POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

ANALOG COMPRESSIVE SENSING FOR MULTI-CHANNEL NEURAL RECORDING: MODELING AND CIRCUIT LEVEL IMPLEMENTATION

FEREIDOON HASHEMI NOSHAHR
Département de génie électrique

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie électrique
Avril 2020

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

ANALOG COMPRESSIVE SENSING FOR MULTI-CHANNEL NEURAL RECORDING: MODELING AND CIRCUIT LEVEL IMPLEMENTATION

présentée par **Fereidoon HASHEMI NOSHAHR**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Yvon SAVARIA, président
Mohamad SAWAN, membre et directeur de recherche
Guchuan ZHU, membre
Frédéric NABKI, membre externe

DEDICATION

To my beloved parents, wife, son,
brothers, and friends . . .

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Prof. Mohamad Sawan for giving me the opportunity to pursue my Ph.D. program under his supervision, and his continuous support and motivation. I also highly appreciate him for his mentoring and helping me in my research. I express my sincere thanks and gratitude to Dr. Yvon Savaria, Dr. Guchuan Zhu, and Dr. Frédéric Nabki for agreeing to be members of the examination board of my Ph.D. thesis.

My appreciation also goes to all my colleagues and friends in Polystim Neurotech Lab. including but not limited to Morteza Nabavi, Mohammad Noghabaei, Mohammad Honarparvar and Armin Najarpour for their valuable interactions in my journey. Likewise, I would like to thank Réjean Lepage and Laurent Mouden for their constant technical helps and supports that backed me to demonstrate my ideas as well as Marie-Yannick Laplante for her kind helps in the administrative level. I also appreciate Dr. Kolta Arlette at the faculty of Dentistry at University of Montreal for performing the neural recordings in an in vitro experiment on the slices of a mouse brain by our prototype neural amplifier.

I am dedicating my thesis to my parents for their endless love and support along the way and I am thankful of my wife and son for their patience and non-stop encouragement towards higher achievements. In the end, my thank goes for CMC Microsystems for supporting me with their design tools and integrated circuits manufacturing facilities.

RÉSUMÉ

Dans cette thèse, nous présentons la conception d'un implant d'enregistrement neuronal multicanaux avec un échantillonnage compressé mis en oeuvre avec un procédé de fabrication CMOS à 65 nm.

La réduction de la technologie affecte à la baisse les paramètres des amplificateurs neuronaux couplés en AC, comme la fréquence de coupure basse, en raison de l'effet de canal court des transistors MOS.

Nous analysons la fréquence de coupure basse et nous constatons que l'origine de ce problème, dans les technologies avancées, est la diminution de l'impédance d'entrée de l'amplificateur opérationnel de transconductance (OTA) en raison de la fuite d'oxyde de grille à l'entrée des OTA. Nous proposons deux solutions pour réduire la fréquence de coupure basse sans augmenter la valeur des condensateurs de rétroaction de l'étage d'entrée. La première solution est appelée rétroaction positive croisée et la deuxième solution utilise des PMOS à oxyde épais dans la paire de l'entrée différentielle de l'OTA. Il est à noter que pour compresser le signal neuronal, nous utilisons le CS dans le domaine analogique.

Pour la réalisation, un intégrateur à capacité commutée est requis. Les paramètres non idéaux de l'OTA utilisé dans cet intégrateur, tels que le gain fini, la bande passante, la vitesse de balayage et le changement rapide de la sortie. Toutes ces imperfections induisent des erreurs et réduisent le rapport signal sur bruit (SNR) total. Nous avons simulé ces imperfections sur Matlab et Simulink pour définir les spécifications de l'OTA requis. Aussi, pour concevoir les circuits analogiques correspondant aux interfaces neuronales requises, tels qu'un amplificateur neuronal, une référence de tension compacte et à faible consommation d'énergie est requise. Nous avons proposé une référence de tension de faible consommation d'énergie sans utiliser le transistor bipolaire parasite de la technologie CMOS pour diminuer la surface de silicium requise. Finalement, nous avons complété l'encodeur de CS et un convertisseur analogique-numérique à approximation successive (SAR ADC) requis pour la chaine d'enregistrement des signaux neuronaux dans ce projet.

ABSTRACT

In this thesis we present the design of a multi-channel neural recording implant with analog compressive sensing (CS) in 65 nm process.

Scaling down technology demotes the parameters of AC-coupled neural amplifiers, such as increasing the low-cutoff frequency due to the short-channel effects of MOS transistors.

We analyze the low-cutoff frequency and find that the main reason of this problem in advanced technologies is decreasing the input resistance of the operational transconductance amplifier (OTA) due to the gate oxide static current leakage in the input of the OTA. In advanced technologies, the gate oxide is thin and some electrons can penetrate to the channel and cause DC current leakage. We proposed two solutions to reduce the low-cutoff frequency without increasing the value of the feedback capacitors of the front-end neural amplifier. The first solution is called cross-coupled positive feedback, and the second solution is utilizing thick-oxide PMOS transistors in the input differential pair of the OTA. Compress the neural signal, we utilized the CS method in analog domain.

For its implementation, a switched-capacitor integrator is required. Non-ideal specifications of OTA of CS integrator such as finite gain, bandwidth, slew rate and output swing induce error and reduce the total signal to noise ratio (SNR). We simulated these non-idealities in Matlab and Simulink and extracted the specification of the required OTA. Also, to design analog circuits such as neural amplifier a low power and compact voltage reference is required. We implemented a low-power band-gap reference without utilizing parasitic bipolar transistor to decrease the silicon area. At the end, we completed the CS encoder and successive approximation architecture analog-to-digital converter (SAR ADC).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

BAN	Body-Area Network
BMI	Brain Machine Interface
RF	Radio Frequency
PA	Power Amplifier
CS	Compressive Sensing
LNA	Low Noise Amplifier
PGA	programable Gain Amplifier
ADC	Analog to Digital Converter
SNR	Signal to Noise Ratio
PMU	Power Management Unit
CU	Control Unit
OTA	Operational Transconductance Amplifier
iEEG	intracranial Electroencephalography
BGR	Band-Gap Reference
CFN	Capacitive Feedback Network
MOS	Metal Oxide Semiconductor
PRG	Pseudo-Random Generator
LFSR	Linear Feedback Shift Register
SAR ADC	Successive Approximation Architecture Analog to Digital Converter
SAR	Successive Approximation Register
DAC	Digital to Analog Converter
AFE	Analog Front-End
EEG	Electroencephalogram
ECoG	Electrocorticography
AP	Action Potential
LFP	Local Field Potential
TDM	Time Division Multiplexing
CMRR	Common Mode Rejection Ratio
PSRR	Power Supply Rejection Ratio
LPF	Low-Pass Filter
IRN	Input-Referred Noise
HPF	High-Pass Filter
PMOS	P-type Metal Oxide Semiconductor

NMOS	N-type Metal Oxide Semiconductor
CMOS	Complementary Metal Oxide Semiconductor
OLN	Open Loop Network
CAFN	Capacitive Amplifier Feedback Network
MCCFN	Miller Compensated Capacitive Feedback Network
NEF	Noise Efficiency Factor
ACCIA	AC-coupled Chopper-stabilized Instrumentation Amplifier
CBIA	Current Balancing Instrumentation Amplifier
CHS	Chopper Stabilization
VGA	Variable Gain Amplifier
AC	Alternating Current
DC	Direct Current
RRL	Ripple Reduction Loop
DSL	DC Servo Loop
PFL	Positive Feedback Loop
SRAM	Static Random Access Memory
DSP	Digital Signal Processor
EMG	Electromyogram
EKG	Electrocardiogram
SC	Switched-Capacitor
CF	Compression Factor
PDF	Probability Density Function
FFT	Fast Fourier Transform
THD	Total Harmonic Distortion
CCPF	Cross-Coupled Positive Feedback
PCB	Printed Circuit Board
IC	Integrated Circuit
ACSF	Artificial Cerebrospinal Fluid
TSMC	Taiwan Semiconductor Manufacturing Corporation
FinFET	Fin Field Effect Transistor
BJT	Bipolar Junction Transistor
CTAT	Complementary To Absolute Temperature
PTAT	Proportional To Absolute Temperature
TSC-VR	Two-Stage Cascode Voltage Reference
VRRS	Voltage Reference with Regulated Supply
PONRST	Power-on-Reset

DFF	D-Flip-Flops
S/H	Sample and Hold
SOC	System On a Chip

CHAPTER 1 INTRODUCTION

1.1 Motivation and Objectives

In the recent decades, collaboration between the medical and engineering fields has brought many advances in the diagnosis and treatment of health diseases. The field of biomedical engineering, as a combination of these two fields, has emerged, grown and proved to be promising to help handicapped people and patients in improving their life-style. Also, increasing healthcare costs are becoming a burden on the economy of most countries and are expected to continue to grow [1]. One method to improve health care is more persistent patient monitoring. In many cases, for these methods to be effective, patient's physiological data should be monitored and analysed all the time. The goal of persistent patient monitoring is healthcare personalization that can result from better collaboration between engineering, technology and medicine. One of the most recent progress in biomedical engineering is the emergence of implantable devices. The pacemaker and cochlear implant are very common prosthetic devices that can restore vital functions.

To diagnose and treat brain disorders such as epilepsy, stroke, Parkinson's disease, movement, head trauma, chronic pain, paralysis, and depression, capturing brain signals is necessary [2–4]. The state-of-the-art and cost-effective solution is wireless monitoring of brain activity through implantable devices. These devices can operate as nodes of wireless body-area networks (BAN). Figure 1.1 shows a wireless BAN [5], which is a sample solution to the needs of continuous monitoring of biomedical signals. This kind of monitoring, as part of a diagnostic procedure or during supervised recovery from a surgical procedure has been presented in [6].

A neural recording implant is a device that captures neural signals of brain from an array of sensors and transfer them out of the body by a transmitter after digitizing. In this thesis, we investigate the challenges in the design of state-of-the-art neural recording implants in widely available manufacturing technologies such as 65 nm CMOS processes. Power consumption and chip area are two crucial factors in neural implants and they must be reduced as much as possible, due to the necessity to avoid heat damage to the brain tissues and using all kind of batteries as well as the spatial limitation in the skull. Increasing the number of electrodes and channels are desirable to develop brain machine interfaces (BMIs). However, it increases the output data rate as well as the power consumption.

Considering that most of the power in this system is consumed in the radio frequency (RF)

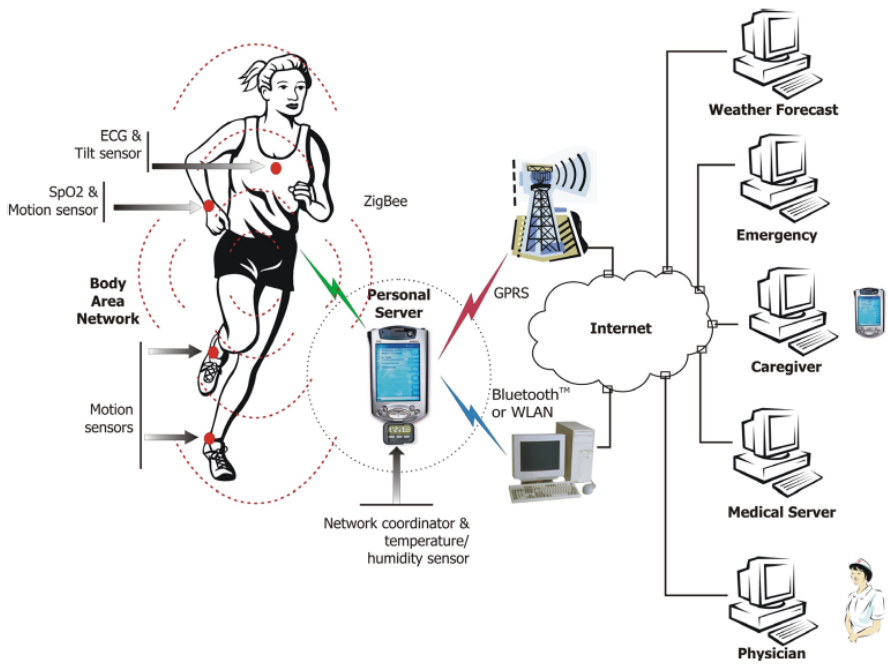


Figure 1.1 Wireless Body Area Network.

power amplifier (PA) of the transmitter compared to other circuits [7], decreasing this vast amount of data is desirable for diminishing the overall power consumption. This can be done by data compression technique. An appropriate method for doing this is compressive sensing (CS). A CS encoder decreases the sampling frequency to less than Nyquist rate and can be achieved by simple circuits in both analog and digital domains in comparison with other compression methods. The sparsity in one domain like time, frequency, etc., is a key condition for signals to utilize CS methods.

