. \tag{2.53}$$

2.5.3 System Overview

The paradigm of the hypothetical chronic wireless neural signal acquisition system is illustrated in Fig. 2.40. The system has a dedicated implantable subsystem and a flexible external subsystem. The implantable subsystem contains the proposed compressed sensing neural recording SoC, an inductive charging module, and a super capacitor. The device will need to be sealed in a biocompatible package. The device can be placed under the skin, above the skull bone. The recording electrode can be placed in any brain area of interest. The external subsystem

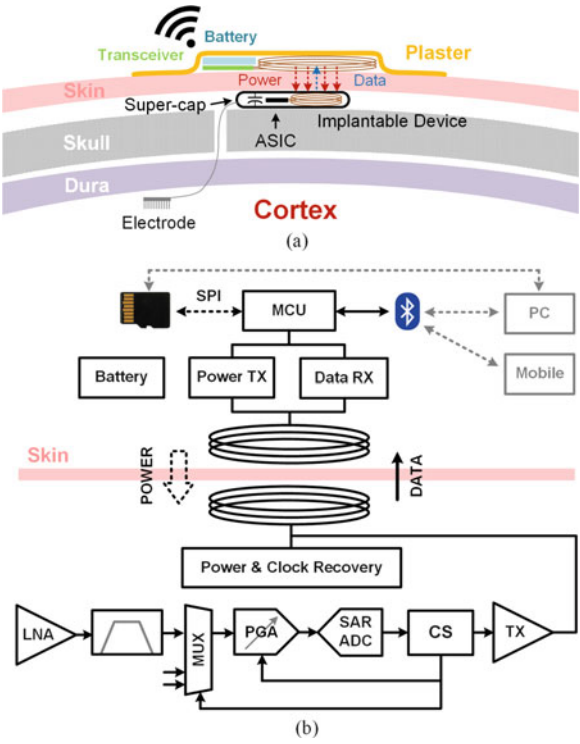


Fig. 2.40 (**a**) Illustration of the hypothetical chronic neural signal recording system using the fully integrated compressed sensing chip, and (**b**) the block diagram of the system

consists of a standard wireless transceiver, a rechargeable battery, and a coil. The external subsystem powers the implanted device and collects data back through back-scattering.

The advantages of the proposed system are threefolds: (1) the implanted wireless device leaves the skin intact, which reduces the risk of infection, (2) the battery is left externally so that the device's lifetime will not be limited by the battery's recharging cycles, and the toxicity associated with batteries will not be a potential danger to the subject, (3) the external transceiver makes the system flexible and versatile, for instance, different wireless solutions or flash memory can be used for different applications. Upgrading the system is also much easier, since the chronic implant can be used for years or even decades while the external digital and wireless electronics can be upgraded easily.

A single pair of coils is used for both power delivery and data read back. A carrier frequency of 13.56 MHz is chosen considering the trade-off between the power transfer efficiency and the data rate. Compressed sensing reduces the data rate of the wireless uplink, which is especially helpful for the multiple channel recordings.

2.5.4 Circuit Implementation

2.5.4.1 Energy Efficient Analog Front-End

The block diagram of one analog recording channel is shown in Fig. 2.41. A fully differential low-noise instrumentation amplifier (IA) is used to amplify the neural signal. The following Gm-C based high pass filter stage (HPF) conditions the