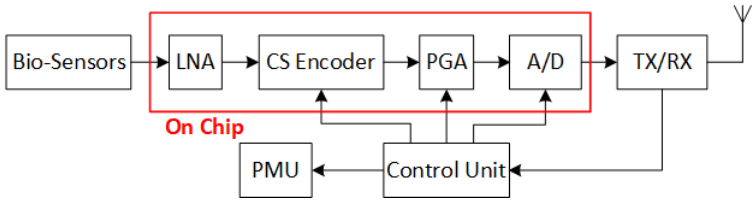


Figure 1.2 Simplified block diagram of a neural recording implant with CS encoder for one channel.

Figure 1.2 shows the simplified proposed block diagram of a neural recording implant with a CS encoder for one channel. In this figure, a neural signal is amplified first by a low noise amplifier (LNA), and then it is compressed by CS encoder in the analog domain. Compressed signals are digitized by an analog to digital converter (ADC) in order to be transmitted

wirelessly to the outside of the body. A programable gain amplifier (PGA) amplifies the signal to a dynamic range compatible with the ADC to maximize the output signal to noise ratio (SNR).

It is anticipated that future neural implants will be interactive with health provider. As the access to the implant after installation is not so easy, it is rational that the implant is designed so that some of its important parameters are configurable. The reasons that justify configurability are as follows:

1. Process variation of microelectronic devices may provoke deviations from desired design specifications.
2. Power consumption is a critical factor in implantable devices which should be managed.
3. In case of monitoring specific circumstances such as sleeping, awakening, working etc., if may be necessary to increase the quality of recorded neural signals by adjusting the compression factor (CF) and the resolution of the ADCs.
4. Variation of physiological parameters in patients as a function of age requires customization and calibration for each patient devices.

The proposed block diagram (Figure 1.2) includes a control unit which receives instructions from the health care provider through embedded interfaces and applies them to the other parts of the implant such as the CS encoder, ADC, programmable gain amplifier (PGA) and power management unit (PMU). It can configure the gain of the PGA, the compression factor (CF) of the CS encoder and the resolution of the ADC. These parameters control the quality of reconstructed signal such as the output SNR. PMU can also be ordered to shut down the whole system by switching it to a standby mode waiting for activation instruction. The focus of this thesis is on the modules in red box presented in Figure 1.2, which includes a LNA, a CS Encoder, a PGA and an ADC. To implement and validate the proposed solution, this system is designed using a 65 nm TSMC CMOS process.

1.2 Contributions

The contributions of this thesis are as follows:

- It is known that scaling down integration technologies degrades key parameters of AC-coupled neural amplifiers. One of these destructive effects is the increase of the low-cutoff frequency. In the literature, there is not a thorough discussion on this problem, its

causes and solutions. In this thesis, we will investigate the problem analytically, find the origin and propose two solutions. We will find that the origin of the increase in the low-cutoff frequency is the decreasing in the input resistance of the OTA due to increasing the gate oxide leakage current of the input differential pair which is addressed as the short-channel effects. We will propose two solutions to reduce the low-cutoff frequency without increasing the value of the feedback capacitors. The first solution is called cross-coupled positive feedback (CCPF), which will use pseudoresistors to provide a negative resistance to increase the input resistance of the OTA. As an advantage, only standard CMOS transistors are used in this method. Moreover, in a second method, we will utilize thick-oxide MOS transistors in the input differential pair of the OTA. We will design and fabricate a prototype exploiting the second method using the 65 nm TSMC CMOS process. We will compare the simulation and measurement results.

- Analog implementations of CS are usually done using switched-capacitor circuits. The OTAs embedded in CS integrators are not ideal. They have finite gain, bandwidth, slew rate and output swing that induce errors and reduce the total SNR of the reconstructed signal. Before implementing a CS system, it is necessary to extract the required specifications of the OTA to minimize the power consumption and silicon area of the CS encoder. As there is no reported detailed modeling in the literature, we will model these non-idealities and simulate them in Matlab and Simulink. The results will demonstrate that the SNR of the whole system is very sensitive to the gain, bandwidth and output swing of OTAs, but not to the slew rate. Also, we will model the finite input dynamic range of the ADC and will extract the gain of the CS encoder. Based on the results of these modeling and simulations, we can optimize the required specifications of the OTA to obtain the compact and low-power CS encoders.

- For proper operation of targeted neural recoding circuits, the effects of power supply reduction or fluctuation must be mitigated and a compact voltage reference is required. In advanced technologies, designing voltage references with low sensitivity to power supply is challenging due to short channel effects. In this thesis, we will propose two new voltage reference circuits for neural implant applications which are optimized for low power and low sensitivity to power supply variation in the 65 nm CMOS technology. In the first design, two stages of voltage references will be used in a cascode architecture and in the second design the voltage reference output will be regulated and fed back to the local power supply. These designs will show low power consumption of 1.72 and 2.77 μW, respectively. In addition, simulation results will show that the achieved PSRR are as low as -26.5 and -63 dB, respectively.

1.3 Thesis Organization

This Ph.D. thesis includes six chapters and its organization is as follows:

Chapter 2 is the literature review. In this chapter, we present the neural signals and their electrical specifications. We investigate the various architectures of neural recording systems in the literature. Then we explore each block of these architectures. One of the most important and challenging blocks of a neural recording implant is the neural amplifier. Therefore we pay more attention to this block than to other blocks. The different topologies of neural amplifiers, their noise reduction techniques and their high gain circuits in advanced technologies are discussed in this chapter. Finally, we cover analog to digital converters (ADCs) and data compression methods that are utilized in neural recording implants.

Systematic design of analog compressive sensing (CS) encoder is discussed in chapter 3. To design analog CS core circuit in 65 nm process, we need operational transconductance amplifier (OTA), which are not ideal. We simulate their non-idealities and with system level simulations using Matlab and Simulink. In this chapter, we investigate the non-ideal effects of OTA, such as finite gain, bandwidth, slew rate and swing of OTA on the output signal to noise ratio (SNR) of the CS core. Due to integration of iEEG samples in a CS core, the output amplitude of the CS core is much bigger than its input amplitude, and the ratio of these signals is called the gain of the CS core. At the end of this chapter, we explore the gain of the CS core for a real iEEG signal recorded from a dog.

Chapter 4 consists of two sections. The first section discusses the neural amplifier design. We design a conventional capacitive feedback network (CFN) neural amplifier. In advanced technologies, the input resistance decreases due to the gate oxide leakage. In this chapter, we explore this fact analytically and propose two solutions to reduce the low-cutoff frequency without increasing the value of the feedback capacitor. The first solution is called cross-coupled positive feedback which uses pseudoresistors to provide a negative resistance to increase the input resistance of the OTA. In the second method, we utilize thick-oxide MOS transistors in the input differential pair of the OTA. We fabricated a prototype based on the second method and the experimental results verified the simulations and our proposal. In the second section, we discuss reference circuit design. In advanced technologies, designing voltage references with low sensitivity to power supply is challenging due to short channel effects. In this section, we explore two new voltage reference designs which are optimized for low power and low sensitivity to power supply variation in the 65 nm CMOS technology.

The design of a CS encoder and of an ADC are given in chapter 5. The first section of this chapter explains the analog CS encoder circuit design. By switching a couple of feedback

capacitors, the gain of the CS core is controlled. The digital part of a CS encoder is designing a pseudo-random generator (PRG). In this section, we explain the design of a 15-bit maximal-length Fibonacci Linear Feedback Shift Register (LFSR) as a PRG which is suitable for the proposed CS core. The second section of this chapter covers the design of a configurable 7 up to 10 bit successive approximation architecture analog to digital converter (SAR ADC). In this section, we explore the design of successive approximation register (SAR) blocks, comparator and the capacitive digital to analog converter (DAC).

Finally, we draw a conclusion in chapter 6 and discuss the future works on the topic. We summarize all the work done, our contributions, and achievements in this thesis. Necessary steps to be taken for further improvements on the present design are also recommended in this section. This conclusion also presents the future direction as an extension to the current research.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

The source of this chapter is [8], our review paper. In the past decade, researchers have worked on the brain to understand its functions and monitor the brain's electrical signals to research, diagnose and treat its disorders, as well as to utilize these signals to control artificial limbs. BMIs can serve people with different clinical disorders. For example, researchers have implemented robotic limbs [9,10], speech synthesizers [11], and human neuroprosthetic control of computer cursors [12–14], utilizing less than 300 electrodes [15]. In addition, monitoring the bio-potential signals is a fundamental and vital part of a medical diagnostics system. For this purpose, patients are generally connected to a massive bio-potential acquisition equipment. However, this limits the patient's daily routine on the one hand, and on the other hand requires the long-term monitoring of diagnostics arduous [16]. One of the most promising solutions is to use neural recording implants as a part of BMI systems, which are in high demand and are being developed and improved as technology develops.

Inability to record from large numbers of neurons has limited the development of BMI. Noninvasive methods are capable of recording millions of neurons through the skull, however this signal is nonspecific and distorted [17, 18]. Utilizing electrodes placed on the surface of the cortex, an invasive method, records proper signals. However, the disadvantage of this is not being able to record deep in the brain and they average the activity of thousands of neurons [19]. Invasive techniques have been utilized by some BMIs. This is because recording single action potentials from neurons in distributed, functionally-linked ensembles is necessary for the most accurate readout of neural activities [15]. Therefore, increasing the spatial resolution and the number of electrodes is essential for developing BMI.

The implementation of a neural recording implant is multi-disciplinary, as it involves the various scientific fields such as electronics, medical, materials, electrodes and system integration. Increasing the number of electrodes and, consequently, the number of channels (in the range of thousands), creates new challenges for neural recording in the various fields mentioned. Microelectrode technology is not appropriate for these large-scale recordings [20]. Recently, Neuralink Company has built arrays of small and flexible electrodes (3072 electrodes per array), which have enabled thousands of channel recordings [21]. In the microelectronics field, large-scale recordings create many challenges with regards to decreasing the power consumption and chip area.

In the design of the neural recording implants, the two constraints, power consumption and chip area, should be addressed. Implantable circuits should consume very low power to avoid any damage to the surrounding tissue due to generated heat. Additional challenges in the design of analog front-end (AFE) of the neural recording systems arise in advanced and scaled technologies. The main reason is due to the short-channel effects of MOS transistors. These effects in the MOS down-scaled technologies decrease the transconductance (g_m) of the transistor on one hand and on the other hand increase the gate leakage current, the flicker and thermal noise power of an MOS transistor. This creates challenges in the design of the high gain and low noise neural amplifier, which will be explained in this chapter.

The remainder of this chapter is organized as follows. Section 2.2 reviews the different types of neural signals and their properties. Section 2.3 presents the essential neural recording architectures in the literature. Section 2.4 surveys the neural amplifiers. The neural amplifiers are the most challenging part of a neural recording implants. They must be compact, high gain, low power, and low noise amplifiers. To satisfy these constraints in the design of the neural amplifiers, various topologies and techniques are proposed which are presented in the subsections of section 2.4. Section 2.5 covers the ADCs that are suitable for neural implant applications based on their various architectures. Finally, section 2.6 discusses the data compression methods in neural recording systems and section 2.7 presents the conclusion.

2.2 Neural Signals

The electrical activities of the brain can be recorded through three different methods: 1) from the scalp which its corresponding signal is electroencephalogram (EEG), 2) the surface of the brain which extracts electrocorticography (ECoG) or intracranial electroencephalography (iEEG) signal, 3) within the brain which captures extracellular activities of neurons. The extracted signals from these methods have a frequency range of a few mHz to 10 kHz and their amplitude is at the range of 20 μV to 10 mV [22].

In the first method, surface electrodes can be used to un-invasively measure the biopotentials of EEG on the scalp. In contrast, in the second method, the electrodes can be placed directly on the brain surface invasively to record electrical activity from the cerebral cortex. The brain signals provided by this method have a dramatically high signal-to-noise ratio (SNR) and are less sensitive to artifacts than EEG. Besides, these signals have a high spatial and temporal resolution. Capturing the signal inside the body by utilizing implantable electrodes is the most effective approach for direct control of prosthetic devices [23].

In the third method, a sharp biocompatible microelectrode is utilized to perform bioelectrical

recordings invasively inside the body. By penetrating microelectrodes into the brain, the bioelectrical activity that is transmitted along the axon of a neuron can either be measured intracellularly in a single neuron or extracellularly from the brain [24]. The extracellular action potentials (APs) generated by depolarization of the membrane of the neuron are at the range of 100 Hz to 10 kHz and their duration is of a few milliseconds. The number of occurrences of the APs are between 10 to 120 times per second. The distance between the active neuron and the recording electrode determines the amplitude of the extracellular APs which are between 50 μVpp to 500 μVpp [2]. In the following, the extraction procedure of the APs (spikes) and the local field potential (LFP) is explained. LFP is the mean field potential generated by neurons in the vicinity of the electrode.

The neural activities are first amplified after sensing, then are low-pass filtered to obtain the LFP and are also high-pass filtered to identify the activity of single neurons using spike detection and performing sorting algorithms [25]. The LFP includes lower-frequency neural waveforms in the range of mHz to 200 Hz with the amplitude of 500 μVpp to 5 mVpp. These potentials are used for brain interfacing applications and carry complementary important information [26, 27]. High-pass filtering extracts the spikes of the nearby neurons on top of background activity. Amplitude threshold methods are used to detect these spikes. In the next step, the features of the spikes are extracted and sorted accordingly. Note that the SNR of the neurons located in the distance of 50-100 μm from the electrode is large enough to easily distinguish the activity of each single unit [28, 29].

For neurons located between 100 μm to approximately 150 μm far from the electrode, spikes can still be detected but their shapes are masked by the noise. These signals are grouped in a 'multi-unit' cluster. However, neurons farther than 150 μm from the tip of the electrode cannot be detected and are added to the background noise.

In addition to extracellular method, the intracellular procedure is explained in [30]. In this method, a sharp glass micropipette penetrates into a neuron of a slice of a brain in the laboratory for neuroscience researches. The neurons of the brain slice die after few hours of recording. A metal electrode located inside the micropipette and in contact with electrolyte is connected to an amplifier. The amplitude of intracellular APs is in the range of 10 to 70 mVpp. The neural recording parameters for different signal modalities are summarized in Table 2.1.

Table 2.1 NEURAL SIGNAL PARAMETERS

Signal Type	Amplitude	Bandwidth
Extracellular action potential	50 - 500 μVpp	100 Hz - 10 kHz
Intracellular action potential	10 - 70 mVpp	100 Hz - 10 kHz
Local field potentials	0.5 - 5 mVpp	1 mHz - 200 Hz
Electroencephalogram	10 - 400 μVpp	1 mHz - 200 Hz
Electrocorticography	10 - 1000 μVpp	1 mHz - 200 Hz

2.3 Neural Recording Architectures

The multi-channel neural recording implants are designed in three main architectures in the literature and are shown in Figure 2.1. The most common architecture illustrated in Figure 2.1(a) exploits one analog to digital converter (ADC) that is shared among all channels. Each channel has a neural amplifier, and the neural signal of each channel is passed to the ADC through an analog multiplexer. The performance of this multiplexing method, which is also referred as time division multiplexing (TDM) method in analog domain, gets limited when the number of channels increase dramatically. In order to improve the neural recording spatial resolution, the number of channels increase. This results in higher sampling frequency of the ADC and the multiplexer which in turn increases the power consumption of the ADC and the driving buffers. Since analog signals are more prone to distortion due to cross-talk noise in analog multiplexers compared to digital signals, careful design considerations have to be accounted in the design process. An example of this design is presented in [31].

The architecture shown in Figure 2.1(b) utilizes one ADC for each channel. Due to the low bandwidth of the neural signals, the sampling frequency and the power consumption of the ADCs are low. In this architecture, a digital multiplexer is used unlike the previous architecture where an analog multiplexer is used. The main advantage of utilizing digital multiplexer is that the power-consuming buffers and ADC drivers are avoided and the inter-channel crosstalk noise is eliminated. This is due to the fact that digital signals have high noise margins and are more stable compared to analog signals against crosstalk and other noises. However, this architecture has higher numbers of ADCs and consequently consumes higher area and power consumption. Therefore, area and power reduction techniques should be applied in the design process. This architecture is shown in [32].

In the third architecture, unlike the other two architectures where there is one ADC for all the channels or one ADC for each channel, one ADC is shared among multiple channels. Figure 2.1(c) shows the block diagram of this architecture where there are m rows and n columns. As shown in this figure, one ADC is assigned to each column with m rows

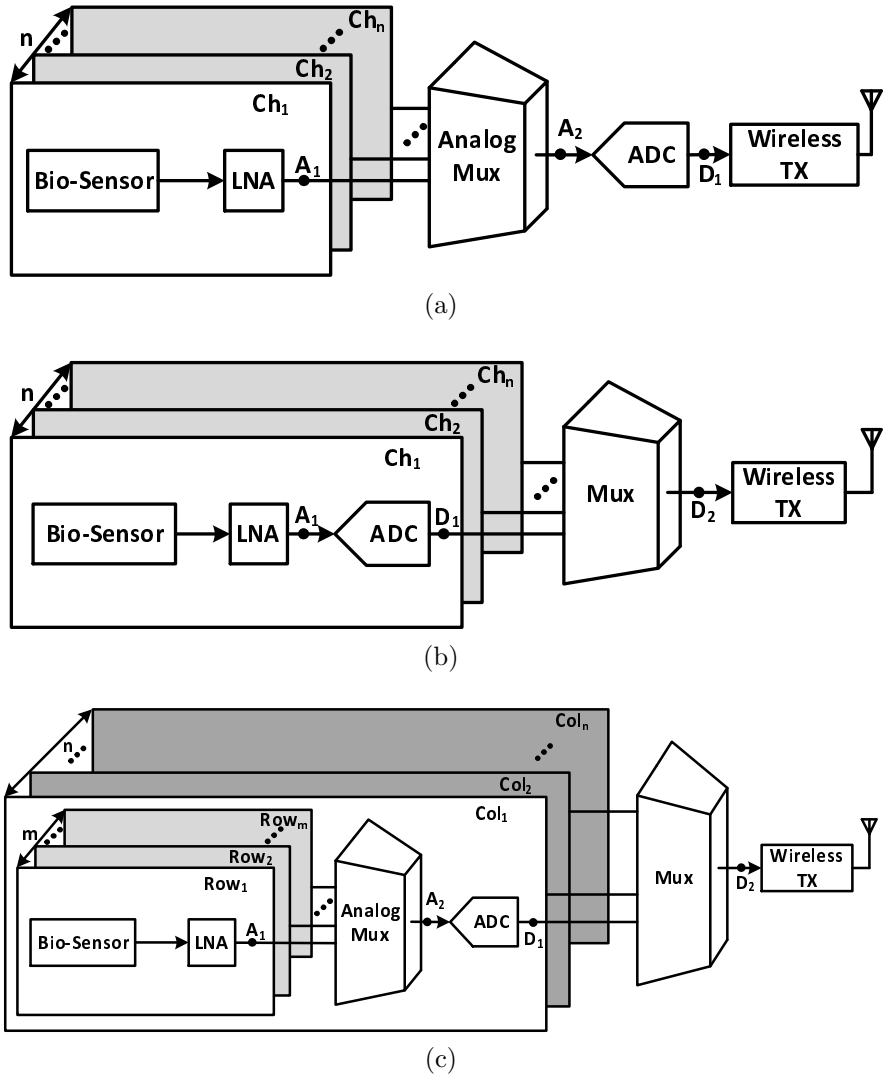


Figure 2.1 Block diagram of different multi-channel neural recording architectures. (a) This architecture shares an ADC among all of the channels. (b) This architecture utilizes an ADC for each channel. (c) This architecture shares an ADC at each column.

through an analog multiplexer. Since these multiplexers are smaller compared to the first architecture, the design considerations to avoid crosstalk is less challenging. As another advantage of this architecture, since the number of ADCs are dependent to the number of columns (n), choosing an appropriate value for n results in minimum value for the power consumption and the area especially when the total number of the channels are very high. Therefore, as a solution for the large-scale recordings, the architecture of Figure 2.1(c) is the best option. This architecture is presented in [33].

In addition to these main architectures, other non-popular architectures are reported in the literature where no ADC is involved. As an example, in [34], the analog signals are converted

to time duration using pulse width modulation (PWM) technique and transmitted to the outside of the body.

Increasing the number of channels in order to increase the spatial resolution is desirable, however, it increases the output data rate and the power consumption especially in the transmitter. In order to decrease the data rate, researchers have proposed different methods to compress the data. One of the main methods to compress the data that is utilized in neural extracellular activities is done in the time domain. In this method, only the APs are detected and transferred out of the body. Since the duty cycle of this method is at the range of 2% to 20%, the data can be compressed by a maximum factor of 50 [32, 35]. This method is applicable in both digital and analog domains.

Another compression method that is used in both analog and digital domains is compressive sensing (CS). This method is simple in implementation and efficient in compression and is suitable for neural signals specially iEEG. In this method, instead of sending all the N samples of the neural signal of each channel, M linear combination of these samples are sent where M is less than N [36–38]. The compression blocks can be implemented in analog in points A1 and A2 or can be implemented digitally in points D1 and D2 in various architectures of Figure 2.1. In Section 2.6, we explore the compression methods and their challenges in more details. In the next sections, we probe the main blocks of these architectures and their challenges.

2.4 Neural Amplifiers

As discussed in Section 2.2, neural signals have very small amplitude and bandwidth and are required to be amplified before converting to digital signals by an ADC. The amplification is done in AFE by neural amplifiers. Different DC offset voltages are generated across various electrodes due to the electrochemical reaction at the electrode-tissue interface on each channel. These voltages vary from 1 mV to 50 mV [33]. The offset voltage of channel can saturate the neural amplifier due to its very high voltage value compared to the amplitude of the neural signal. Therefore, these offset voltages should be eliminated. Besides, to design multichannel neural amplifiers for implantable applications, the power consumption and chip area should be minimized.

Different noise sources also impose challenges in the design of neural amplifiers. Neural amplifiers have very low bandwidth. Therefore the main sources of noise are flicker and thermal noise of the neural amplifiers which decreases the SNR in the output of the amplifiers. To achieve the adequate output SNR, the neural amplifier is designed as an LNA.

Generally, in the design of neural amplifiers, to provide required signal quality, several factors should be considered. These factors are sufficient gain, high SNR, appropriate bandwidth, high common mode and power supply rejection ratio (CMRR and PSRR), low power consumption and low chip area.

2.4.1 Neural Amplifier Topologies

In this section, we classified the neural amplifier topologies to AC-coupled and DC-coupled neural amplifiers, based on their tissue DC offset cancelation methods. In AC-coupled neural amplifiers, the DC offset are blocked using one or two AC-coupling capacitors placed at the input of the amplifier. To achieve high gain in the amplifier, these capacitors are usually huge. As a drawback, the chip area is increased significantly and the input impedance is decreased. A small-frequency and well-defined high-pass pole is required to record low-frequency signals while rejecting the tissue DC offset voltage.

An alternative approach to cancel the tissue DC offset effect is to utilize the DC-coupled neural amplifiers. This type of amplifier uses a low-pass filter (LPF) in the feedback path (shown in Figure 2.2) to generate a high-pass pole as the overall transfer function.

AC-coupled neural Amplifiers

One of the popular neural amplifier topologies is the closed-loop capacitive feedback amplifier introduced in [39] which is also known as conventional capacitive feedback network (CFN) topology [34,40–44]. Figure2.3(a) shows the conventional circuit architecture of this topology. A large capacitor C_I at the input is used to block the tissue DC offset. The gain of this amplifier is equal to $\frac{C_I}{C_F}$. To implement the high-pass pole, a capacitor C_F is set in parallel with a highly resistive element R_F in the feedback path.