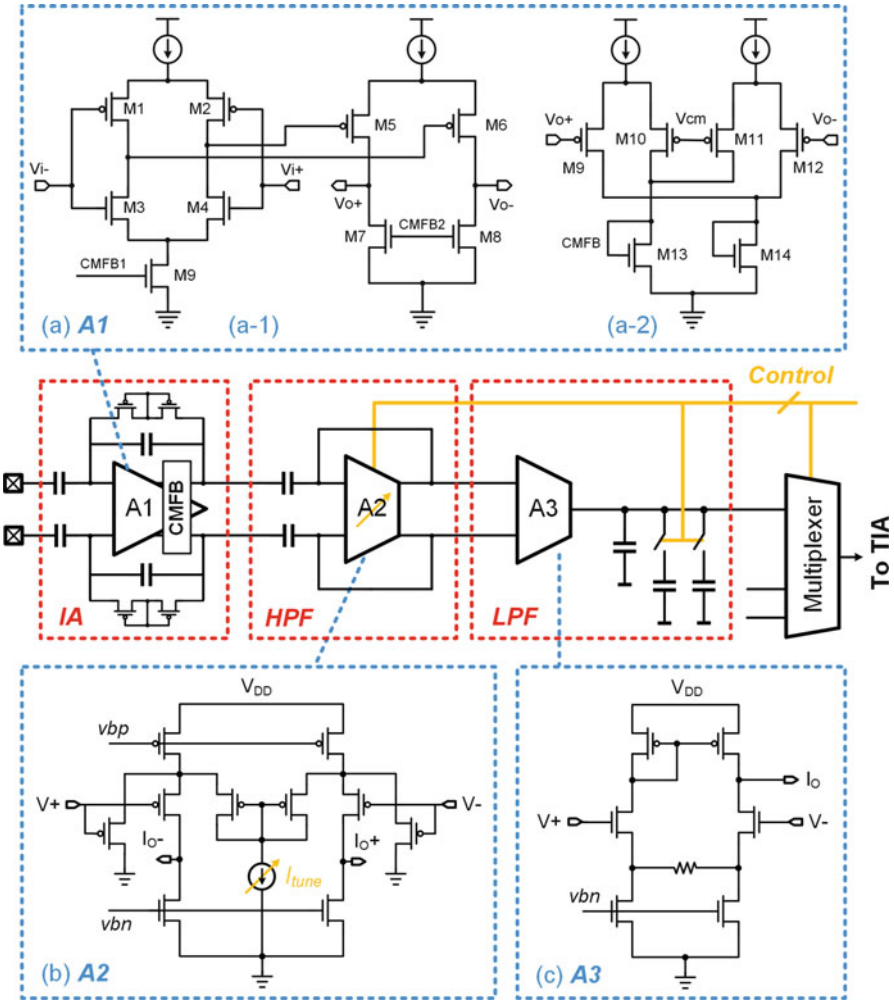


Fig. 2.41 The circuit schematic of one analog front-end channel of the proposed system (Part I). The signal chain includes: amplification, filtering, voltage-to-current conversion, multiplexing and digitization. The circuit schematic of (**a**) the low-noise neural amplifier, (**b**) the OTA with an extended linear range, and (**c**) the OTA with a programmable transconductance

signal with a tunable cut-off frequency. The next stage (LFP) is an operational transconductance amplifier (OTA) that converts the voltage signal into a current in a programmable low-pass frequency corner.

The IA in this work is a fully differential capacitor-coupled neural amplifier, which amplifies the weak neural signal in a wide frequency band. The input capacitors block the large electrode offset and half-cell potential from the interface, giving a maximum input range. The closed-loop differential gain is set to be 34 dB to relieve the noise requirement of the following stages. The core of the IA is a low-noise OTA, as shown in Fig. 2.41a-1. The OTA has been designed to maximize the noise and power efficiency. Compared with the design presented in Sect. 2.2.2, a two-stage topology is used to provide a sufficient open-loop gain. A complementary input stage (M1–M4) is used to increase the overall transconductance without increasing the quiescent current. The complementary input amplifier suffers from PVT variations [110], thus additional common-mode feedback circuit, as shown in Fig. 2.41a-2, is adopted to stabilize the DC output at half of the supply voltage. All input transistors are biased in the sub-threshold region to achieve a high energy efficiency. Since the complementary stage has a limited input range, a fully differential structure is chosen. The first stage dominates the noise, and the input-referred noise of the OTA can be expressed as:

$$\overline{v_{i,n,\text{tot}}^2} = \frac{1}{(g_{m1}+g_{m3})^2}\left[8KT\gamma(g_{m1}+g_{m3}) + 2\left(\frac{K_N g_{m3}}{C_{\text{ox},N} f W_N L_N} + \frac{K_P g_{m1}}{C_{\text{ox},P} f W_P L_P}\right)\right]\Delta f \tag{2.54}$$

where g_{m1} (=g_{m2}) are the transconductance of the transistor M1 (M2), and g_{m3} (=g_{m4}) are the transconductance of the transistor M3 (M4). The flicker noise can be reduced by increasing the widths and lengths of the input transistors. A biasing current of 1 μA is used in the first stage as a trade-off between power and noise. A biasing current of 20 nA is used in the second stage. The dominant pole is set at the second stage, and the stability is guaranteed by adding an additional capacitive load.

An ultra low-power programmable bandpass filter is integrated into each channel for selecting the frequency band of interest. The first stage is a fully differential Gm-C highpass filter. The circuit schematic of the Gm block is shown as A2 in Fig. 2.41b. Current division and local feedback are used to achieve a low transconductance and an extended linear input range. The cut-off frequency can be programmed by tuning the transconductance. The second stage of the filter is a single-ended Gm-C based lowpass filter. The circuit schematic of the Gm block is shown as A3 in Fig. 2.41c. Source degeneration is used to achieve a high linearity. The differential voltage signal is converted into a single-end current signal. Since a standard current mirror load is used, no extra power is wasted for this conversion, but the single-ended operation reduces the capacitor array size by half, which is important for

this design to be implemented at the channel level. The lowpass frequency can be programmed by selecting the load capacitor.

The shared part of the analog recording front-end is shown in Fig. 2.42. The current output from each channel is multiplexed and then converted to a voltage using a transimpedance amplifier (TIA) with a programmable gain. A single-to-differential (S2D) converter is used to drive the differential input ADC with an

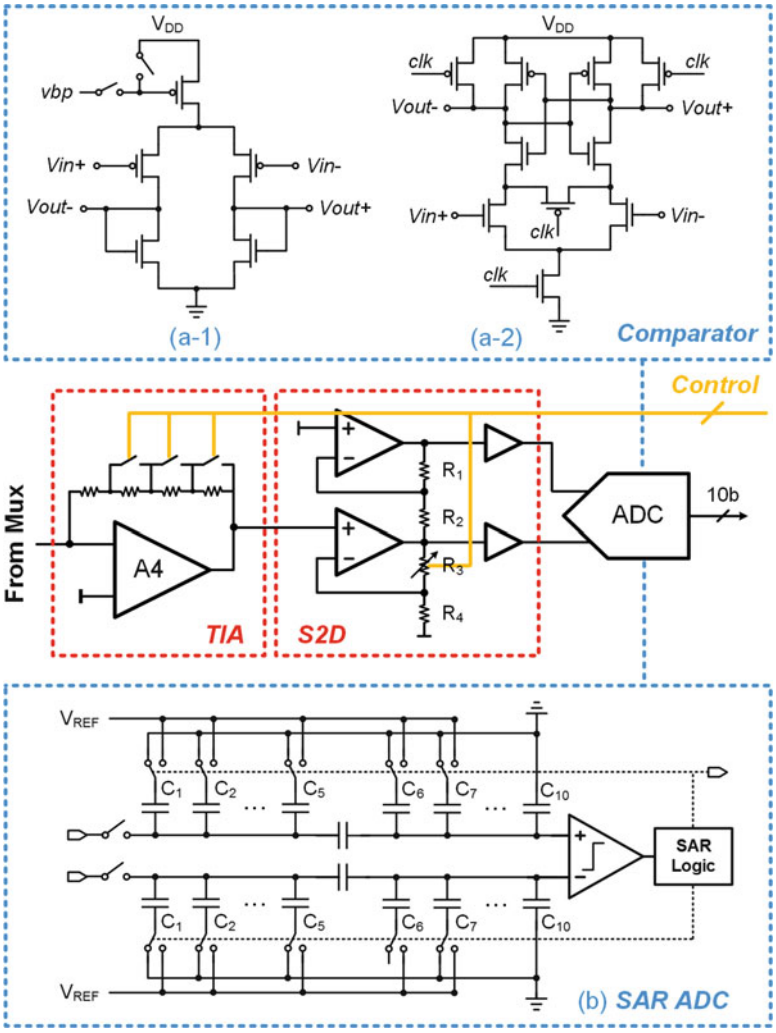


Fig. 2.42 The circuit schematic of the analog front-end of the proposed system (Part II). A current-to-voltage conversion with a programmable gain and a 10-bit SAR ADC is used to digitize the signal. The boxed windows show the circuit schematic of (**a**) the comparator and (**b**) the SAR ADC

additional programmable transimpedance. A 10-bit SAR ADC digitizes the signal. The design details of the ADC have been presented in Sect. 2.4.

The single-ended current output from the 16 channel is selected by a multiplexer. The single-ended signal reduces the effort in routing, and the R-I drop in the long routing trace doesn't corrupt the current signal, thus making it less susceptible to noise. The following TIA stage is used to convert the current signal back to a voltage in a programmable gain, as shown in Fig. 2.42. The gain can be set to be 5×, 6×, 7×, 8× by the compressed sensing digital processor. The gain of 2×, 4× can be easily achieved by shifting bits in the binary digital processor, and the 3× can be achieved from shifting the 6× signal by 1 bit.

2.5.4.2 Compressed Sensing Module

The compressed sensing processing is implemented in the digital domain. The digitized neural signal, x_{in}, of a single channel is fed into the digital processing module.

$$y = \Phi x_{\text{in}} \tag{2.55}$$

that can be written as,

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1N} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{M1} & \Phi_{M2} & \cdots & \Phi_{MN} \end{bmatrix} \begin{bmatrix} x_{\text{in}_0} \\ x_{\text{in}_1} \\ \vdots \\ x_{\text{in}_N} \end{bmatrix} \tag{2.56}$$

Equation (2.56) can be rewritten in the form of a sum of vector multiplications, as:

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_M \end{bmatrix} = \sum_{i=1}^{M} \left(\begin{bmatrix} \Phi_{1i} \\ \Phi_{2i} \\ \vdots \\ \Phi_{Mi} \end{bmatrix} x_{\text{in}_{i-1}} \right) \tag{2.57}$$

There are two modes of operation. In the simple mode, the entries of the sampling matrix Φ are assigned to be 0, $+1$, or -1; in the high resolution mode, the entries can be assigned to be $0, \pm\frac{1}{8}, \pm\frac{2}{8}, \ldots, \pm\frac{7}{8}$.