The main drawback of this topology is its large area due to the huge input capacitor C_I. To achieve high gain and low-cutoff frequency at the range of 1 Hz and lower, a huge capacitor at the input of the amplifier is required. This large capacitor (C_I) occupies large chip area and results in reduced input impedance of the neural amplifier. For this reason, this structure is not suitable for multi-channel applications. Using two or three gain stages can reduce the size of the capacitors and consequently the chip area at the cost of increased power consumption.

The total input-referred noise of the amplifier is presented as [39]

$$\overline{v_{ni,amp}^2} = \left(\frac{C_I + C_F + C_{in}}{C_I}\right)^2 .\overline{v_{ni}^2} \tag{2.1}$$

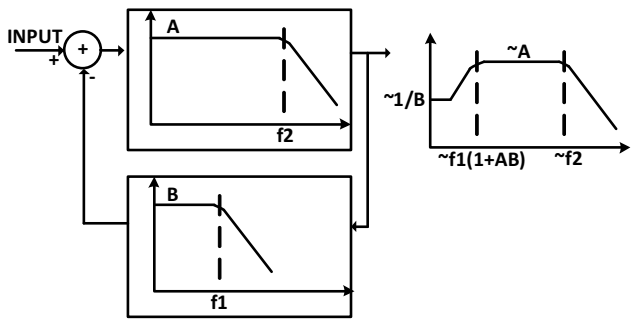


Figure 2.2 Implementing a high-pass pole using a low-pass filter in the feedback.

where C_F is the feedback capacitor, C_I is the input capacitor, C_{in} is the OTA input terminal capacitance, $\overline{V_{ni}^2}$ is the OTA input-referred noise power and $\overline{V_{ni,amp}^2}$ is the input-referred noise power of the whole neural amplifier.

Equation (2.1) shows that increasing the gain of the neural amplifier, or in other words increasing the value of C_I, for a constant C_F, reduces the overall input-referred noise (IRN). In addition, for a constant gain, increasing the sizes of the differential pair transistors, on one hand decreases the IRN power of the OTA ($\overline{v_{ni}^2}$), and on the other hand, increases the size of the OTA input capacitance C_{in} which according to (2.1) results in increasing of the noise multiplication factor. As a trade off, there is an optimum point for the size of the differential pair to minimize the overall IRN power of the neural amplifier for a specific gain.

By replacing the feedback capacitor C_F in the conventional CFN topology shown in Figure 2.3(a) to a T-capacitor network topology in Figure 2.3(b), the total equivalent feedback capacitor is reduced [45]. Therefore, the similar gain is maintained with a smaller C_I capacitor. However, this comes at the cost of increased low-cutoff frequency due to the reduced effective feedback capacitance. In order to maintain the same low-cutoff frequency, the feedback resistor has to increase. Increasing the feedback resistor, increases the input referred noise of the whole neural amplifier.

Another topology of the AC-coupled neural amplifier is shown in Figure 2.3(c). In this topology, a high-pass filter (HPF) followed by a resistive feedback non-inverting amplifier is used to cancel the DC offset [46]. The HPF is composed of an electrode capacitance and a high resistive PMOS where the bias current is programable. The operating point of M_P in Figure 2.3(c) determines the amount of resistance and the low-cutoff frequency. Since the value of the electrode capacitance varies significantly from one electrode to the other, the low-cutoff frequency is not accurate. In [47], the same structure is used to design multichannel AFE to construct a neural signal recording utilizing off-the-shelf components. This structure is used with on-chip AC-coupled capacitor in the literature. However, in these designs, large

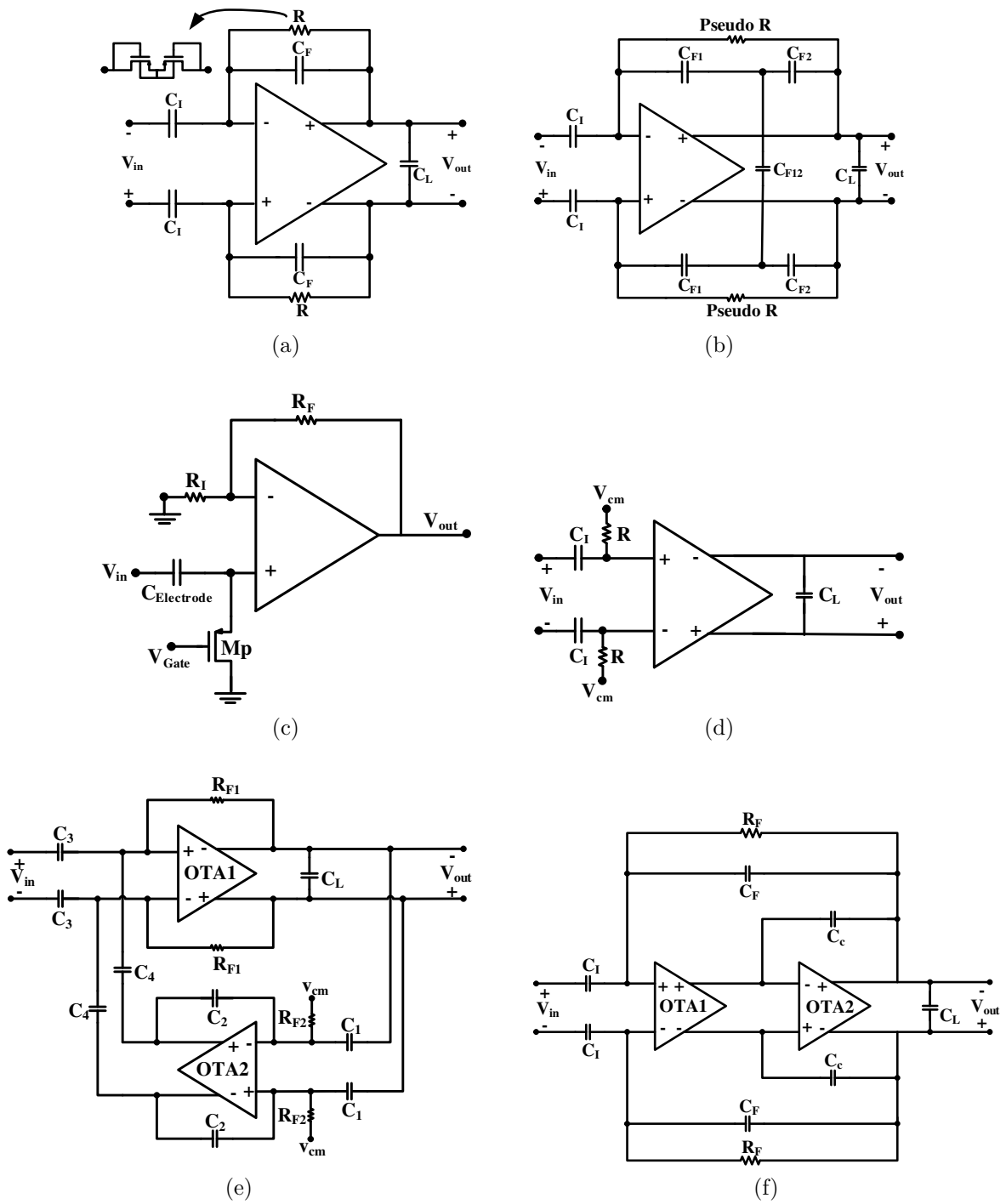


Figure 2.3 AC-coupled neural amplifier topologies. (a) Conventional capacitive-feedback network (CFN) topology. (b) CFN amplifier using T-capacitor feedback network topology. (c) AC-coupling utilizing the electrode capacitance and a resistive element. (d) Open loop network (OLN) topology. (e) Capacitive amplifier feedback network (CAFN) topology. (f) Miller compensated CFN (MCCFN) topology.

off-chip resistors or active on-chip resistors implemented with MOS transistors biased in the subthreshold region are used which increase the IRN.

Figure 2.3(d) shows open loop network (OLN) topology. It is similar to that of Figure 2.3(c) while utilizing an open loop amplifier [48]. Therefore this topology in addition to have the problems of the previous architecture, suffers from non-accurate gain and also sensitivity to the technology process deviation.

Capacitive amplifier feedback network (CAFN) topology is another topology that removes the tissue voltage offset using a coupling capacitor shown in Figure2.3(e) [49]. This structure is a band-pass amplifier where its midband gain (A_m) is calculated by $\frac{C_2C_3}{C_1C_4}$ and its low-cutoff frequency (f_L) is calculated by $\frac{C_2}{R_{f1}C_1C_4}$. As explained in [49] and [50], C_3 has to be maximized to decrease the input referred noise. For a given gain, the ratio of C_3/C_4 has to increase and the ratio of C_2/C_1 has to decrease. By increasing C_3/C_4 and decreasing C_2/C_1, the input referred noise gets close to the CFN topology of Figure2.3(a). As a conclusion, the added complexity to this structure does not significantly improve its parameters. Even in some cases with the same condition, the noise, power consumption and chip area of this neural amplifier is deteriorated compared to the CFN structure of Figure 2.3(a).

The topology shown in Figure 2.3(f) is Miller compensated CFN (MCCFN) topology which is similar to the conventional CFN topology, but it uses two OTAs in series. In some implementations such as in [51, 52], no Miller compensation capacitors are utilized in the OTAs. However, in other implementations such as [53], a Miller compensation capacitor is used in the second OTA (OTA_2) to make the non dominant pole far away for higher stability. In addition, in the design presented in [50] a Miller compensation capacitor is used in OTA_1 to create a double pole in the low-cutoff frequency. In design of a single stage neural amplifier with high gain, this topology can provide a fair trade off between output swing, DC gain, noise, and power consumption. This advantage is realized by utilizing the OTA_2 in this structure which increases the open loop gain. In addition, by designing the OTA_2 as a high-swing OTA in this structure, we can increase the output swing. Designing neural amplifiers with this topology (MCCFN) can result in low power and low noise with low noise efficiency factor (NEF) compared to the other topologies. However, in advanced technologies where the neural amplifiers are designed in multistage architecture, there is no significant advantage compared to the conventional CFN topology. Note that high-cutoff frequency in all AC-coupled topologies of Figure 2.3 is determined by the frequency response of OTA_1.

DC-coupled neural Amplifiers

The schematic shown in Figure 2.4(a) is the first neural amplifier that uses a LPF in the feedback path which is based on the block diagram of Figure 2.2 [54]. The high-cutoff frequency of this architecture is also realized by the frequency response of OTA_1 and its midband gain is obtained from the DC gain of this OTA. Due to this feature, no large capacitor ratio is required to obtain high midband gains. However, the midband gain is affected by strong process variations.

Utilizing an integrator as an active LPF in the feedback path results in lower area since a smaller capacitor is required due to the Miller effect compared to the structures that utilize passive LPF. However, to reduce the IRN, the capacitors C_I and C_L should be increased which results in additional area. In addition, this topology consumes huge amount of power in OTA_2 (feedback path) which in turn reduces the NEF in this topology.

A single ended configuration is utilized in Figure 2.4(a). This topology has a lower CMRR and PSRR compared to a fully differential configuration, however, it occupies lower area since it has only one miller capacitor. Moreover, the resistance of the pseudoresistor shown in Figure 2.4(a) varies due to the high voltage swing at the output and consequently alters the high-pass pole. The variation in the open-loop gain of the OTA_1 can also change the high-pass pole.

A DC-coupled neural amplifier topology similar to Figure 2.4(a) is shown in Figure 2.4(b) [55]. Unlike Figure 2.4(a), the architecture in Figure 2.4(b) utilizes a fully deferential architecture and a passive LPF. Since this architecture does not exploit the miller effect, a huge feedback capacitor is used to obtain a high-pass pole at low frequencies. Passive off-chip elements are used in the architecture of [55] that is not suitable for multichannel and implantable applications. Since no OTA is utilized in the feedback path, the noise and NEF are reduced at the cost of huge feedback capacitor. Note that the feedback capacitor is implemented off chip due to the high area.