In order to avoid a large on-chip storage for the sampling matrix, a shift register chain is used to preload the coefficients at the beginning of each sampling loop. Figure 2.43 shows the block diagram of the compressive sensing processing unit. Parallel output values from the ADC are fed into the digital model. A simple sign control is applied before sending the ADC output to the adder under the simple mode. Under the high resolution mode, the entries coefficients $\pm\frac{3}{8}$, $\pm\frac{5}{8}$ and $\pm\frac{7}{8}$ are

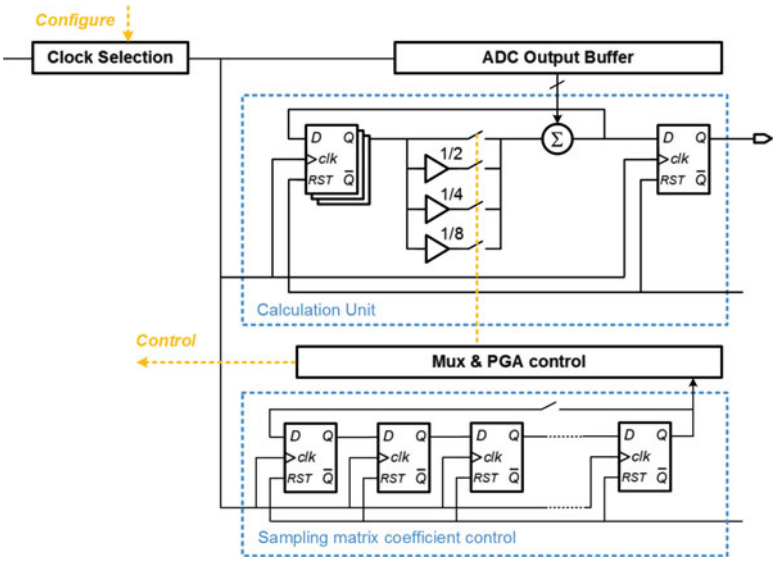


Fig. 2.43 The block diagram of the compressive sensing processing module. A linear congruential pseudo random number generator is used to generate all the entries of the sampling matrix

realized by configuring the gain of the analog amplifier to 3, 5, and 7, respectively, while shifting the ADC output by 3-bit before sending the ADC's output to the adder.

There are M vector multiplication units integrated in the system. The entries of Φ are randomly generated off-line and used for the logic control inside of each vector multiplication unit. The output measurement y is reset after every N iteration. The dimension of x_{in} is controlled by the iteration times. A parallel to serial convertor is integrated in the system for the readout of the measurements.

According to the CS theory, a dictionary for sparsifying neural signals is required for sparse recovery. In this section, neural data recording without compression is performed to generate a database for the algorithm analysis at the very beginning. The database is divided into two halves, where one half is used for training signal dependent dictionary $\mathbf{D}$ by an unsupervised dictionary learning algorithm [150] and another half is used for testing the recovery performance. In the proposed CS framework, we adopt an on-chip Bernoulli sensing matrix $\mathbf{\Phi}$ to compress the neural spikes or LFP $\mathbf{x}$ of length N into measurements y of length M, where normally $M \ll N$ and compression ratio is defined by $\frac{M}{N}$, as in Eq. (2.5.2.3). The recovery problem can be solved by Orthogonal Matching Pursuit [154],

$$\min_{\mathbf{a}} ||\mathbf{y} - \mathbf{\Phi D a}||_2^2 \quad s.t. \ ||\mathbf{a}||_0 \leq \Phi, \tag{2.58}$$

where **a** is the sparse coefficient vector and S indicates the sparse level. The recovered signal is defined as $\hat{\mathbf{x}} = \mathbf{Da}$ and the recovery quality is quantitatively evaluated by the SNDR, as defined in Eq. (2.5.2.3).

2.5.4.3 On-Chip Wireless Power and Data Link

A low-power backscatter based wireless transmitter is designed to communicate with an external transceiver. The backscatter transmitter consists of a PWM encoder and a buffered transistor for the antenna impedance modulation [48].

An active rectifier is used to achieve a high power efficiency [155]. Coupling coils are implemented off-chip. The system clock is recovered from the power waveform [86]. The circuitry of the clock recovery and division module is shown in Fig. 2.44. The module consists of a Schmitt trigger and several D flip-flops. The Schmitt trigger makes the circuit more resistant to the noise in the power waveform. Figure 2.45 shows the circuit schematic of the Schmitt trigger [156]. The D flip-flop guarantees the clock has a 50% duty cycle. Several different clocks can be divided from the following D flip-flops. The clock frequency selection is configured in a register.