Figure 2.4(c) shows a neural amplifier that utilizes the block diagram of Figure 2.2 as well to remove the DC offset of the input signal [16]. However, in this method the currents are compared in the input instead of voltages unlike the two previous topologies. This figure shows an AC-coupled chopper stabilized instrumentation amplifier (ACCIA) which utilizes current balancing instrumentation amplifier (CBIA) block. The coarse-fine servoloop is implemented using a coarse transconductance (CGM), a fine transconductance (FGM), and an integrator as shown in Figure 2.4(c). In [16], the authors have minimized the output range of analog fine servo by designing the coarse servo digitally. By this technique, the

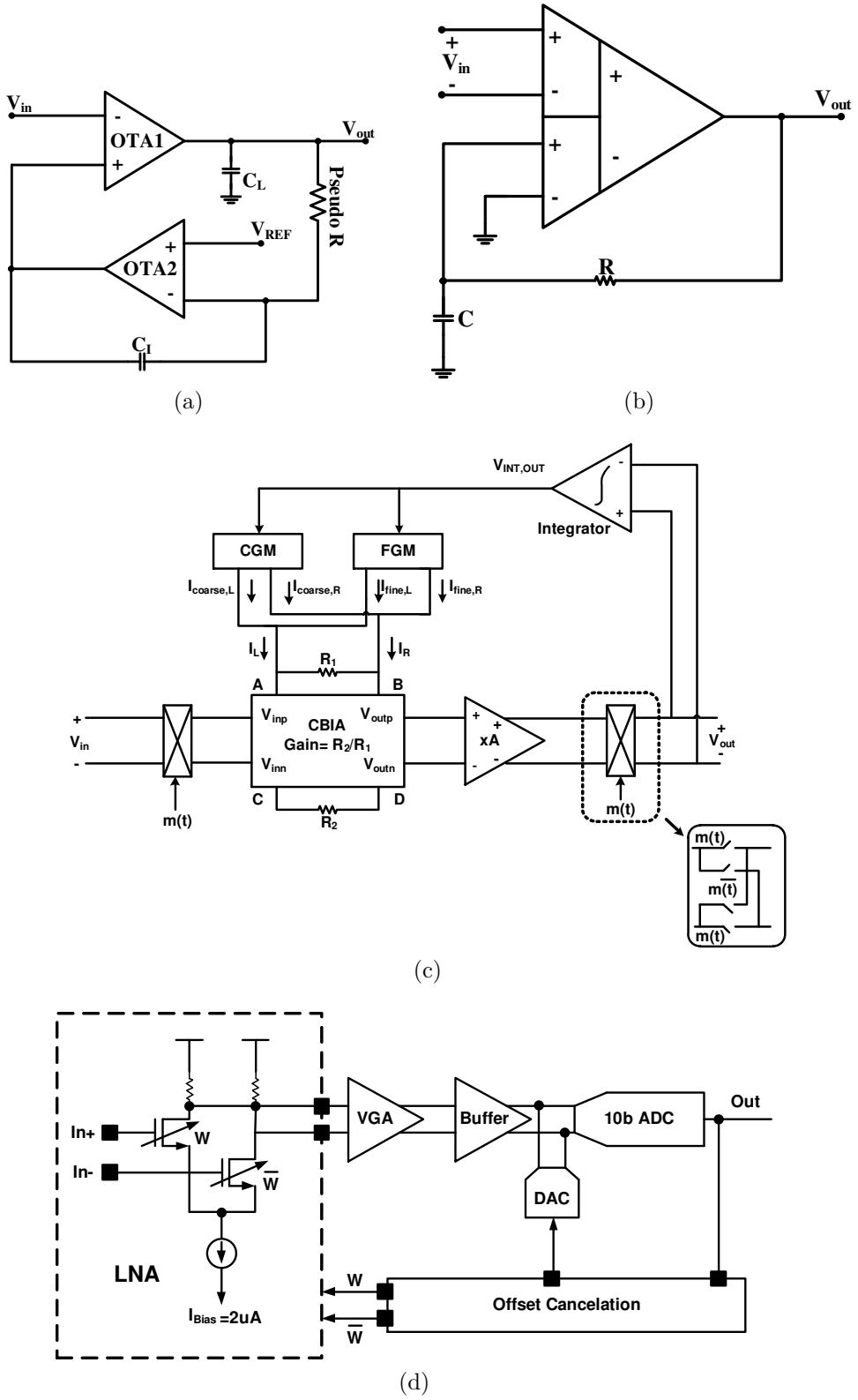


Figure 2.4 DC-coupled neural amplifiers utilizing (a) an analog active LPF. (b) an analog passive LPF and a differential difference amplifier, (c) analog and digital DC servo loops (DSL), and (d) input differential pair width modulation by an offset cancellation feedback.

power consumption of the fine servo is reduced and the power-noise performance of the ACCIA is improved compared to the previous ACCIA reported by the same group in [56]. This design also uses a chopper stabilization (CHS) technique to reduce the noise. However, this design still consumes large area and high-power consumption which is not attractive for multi-channel applications.

The idea of utilizing a digital LPF instead of analog in the feedback path of Figure 2.2 in order to avoid utilizing huge capacitors and high-power OTA is proposed in [57, 58] and shown in Figure 2.4(d). In this method, the low cutoff frequency can be determined more accurately. Also, it can result in less area and lower power consumption in advanced technologies and the comparison between the input signal and the feedback signal is performed by utilizing a DAC and an array of transistors in differential pair. This structure modulates and changes the width of the transistor based on the offset voltage to maintain constant IRN and CMRR. Nevertheless, measurment results show that these parameters vary with the offset voltage variation. The digital LPF in [57] is designed off-chip. The off-chip implementation of the filter creates an undesired delay to the low-frequency signal path. This limits the order of the filter to first order and makes it difficult to stabilize the feedback loop. In [58], a digital LPF is implemented on chip for 4 channels to eliminate the delay issue. The low cut-off frequency of this design can be programmed down to 40 Hz.

The authors in [33] also exploit a digital on-chip filter to implement a 56 channels neural recording implant. In this design, comparing the input signal and the feedback signal is performed directly in the currents passing through the differential pair by a current streering I-DAC. This design uses a CHS technique to reduce the flicker noise, but due to the high power consumption of the blocks and relatively high bandwidth ($10\ Hz - 5\ KHz$) compared to f_C ($\simeq 100\ Hz$, flicker noise corner frequency), a high NEF ($= 7$) is observed. Also, in addition to the complexity of this design, the midband gain is very sensitive to the technology process deviation.

2.4.2 Multistage Amplifiers

In the literature, there are reports of single stage amplifier designs for biomedical implants that are mostly implemented as capacitive coupled with the microelectrode sensors [59–63]. Single-stage neural amplifiers face the following challenges. In the case of DC-coupled neural amplifiers that are generally designed as open loop structures with high gain, the midband gain is very sensitive to process variation. In addition, the gain variation causes undesired variation in the low-cutoff frequency (f_L). On the other hand, in the case of AC-coupled, very popular and commonly-used conventional CFN topology shown in Figure 2.3(a), has

the following drawbacks. The midband gain (A_M) of such amplifier is obtained by $\frac{C_I}{C_F}$. For a specific and fixed value of A_M, the minimum possible value is chosen for C_F to minimize the C_I and consequently the chip area. The C_F is usually chosen in the range of $100 - 200\ fF$ to be larger than the neighborhood parasitic capacitances. Therefore, to have a high gain, the C_I has to be increased A_M times greater than the C_F which results in huge chip area and very low input impedance. Also, the latter factor, itself, causes attenuation of the neural signal in the input of the amplifier which in turn reduces the total gain of the amplifier and the CMRR. Furthermore, in the scaled and advanced technologies, the transconductance (g_m) reduces due to short-channel effects of MOS transistors which creates challenges and difficulties in designing high gain OTAs.

To overcome the mentioned challenges, amplifiers are designed in two or three stages in the AFE of the neural recording systems. Utilizing AC-coupled multistage amplifiers, reduces the input capacitors' value that leads to reduction of the chip area at the cost of increasing the power consumption slightly. Few examples of such multistage amplifiers are provided in [64–68].

In order to have low IRN, the first stage is designed as a low noise amplifier (LNA). The other stages (i.e., the second and third stages) are designed so that provide enough gain and linearity. Also, tuning the gain as well as the f_L and f_H are carried out in the second or third stages. Although it is seen in [42] that f_L and f_H are tuned in the first stage of the neural amplifier, it is not desirable as it affects the IRN.

It is proven by *Isoperimetric Theorem* in mathematics that for two or three-stage AC-coupled amplifiers (assuming that the main area of the neural amplifier is consumed by the capacitors), the maximum total gain for a specific area or the minimum area for a specific total gain is obtained when the gains at all the stages are equal. However, further design considerations such as lowering the IRN, causes the first stage of the neural amplifier which is an LNA, to be designed with higher gain compared to the other stages [42, 69].

Another challenge seen in both single-stage and multistage amplifiers (in the second or third stage) is that the high output swing of the OTA varies the feedback resistance of the MOS pseudoresistor in the CFN topology of Figure 2.3(a). This variation increases the non-linearity of the amplifier and consequently increases the distortion and makes the high-pass pole frequency variable with time [70].

Figure 2.5(a) shows the solution of this problem that is exploited in [71–73]. In these designs, the front-end amplifiers are composed of two stages in which the second stage utilizes a source follower like in Figure 2.5(a) to provide a constant voltage on the gate-source terminals of the MOS pseudoresistors, while the output swing is high. This technique increases the linearity

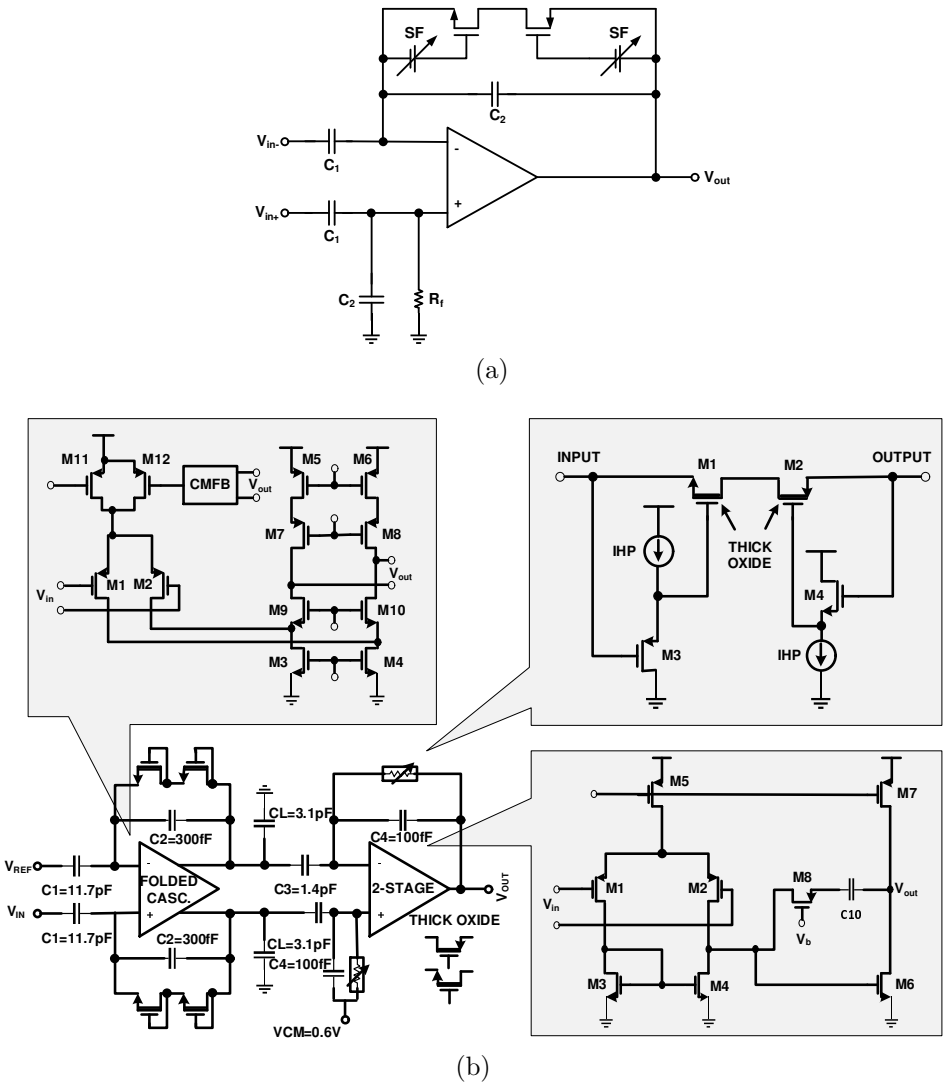
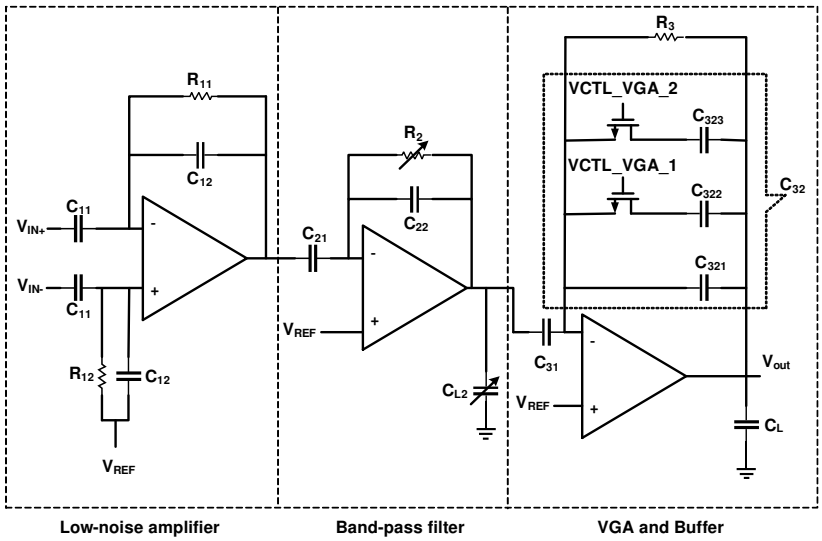


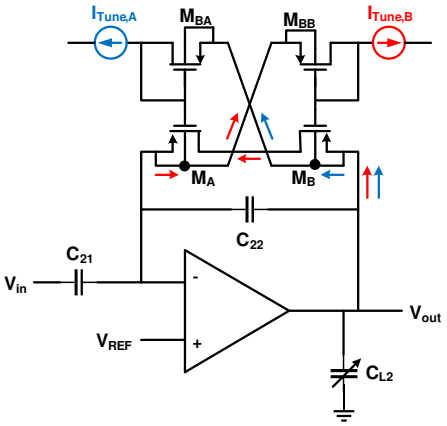
Figure 2.5 Improving the linearity of CFN amplifier utilizing (a) source-followers (SF). (b) two gain stages with employing the proper NMOS and PMOS pseudoresistor and SFs.

of the amplifier significantly.

Nowadays, proposed and designed front-end amplifiers have two or three stages. The design in Figure 2.5(b) utilizes of two stages [71, 72]. The gain of the first and second stages are as 39 and 14, respectively. To increase the linearity in the first stage, NMOS transistors are used as pseudoresistors and in the second stage as shown in this figure, NMOS and PMOS transistors accompany with a source follower are used. The pseudoresistors are all thick-oxide transistors to provide higher resistance compared to the standard CMOS. The high-pass pole frequency is adjusted by the bias current of the source followers in the second stage. This is while, the low-pass pole frequency is determined by C_L. This extra and large capacitor is used in [39, 41, 74] which in all these designs causes an overhead in the consuming area.



(a)



(b)

Figure 2.6 (a) Schematic of a three-stage amplifier, (b) Second-stage amplifier tunes the high and low-cutoff frequency.

The architecture shown in Figure 2.6(a) utilizes three stages for amplification [69]. The first stage is an LNA, the second stage is band-pass filter where adjusts the low and high-cutoff frequencies. The third stage, behaves as a variable gain amplifier (VGA) and a buffer. The gains of the first, second, and third stages are 50, 2, and 5, respectively. As illustrated in Figure 2.6(b), the second stage utilizes a current-controlled pseudoresistor, that uses a cross-coupled architecture to provide a symmetrical resistance with high linearity in the range of 0.2 V. This voltage is at the voltage swing range of the second stage. The low-cutoff frequency of these circuits are determined by the bias current of the current sources while altering the value of C_{L2} tunes the high-cutoff frequency.

2.4.3 Noise Reduction Techniques

As shown in Table 2.1, the amplitude of the neural signals are very low, therefore, To achieve a high SNR, the first stage of a multistage neural amplifier should be an LNA. Since the noise in the second and third stages of the amplifiers are divided by the squared gain of previous stages, the input referred noise of these stages are less important [75]. The noise efficiency factor (NEF) is a widely used figure of merit and is presented as [76]

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

where $V_{ni,rms}$ is the input referred rms noise voltage, I_{tot} is the total supply current of the amplifier, and BW is the amplifier bandwidth (in Hz). To compare LNAs with different IRN, power consumption and band width, NEF is utilized and the smaller NEF is the better.

To reduce the noise of LNAs, we explain circuit and systematic approaches in the following subsections. Both of these approaches are applicable on AC and DC-coupled neural amplifiers. IRN and NEF in different topologies are calculated and compared together in [50]. However, since the conventional CFN topology of Figure 2.3(a) is more appropriate and popular for multichannel neural recording compared to other topologies, we focus more on this topology in this chapter.

Circuit Techniques

The IRN value of an OTA varies based on its architecture. For example, a 2-stage OTA (or Miller OTA) can have a lower IRN compared to the folded cascode architecture due to the less number of transistors in the first stage. However, in general and in the same condition of bias current and transistor sizes, the differential pair transistors have the maximum contribution in the IRN value of various OTA architectures. Also the transistor of the tail current source (the current source that is connected to the differential pair) in all architectures as well as the cascode transistors in the telescopic or folded cascode architectures has the minimum contribution in the IRN value.

As presented in Table 2.1, neural signals have very low frequency components and small band width. For the frequencies higher than the corner frequency (f_c) in an LNA, the thermal noise is dominant and for the frequencies lower than (f_c), the flicker noise is dominant where both of these noises should be mitigated. For example, in [39] and [54] , (f_c) is reported as 100 and 300 Hz, respectively. In [77], the IRN of the thermal noise is calculated and tabulated for different OTA architectures. Equation (2.3), presents the flicker noise power of

a MOS transistor [78].

$$\overline{V_n^2} = \frac{K}{C_{ox}WL}.\frac{1}{f} \tag{2.3}$$

Where C_{ox} is the gate oxide capacitance per unit area, and K is a process dependent constant for a MOS transistor. W and L are the width and length of a MOS transistor, respectively. To reduce the flicker noise power of an OTA, the differential pair transistors should be of PMOS type where they have lower K compared to the NMOS transistors and according to (2.3) their sizes should be chosen large.

In addition, based on the IRN equations of the thermal and flicker noise in various OTA architectures, increasing the transconductance of the differential pair (g_m) compared to other transistors, reduces the IRN corresponding to both thermal and flicker noise. Therefore, by biasing the differential pair in the subthreshold region, the g_m of these transistors can be maximized. This can be achieved by increasing the W/L ratio of the differential pair for a constant bias current. Increasing the W/L ratio decreases the flicker noise of the differential pair as well.

As explained in the Section 2.4.1, for the CFN topology increasing the C_I based on (2.1), increases the midband gain of the neural amplifier and results in the reduction of the IRN. This is why the gain of the first stage (LNA) is usually designed significantly higher than the next stages. Also, for a given gain, increasing the width of the differential pair, decreases the flicker noise power of the OTA in (2.1) on one hand, and increases the C_{in} on the other hand. Therefore, as a trade off, there is an optimum point for the W where IRN is minimized. The low noise neural amplifier in [39, 69, 71, 79] have been designed considering the circuit noise reduction techniques explained in this section and without considering the systematic technique of the next section.

Systematic Technique

The neural signal contains important information in the low frequency range (i.e., less than f_c). In this frequency range, the flicker noise is dominant and to reduce its effect, the chopper stabilization (CHS) technique is usually utilized [80, 81]. The CHS technique operates as follows.

Utilizing up-modulation by the first chopper, CHS transposes the neural signal to a higher chopping frequency (f_{ch}) where the 1/f noise is not available. In the next step, the amplifier

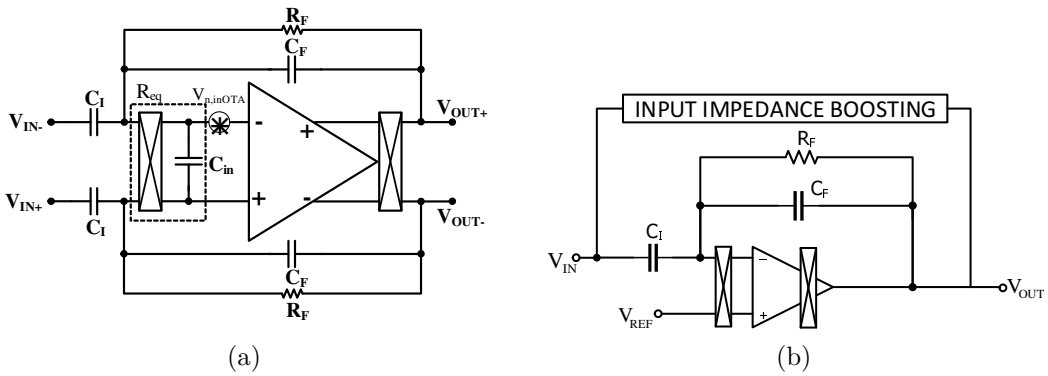


Figure 2.7 CHS technique in the CFN neural amplifiers. (a) Noise model of neural amplifier when the first chopper is placed in front of the OTA. (b) Utilizing an impedance boosting feedback circuit to increase the amplifier input impedance.

amplifies the signal while adding its 1/f noise and offset. The second chopping, modulates up the offset and 1/f noise to f_{ch} while the signal is demodulated back to the baseband. A LPF extracts the original signal with much higher SNR.

CHS can be used in both AC and DC-coupled amplifiers. Examples of utilizing this technique for DC-coupled amplifiers are shown in [16,33,56]. However, in this section, we explore more on the CNF topology of Figure 2.3(a). This is because this topology is more appropriate for multichannel neural recording and is more commonly used.

In practice, implementing CHS technique is challenging. The fact that the chopping switches are not ideal as well as the amplifier which has offset, limited gain and bandwidth, creates challenges in the design process. Utilizing the CHS technique in a designed LNA with the CFN topology of Figure 2.3(a) for the same capacitor size does not reduce the noise in low frequencies, but increases it. The reason behind this, is explained in [82] . Figure 2.7(a) shows the amplifier utilized in [82] with the input and output chopper switches and the IRN of OTA. The input chopper switches accompanied by the OTA input parasitic capacitance (like the switched capacitor circuits) can be modeled as a resistor which its value can be calculated as

$$R_{eq} = \frac{1}{f_{ch} C_{in}} \tag{2.4}$$

Where C_{in} is the OTA input parasitic capacitance and f_{ch} is the chopping frequency. The total noise of the neural amplifier when transferring the OTA noise to the input of the neural amplifier by considering the effect of the resistor and other capacitors can be presented as

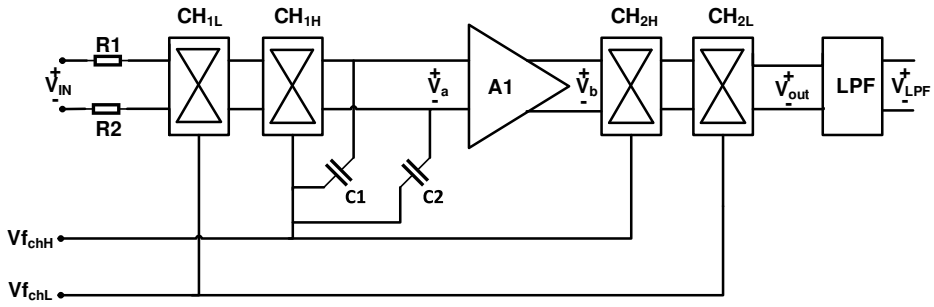
$$\overline{V_{in}^2} = \overline{V_{in,OTA}^2}\left(1 + \frac{C_F}{C_I} + \frac{2\pi f_{ch} C_{in}}{sC_I}\right)^2 \tag{2.5}$$

This equation shows that to reduce the total IRN of the amplifier requires a huge C_I capacitor. The capacitance value of the C_I in [82] and [83] are reported as 300 pF and 1 nF, respectively. These large capacitors increase the area and also decrease the input impedance of the neural amplifier. The latter factor reduces the CMRR and also increases the effect of mains interference on the system [84]. To improve the input impedance of this neural amplifier, an input impedance boosting circuit is utilized in [82,83] as shown in Figure 2.7(b). The operating of this circuit is so that a larger portion of the input capacitor current is provided by the output of the OTA through the positive feedback. This technique increases the input impedance in [82] from 400 MΩ to 2 GΩ at 1 Hz.