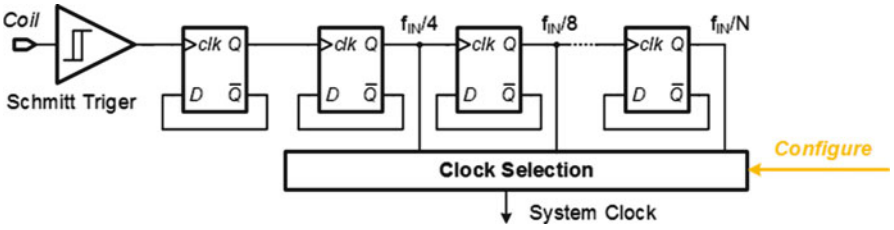
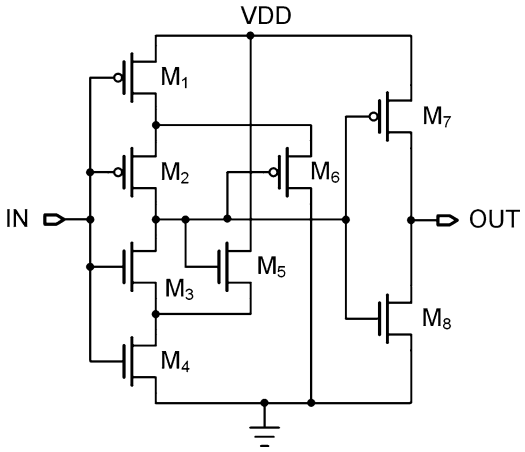


Fig. 2.44 The circuit schematic of the clock recovery and clock division module

Fig. 2.45 The circuit schematic of the CMOS Schmitt trigger



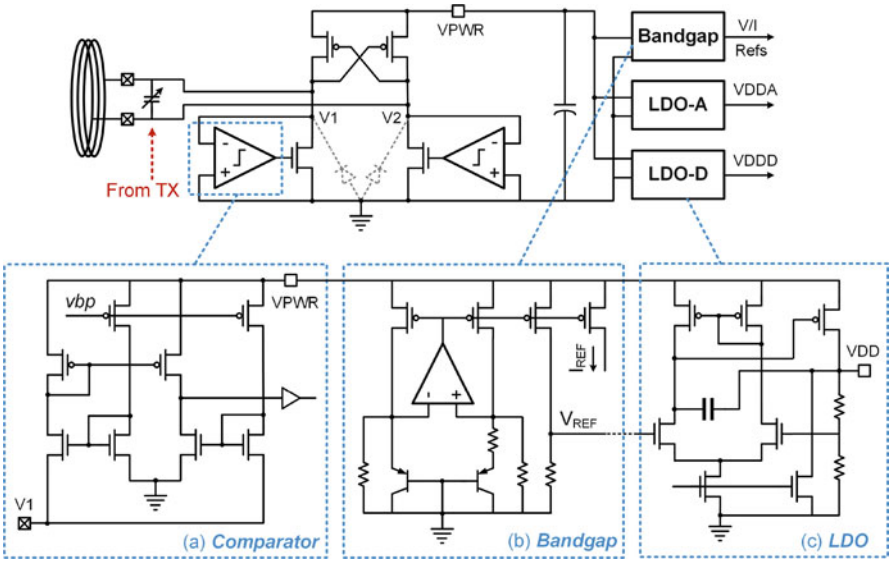


Fig. 2.46 Inductive power management module, including active rectifier and LDOs for analog and digital power supplies. (**a**) Circuit schematic of the comparator, (**b**) bandgap reference, and (**c**) LDO (start-up circuits are not shown)

Standard bandgap reference and low drop-out (LDO) circuits are used in the power management unit. The block diagram and the circuit schematics of the power management module are shown in Fig. 2.46. A push-pull comparator with source input is used to drive the active diodes. The design details of the active rectifier can be found in references [155, 157, 158].

2.5.4.4 External Wireless Relay Board

An external wireless relay board has also been designed to demonstrate the proposed paradigm. The external subsystem consists of a microcontroller with an integrated wireless transceiver, envelope detection circuits for reading the backscattered signal, power transmitter circuits, and a battery management module.

A 32-bit ARM Cortex-M0 based wireless transceiver (Nordic Semiconductor nRF51822) is used as the central processor and the wireless transceiver. It features a 2.4 GHz transceiver with an integrated Bluetooth 4.0 low-energy protocol framework, which provides an easy interface to the computer or mobile devices. A reliable wireless communication up to 5 meters was measured in normal indoor environment. A Serial Peripheral Interface (SPI) based microSD card interface is optional in the system to allow a long-term wireless recording without a limited receiver range.

A computer user interface has been developed in Matlab to configure the device and read back the data. The signal conditioning and off-line analyses can also be performed in the user interface.

2.5.5 *Measurement Results*

The proposed SoC design has been fabricated in IBM 180 nm standard CMOS technology, occupying a silicon area of 2.1 mm × 0.8 mm, excluding the IO pads. A microphotograph of the fabricated chip is shown in Fig. 2.47, with major building blocks highlighted.

Bench testing was conducted to verify the functions of the chip and the system. The measured frequency response of the low-noise amplifier is shown in Fig. 2.48. The frequency response was measured point by point using a function generator 33521A and an oscilloscope MSO7034B from Agilent. The phase shift was calculated in the oscilloscope. The measured midband gain is 34.1 dB. The measured CMRR and PSRR of the analog front-end in the frequency range of 0.5 Hz to 7 kHz are >80 dB and >67 dB, respectively.

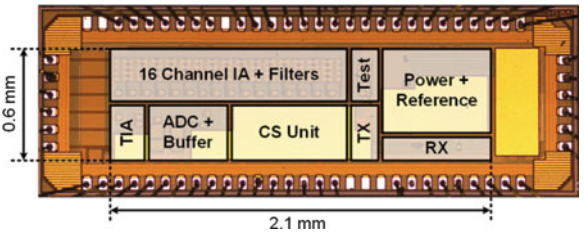


Fig. 2.47 The micrograph of the fabricated fully integrated compressed sensing neural recording front-end chip

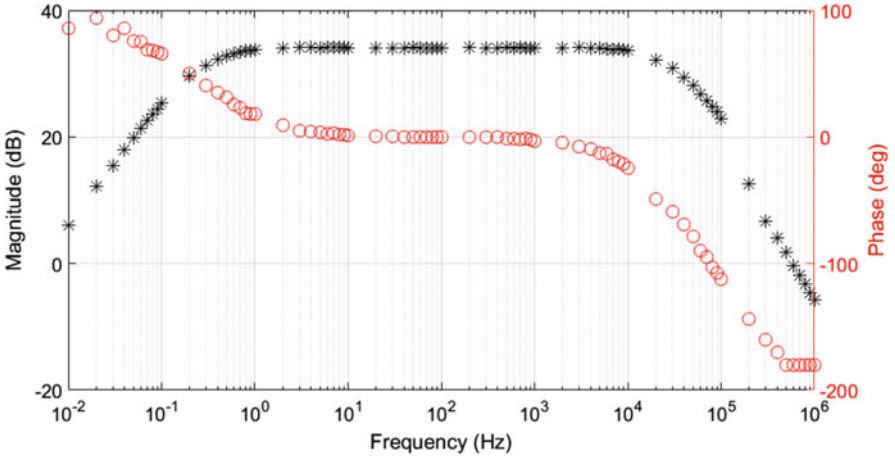


Fig. 2.48 The measured frequency response of the low-noise amplifier (without filtering stages)

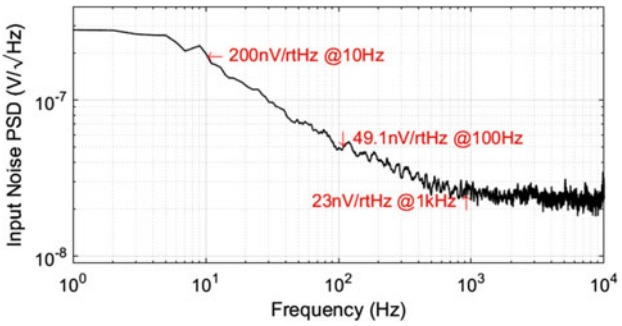


Fig. 2.49 The measured input-referred voltage noise spectrum

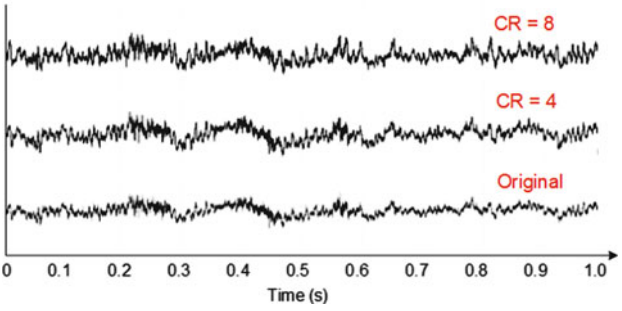


Fig. 2.50 A time-domain comparison between the uncompressed recording and the data reconstructed from recordings in different compression ratios (CR)

The input-referred noise spectrum is shown in Fig. 2.49. The noise was measured with the inputs shorted by an internal switch. The noise spectrum density is 200 nV/$\sqrt{\text{Hz}}$ at 10 Hz, 49.1 nV/$\sqrt{\text{Hz}}$ at 100 Hz, and 23 nV/$\sqrt{\text{Hz}}$ at 1 kHz. An integration under this curve from 1 Hz to 7 kHz yields an rms noise floor of 2.85 μV. The total harmonic distortion of the amplifier was measured to be −63 dB, with an input amplitude of 1 mV.

An invasive neural recording was performed in an anesthetized rat with a tungsten microelectrode placed in its motor cortex. Action potential data was extracted by configuring the filter with a passband of 300 Hz to 7 kHz. Different compression ratios from 2, to 4, to 8 and to 16 have been applied, respectively. Signal-to-noise distortion ratio (SNDR) of 3.60 dB, 9.78 dB, 30.60 dB, and 52.99 dB are achieved for compression ratios 16, 8, 4, and 2, respectively. Dual-threshold level-crossing action potential detection was used for both the uncompressed data and the restored data. A near-lossless action potential detection can be achieved while a compression ratio lower than 8 was applied.