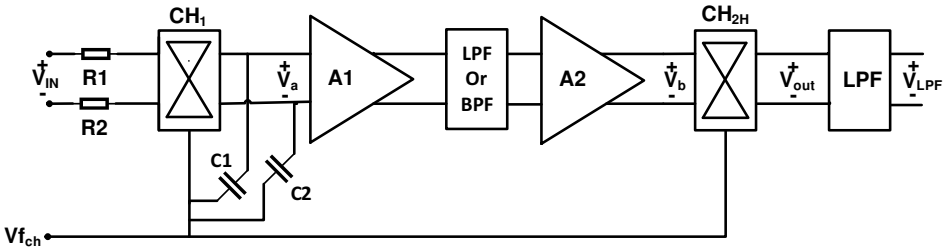
Two other challenges regarding to the chopper amplifiers are reducing the output ripple and their residual offsets [82, 85]. The offset voltage of the OTA causes the ripple. When the offset voltage gets amplified after up modulation, it can even have amplitude more than biopotential signals, therefore, it can limit the amplifier's output headroom. The amplifier designed in [85], senses the ripple in the output using a ripple reduction loop (RRL), and compensates the ripple by applying it to the input through a feedback. Note that since RRL is implemented in analog and works continuously, increases the power consumption significantly.

The residual output offset is mainly due to the non-idealities of the CMOS switches in the input chopper [80]. Clock feedthrough and charge injection creates spikes in the input of the OTA and they are amplified and presented in the output of the OTA. Then these spikes are down modulated with the output chopper and increase the DC at the output which is called the residual output offset. The authors in [83], have compensated the residual offset by embedding the circuit called DC servo loop (DSL) in the feedback path. DSL takes samples from the output and provides almost equal current, but in the opposite direction of the offset current by utilizing a $G_m - C$ filter and apply it to the input of the chopper. The main drawback of this circuit is that it requires a huge off-chip capacitor (greater than 10 μF) to decrease the low-cutoff frequency to the appropriate value.

To reduce the power consumption while removing the ripples in [85] and to avoid huge off-chip capacitor in [83], a new mechanism called digitally-assisted calibration loops is proposed in [82] that removes both the ripple and the offset. This mechanism provides a digital implementation of both RRL and DSL. There are also more examples in the literature that

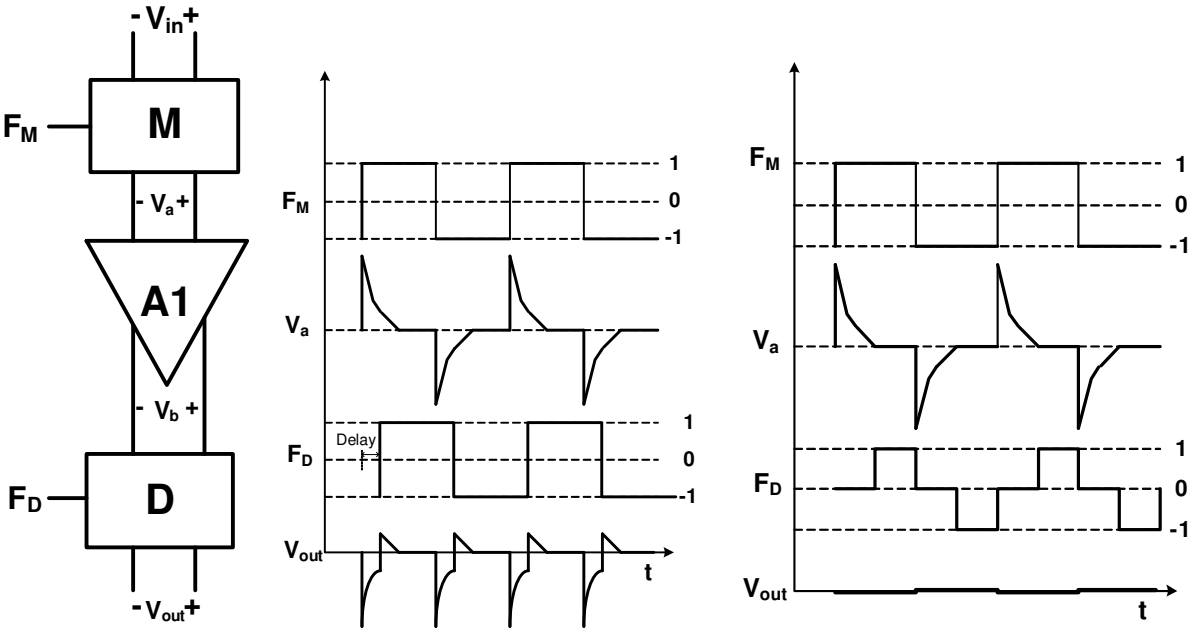


(a)



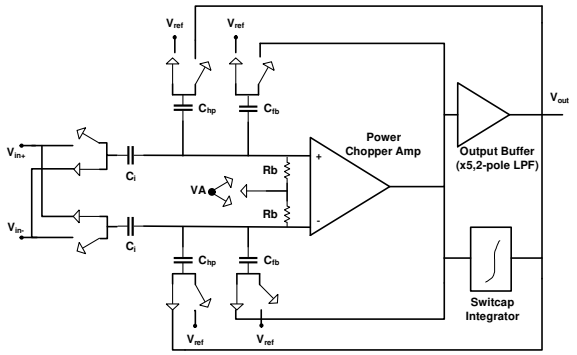
(b)

Figure 2.8 (a) Schematic of nested chopper amplifier technique. (b) Schematic of Spike filtering technique.

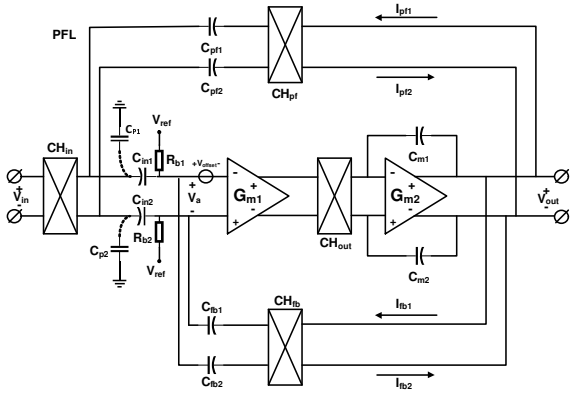


(a) (b) (c)

Figure 2.9 (a) Delayed modulation technique, (b) delayed demodulation clock diagram, (c) dead band clock diagram.



(a)



(b)

Figure 2.10 Chopper switches placed in front of the AC-coupling capacitors. (a) Circuit architecture of neural amplifier, illustrating two loop feedback paths to define the midband gain and high pass corner, (b) Schematic of the capacitively-coupled LNA with the positive feedback loop for input impedance boosting.

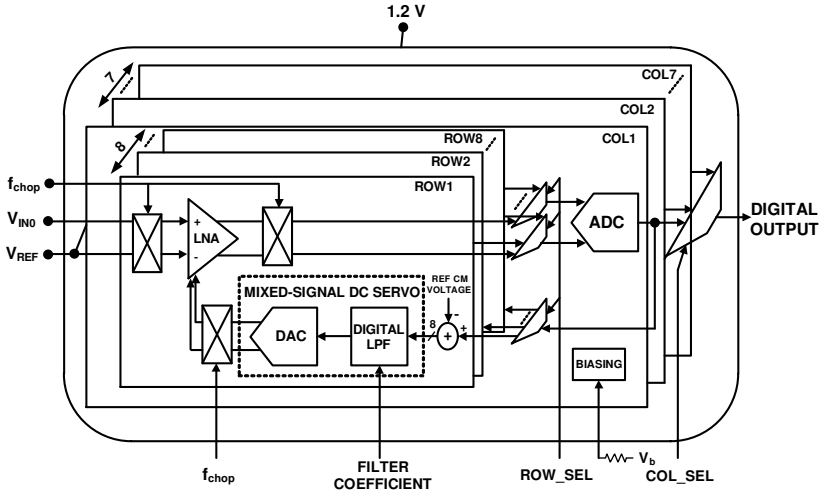


Figure 2.11 Block diagram of the 56-channel neural recording implant utilizing DC-coupled FEA and digital DSL.

utilize techniques to reduce the charge injection [86]. Three samples of these techniques which employed in the chopper amplifiers are as nested chopping, spike filtering, and the use of delayed modulation or dead band.

Figure 2.8(a) shows the schematic of the nested chopper amplifier [87–90]. As illustrated in the figure, this architecture utilizes a pair of internal and external chopper where the frequency of the internal chopper should be chosen greater than 1/f noise corner frequency (f_C). The frequency of the external chopper (f_{chL}) which is relatively a low frequency can be optimized for the input signal efficiently. With this technique, [87] has achieved 100 nV offset, with f_{chH} = 2 KHz and f_{chL} = 15.6 HZ.

Spike filtering technique is another approach for reducing the residual offset. As explained earlier, the charge injection in the MOS switches of the input chopper creates spikes in the output of the amplifier which eventually appear in the output of the second chopper as the residual offset in the frequency spectrum. Therefore, we can assume that these spikes have odd harmonics in the frequency domain [91]. By filtering these spikes, the residual offset can be reduced significantly. Figure 2.8(b) shows the schematic of this technique used in [92, 93].

Unlike the previous method where a filter is used to reduce the residual offset in the frequency domain, in the delayed modulation or dead band technique, the residual offset is mitigated in the time domain. Figure 2.9(a) shows the schematic of the delayed modulation technique. As shown in the figure, this technique utilizes a modulator M, amplifier A_1, and demodulator D. In [94], an amplifier is used followed by a LPF to shape the spikes and the second chopper with a constant delay is used (as shown in Figure 2.9(b)) to demodulate the spikes. This causes the average of the signal to be zero which is extracted by a LPF. Another method which is called dead-band or guard-band and its implementation is relatively simpler is shown in Figure 2.9(c). In this method, there is no modulation when there is spike in the output signal. This technique is used in [95, 96].

Another approach to implement CHS technique in the AC-coupled CFN topology as shown in Figure 2.10(a) is to use a chopper in front of the input coupling capacitors [66, 97]. In this figure, there are two feedback loop to determine the midband gain and high pass corner of the amplifier. However the main drawback of this architecture is its low input impedance which is calculated as

$$Z_{in} = \frac{1}{j\omega_{sig}(1 + f_{ch}/f_{sig})C_I} \tag{2.6}$$

where f_{ch} is the chopping frequency and f_{sig} is signal frequency. Increasing the f_{ch} causes a

more relaxed implementation of LPF on one hand and on the other hand reduces the input impedance according to (2.6). The input impedance is reported about 8 $M\Omega$ in [66]. In order to improve the low impedance of the design in [66,97], a positive feedback loop (PFL) shown in Figure 2.10(b) is proposed in [98]. This circuit takes samples from the output and accordingly provides a large portion of the current of the input coupling capacitors. The measurement results show that the input impedance increases from 6 MΩ to 30 MΩ. Note that, due to the presence of parasitic input capacitances, the boost factor and the original input-impedance are lower than the expected (100 and 8 MΩ, respectively).

Finally, at the end of this section, we analyze [33], as a sample of the DC-coupled neural amplifier. Figure 2.11 shows the block diagram of this amplifier which has 56 channels. In this architecture, a digital DSL is used in each channel to remove the input DC offset (caused by recording electrodes). In addition, CHS is used to reduce the $1/f$ noise of the LNA. Avoiding huge capacitors in the DC coupled architecture due to utilizing a digital DSL reduces the chip area. However, due to utilizing CHS at a much higher bandwidth ($10\ Hz - 5\ KHz$) than f_C ($\simeq 100\ Hz$, flicker noise corner frequency), it does not present a good noise reduction performance. The reported IRN and NEF are 5.4 μV_{rms} and 7, respectively, with a offset voltage of 50 mV.