Figure 2.50 compares the time-domain waveforms of the uncompressed and the reconstructed local field potential (LFP) sampling data sets. And Fig. 2.51 shows the comparison of the spectrums of the original uncompressed and reconstructed LFP sampling data sets. The LFP exhibited rhythmic bouts of broadband power

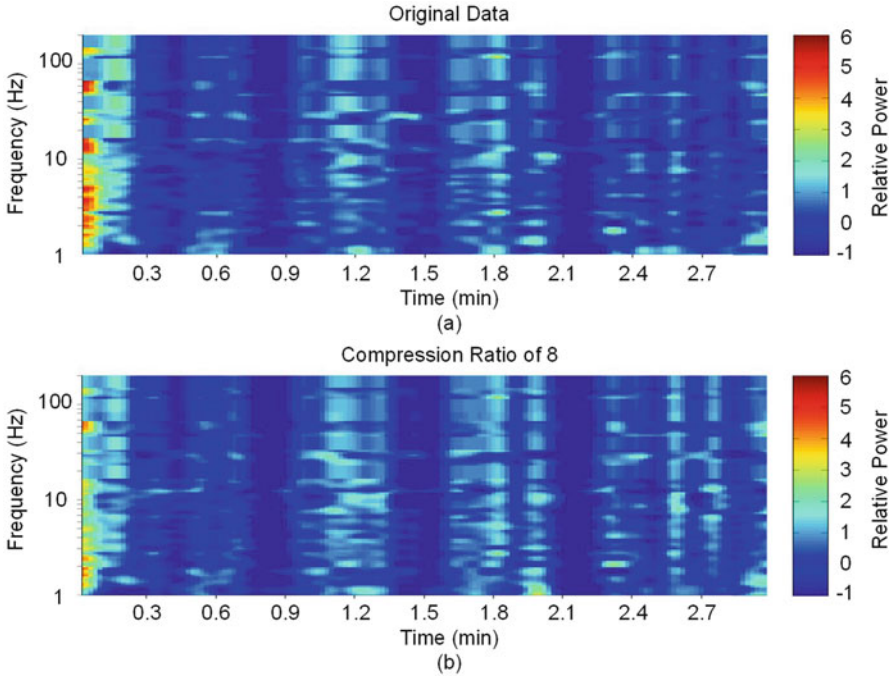


Fig. 2.51 A comparison of the spectrograms of (**a**) the uncompressed recording and (**b**) the data reconstructed from the recording with a CR of 8

interleaved with low power epochs. According to Fig. 2.51, the time-frequency content of the restored signal was very similar to the uncompressed LFP. Signal-to-noise distortion ratio (SNDR) of 9.04 dB, 4.85 dB, and 3.78 dB were achieved for compression ratios 4, 8, and 16, respectively.

A demonstration system was developed to show the proposed concept, as shown in Fig. 2.52. An open cavity plastic package was used for packaging the chip, and the size of the demonstration implantable system was limited by the package. Commercial coils were used for the inductive power and data transfer. An additional ceramic capacitor was used to improve the impedance matching. Two LEDs were used only for debugging purpose. A couple of programming and debugging pads are left (not shown). No other off-chip components were required.

In vivo evaluation of the device for a long-term operation was conducted in a rhesus macaque. An electrode was chronically implanted in the hippocampus. The recording device, including an external transceiver, was housed in a small chamber that was fixed to the skull. Figure 2.53 shows the spectrogram of a 24-h continuous recording while the monkey was freely behaving in his home cage. The recording shows the states of hippocampal activities throughout the day. Greater power at higher frequencies (>20 Hz) was associated with periods in which the animal was awake and freely moving about his home cage (hours 0–7.5 and 19–24). Greater power at low frequencies (<20 Hz) was associated with sleeping (hours 7.5–19). Individual sleep cycles can be seen. Some broadband chewing artifacts were also

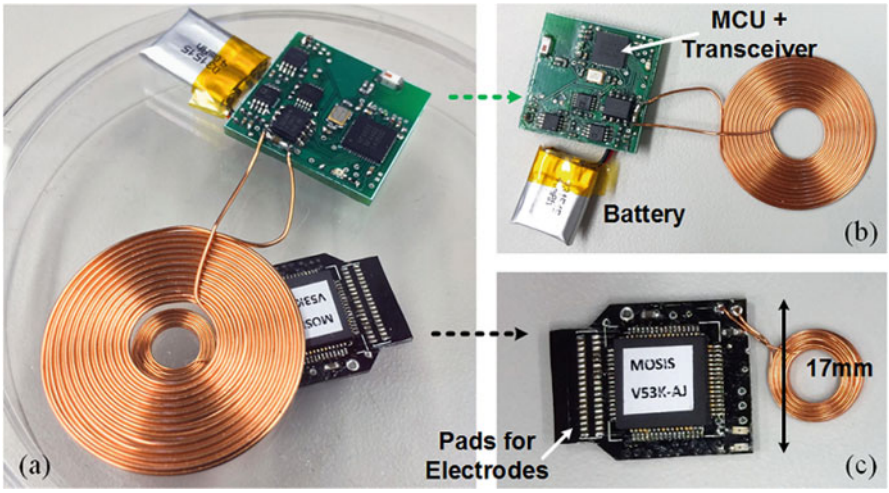


Fig. 2.52 Photograph of an assembled demonstration system. (**a**) Power and data transmission testing setup across a 5 mm plastic cap, (**b**) the external transceiver board, and (**c**) the implantable device

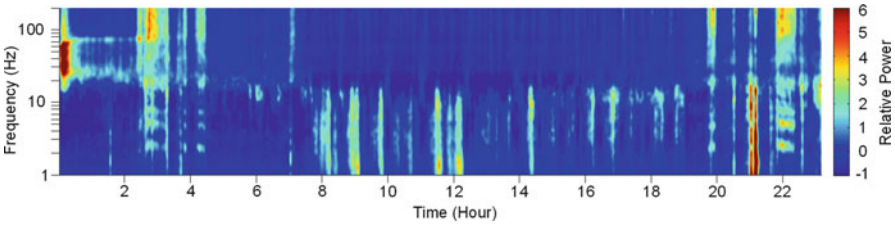


Fig. 2.53 A 24-h continuous recording in the hippocampus of a rhesus macaque during free behavior

Table 2.6 The chip specifications summary

Neural amplifier		*CS processor*	
Midband gain	34.1 dB	Input channel	Up to 16
Bandwidth	0.5 Hz–7 kHz	CS ratio	up to 8×
LNA noise	2.85 μVrms	Clock freq.	4 MHz
THD (1 mV)	−63 dB	*Wireless power and data*	
NEF/PEF	1.58/4.5	Carrier freq.	13.56 MHz
CMRR	>80 dB	Power efficiency	up to 73%
PSRR	>67 dB	Distance	up to 10 mm
SAR ADC		*Power*	
ENOB	9.1	Analog front-end	2.5 μW (per ch.)
Sampling rate	1 MSps	ADC	35 μW(@1 MSps)
INL (LSB)	+0.62/−0.85	CS processor	77 μW
DNL (LSB)	+0.69/−0.92	TX transmitter	27μW
FoM (fJ/step)	34.2	Total (avg.)	254 μW

Table 2.7 Comparison with state-of-the-art works

Reference	Deepu [106]	Gangopadhyay [159]	Zhang [160]	Biederman [56]	This work
Publication	2014 JSSC	2014 JSSC	2015 JNE	2015 JSSC	–
CMOS technology	0.35 μm	130 nm	180 nm	65 nm	180 nm
No. of channels	4	64	4	64	16
Signal type	ECG	ECG	Extracellular	Extracellular	LFP/extracellular
Input-referred noise	1.46 μV	<2 μV	3.1 μV	7.5 μV	2.8 μV
Sampling rate/ch	256/512 Hz	2 kHz	20 kHz	20 kHz	20 kHz
Front-end NEF/PEF	3.31/26.3	–	–	3.6/12.9	1.58/4.5
ADC ENOB	9.3	6.5	–	8.2	9.1
AFE + ADC power/ch	0.54 μW (512 Hz)	28 nW (2 kSps)	15 μW (20 kSps)	1.84 μW (20 kSps)	3.2 μW (20 kSps)
Compression method	Lin slope predict.	CS	CS	Spike Dect.	CS
Compression ratio	2.55×	up to 6×	8×–16×	8.3× (epochs)	8×–16×
Reconstruction SNR	–	–	<9 dB (16×)	–	9.78 dB (8×)
Wireless	–	–	–	–	Backscattering
In vivo experiment	Yes	–	Yes	Yes	Yes
System integration	–	–	–	Complete	Complete

present (around hours 3–4.5 and 20–22) corresponding to the times when the animal was fed. The overall activity pattern matches previous observations of sleep–wake changes in neural activity. The measured specifications of the chip are summarized in Table 2.6.

In this work, a fully integrated wireless neural signal acquisition system is presented. A high efficiency wireless neural signal recording SoC with integrated compressed sensing processor was designed and fabricated in 180 nm CMOS technology. An external wireless relay was used to power the implantable SoC, read back the data through backscattering, and transmit the data through a universal wireless link. The system features high energy efficiency, high flexibility, compatibility, upgradability without compromising the signal recording quality. By performing on-chip compressive sampling, the data rate is significantly reduced, which allows the system to support more recording channels without a power penalty. According to the experimental results, a compression ratio up to 8× will cause negligible loss of the data quality and/or information contained in the data. A pre-implantable system was assembled and successfully demonstrated the proposed paradigm. Bench testing and in vivo experimental results are presented. Table 2.7 compares the performance of the proposed work with prior published compressed neural signal recording front-end designs. The system shows a promising chronic neural signal recording paradigm for neuroscience research and BMI applications.

http://www.springer.com/978-3-319-67939-6

Brain-Machine Interface
Closed-loop Bidirectional System Design
Liu, X.; Van der Spiegel, J.
2018, XXXIII, 242 p. 213 illus., 158 illus. in color.,
Hardcover
ISBN: 978-3-319-67939-6