2.4.4 Advanced Neural-signal Amplifiers

In this section, we compare and summarize the various mentioned topologies of neural amplifiers for multichannel neural recording application in Table 2.2. Also, we investigate the noise reduction techniques in advanced technologies. DC-coupled neural amplifier is not appropriate for large-scale recording application due to following drawbacks. The high gain value of this amplifier has much variation due to its open loop implementation. Also, it needs a huge capacitor in the feedback path where it implements a passive analog integrator and dissipates large power where it implements an active analog integrator. In case of digital implementation of the integrator for each channel, it consumes relatively large silicon area and power consumption. The best choice is to design the neural amplifier in two or even three gain stages with AC-coupled CFN topology to obtain the necessary gain in lower area. The gain of the first stage should be higher compared to the other stages to decrease the IRN of the whole amplifier. If achieving the required SNR in the output of the amplifier is possible by applying the circuit noise reduction techniques, it is not necessary to utilize the systematic techniques due to its silicon area and power overhead. Otherwise, the first stage of the neural amplifier as an LNA needs CHS technique to reduce the flicker noise of the LNA.

Table 2.2 COMPARISON BETWEEN DIFFERENT TOPOLOGIES OF NEURAL AMPLIFIERS

Amplifier Topology	**Figure**	**Ref.**	**Pros**	**Cons**
CFN	3(a)	[39]	Accurate gain, suitable for multistage amplifier	Large cap. area to obtain high gain
CFN with T-network	3(b)	[45]	Input and total cap. reduction	Low-cutoff frequency increase
Electrode cap. and resistive feedback	3(c)	[46]	No need to input cap.	Inaccurate and not adjustable high pass pole
OLN	3(d)	[48]	Small input cap. area	Inaccurate gain
CAFN	3(e)	[49], [50]	Smaller total cap. area compared to CFN	Higher power consumption and noise compared to CFN
MCCFN	3(f)	[51], [52]	Higher swing as a single stage high gain amplifier compared to CFN	Higher power consumption compared to CFN, Higher area consumption compared to multistage CFN
Analog Integrator	4(a)	[54]	Elimination of input cap.	Inaccurate gain and Low-cutoff frequency, Large power consumption
Differential difference amplifier	4(b)	[55]	Elimination of input cap.	Inaccurate gain and Low-cutoff frequency, large off-chip passive components
Analog & digital DSL	4(c)	[16], [56]	Relax analog DSL requirement due to digital DSL	Large area and power consumption
Differential pair width modulation	4(d)	[57], [58]	Fully-digital DC offset rejection	IRN and CMRR variation with input offset variation, complexity overhead
Fully-digital DSL	11	[33]	Fully-digital DC offset rejection	Inaccurate gain, high power consumption and NEF

Applying CHS to decrease the IRN needs to increase the input capacitors value which causes the input impedance reduction and consequently, CMRR reduction. Input impedance boosting circuit increases the input impedance by utilizing a positive feedback as mentioned in section 2.4.3. Furthermore, offset voltage of OTA causes ripple at the output of the amplifier which can be reduced by RRL circuit. Also, the non-idealities of the first chopper switches (charge injection and clock feedthrough) cause the residual offset which can be compensated

by DSL circuit in the feedback path. However, we presented three methods to decrease the residual offset at the origin in section 2.4.3. Among these methods, we suggest the nested chopper technique (Figure 2.8(a)), because of its simple implementation and high performance.

2.5 Analog to Digital Converters

The amplified neural signals are converted to digital by an ADC as shown in Figure 2.1, before transmitting out of the body. This is because, the digital data are so tolerant to the noise and other interferers compared to the analog signals. As the neural signals have very low bandwidth and require to consume low power for implant applications, successive approximation architecture analog to digital converters (SAR ADCs) are one of the best options. This ADC utilizes digital to analog converter (DAC) and successive approximation register (SAR) blocks. Most of the neural recording systems utilize an 8 to 10 bit resolution SAR ADC [99], however in [100], the resolution of the ADC is adaptively configured by the activity of the input neural signal, to save the power and compress the data. Figure 2.12 shows a SAR ADC which uses charge redistribution between the capacitors. Also, two sub DACs are utilized to decrease the total amount of capacitance and consequently reduces the total area [33, 71, 101]. This ADC is suitable for neural recording architectures shown in Figure 2.1(a) and Figure 2.1(c) which use one or several ADCs for all the channels. Due to the large area and high-power consumption, this ADC is not suitable for the architecture shown in Figure 2.1(b) where each channel uses one ADC. For this purpose, the authors in [102] propose an improved SAR ADC shown in Figure 2.13(a). The output bits of this ADC are extracted in serial. In addition, this structure exploits binary search algorithm as shown in Figure 2.13(b).

Although, in general, SAR ADC is a common architecture used in neural recording implants, the use of other ADC architecture is sometimes presented in the literature. For example, a logarithmic pipeline ADC with 8 bit resolution is used in [44]. The logarithmic encoding is used to represent the high dynamic range with a short word length. Another neural recording system with 256 channels is shown in [103]. In order to reduce the area and power consumption, an 8 bit resolution single slope ADC is used for each channel. The ramp generator and counter are shared between all channels in this architecture. The authors in [104] introduce a 10-bit resolution dual mode SAR and single-slope ADC for neural recording application. In the normal mode, the ADC works as a SAR ADC to quantize the extracellular action potential. In the compression mode, to reduce the dynamic power, the ADC is configured to single-slope and it processes just essential parts of spike waveforms.

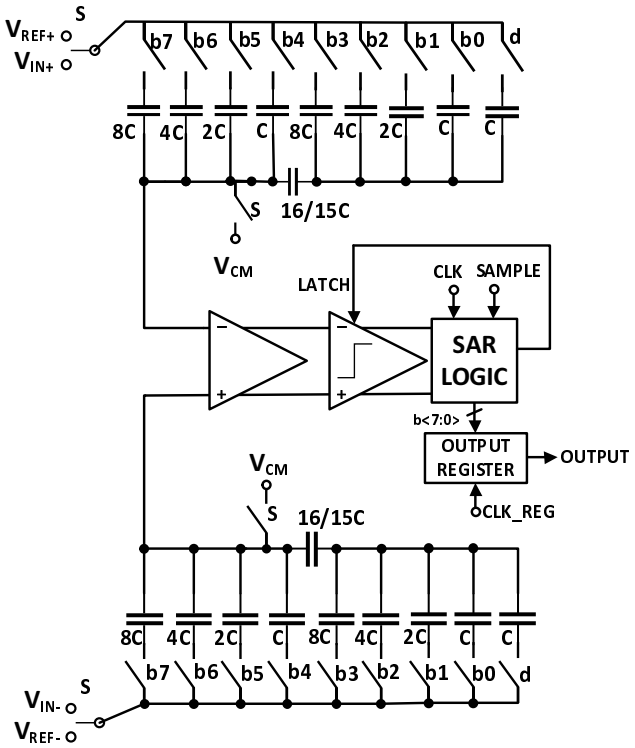


Figure 2.12 Differential 8-bit SAR ADC

Sigma-Delta modulators are also utilized in neural recording applications due to the low-frequency bandwidth of neural signals [105, 106]. Sigma-Delta modulators operate based on oversampling data conversion. Therefore, although this data converters can be low power, they can increase the power consumption in the subsequent circuits especially in the wireless transmitter due to increasing the output data rate of $\Sigma\Delta$ modulators much higher than Nyquist-rate ADCs. However, as mentioned earlier, the charge redistributed SAR ADC is the best data converter option for multi-channel neural recording implants.

2.6 Data Compression

Increasing the number of channels improves the spatial resolution on one hand, and on the other hand increases the output data transfer rate and power consumption specially in the wireless transmitter. One of the methods to decrease the power consumption is to decrease the data transfer rate by reducing the data redundancy which is called data compression. Since the encoders of the compression methods are implanted on the brain, the method that has a simpler encoder in terms of consuming less area and power is desirable.

As mentioned earlier in Section 2.3, CS is a compressing technique that efficiently acquires

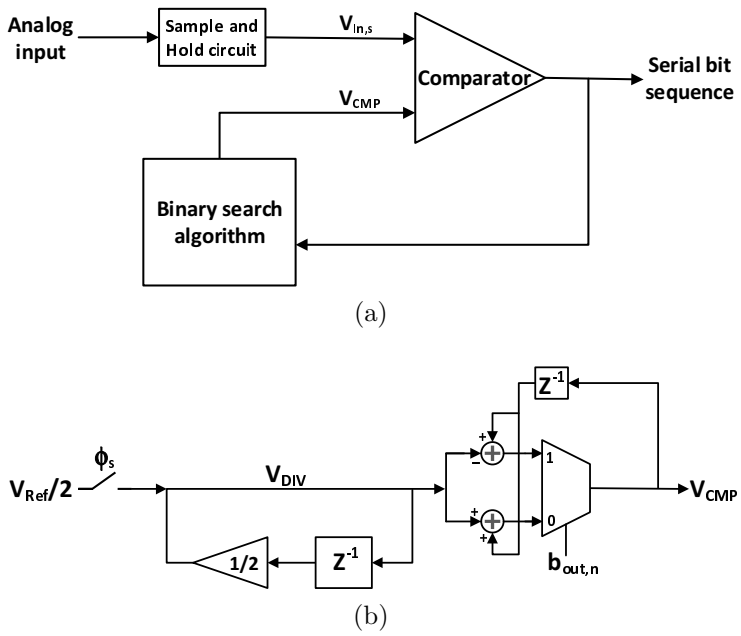


Figure 2.13 Modified SAR ADC for single-channel application. (a) Block diagram of the modified SAR ADC. (b) Schematic of binary search algorithm.

and reconstructs a bio-signal especially EEG and iEEG signals. CS is only used for bio-signals that are sparse in time or other domains. Luckily, most of the bio-signals are sparse in the time, Gabor or wavelet domains [107,108], which makes them suitable to use CS. Data reconstruction can be achieved with far less samples compared to Shannon-Nyquist sampling theorem by utilizing optimization methods and exploiting the sparsity of the signal.

The core of CS encodes an N-dimensional sampled input signal (X) into an M-dimensional sequence of measurement (Y) through a linear transformation by the $M \times N$ measurement matrix Φ, where $Y = \Phi X$. In this matrix equation, M is less than N ($M < N$) that represents the compressing of data from N sample to M sequence. There are infinite number of feasible solutions for X, as the equation is underdetermined. Assuming X is sparse, the sparsest solution is often the correct solution with high probability. As presented in [109], employing a random measurement matrix Φ, as a universal encoder as well as large enough input samples X is required to perform signal reconstruction of any sparse signal. A general approach to facilitate an efficient circuit implementation of Φ is to utilize pseudo-random Bernoulli matrix where each entry $\phi_{m,n}$ is ± 1 [110, 111].

As shown in Figure 2.14, CS encoder can be implemented in both analog and digital domain. Figure 2.14(a) demonstrates the block diagram of an analog implementation of CS encoder presented in [79,101,112,113]. The sparsity of EEG signal in Gabor domain is utilized in [101] and the design in [79, 112, 113] exploits the spatial sparsity of the iEEG signals recorded from

the electrodes of the sensor array. The CS core block diagram in digital domain is shown in Figure 2.14(b). The sparsity of the neural signal in Gabor domain is utilized in [114–119] and implement the CS encoder digitally.

Other signal compression method which is suitable for extracellular recordings is presented in the literature [32, 35]. This compression method is based on the sparsity of APs in the time domain. Since most of the information corresponding to the extracellular activities that are sensed and captured by the microelectrodes are in the APs, and the duty cycle of the APs are between 2% and 20%, therefore, it is sufficient to detect only the APs and transfer them out of the body [77]. As the waveform of the APs captured by different neurons in a microelectrode are different, we can exploit the features of these waveforms for the subsequent processes such as spike sorting. This waveform features can be such as the time of occurrence and the maximum and minimum amplitude value. Note that the best performance is achieved when complete waveform representations are available [120, 121].

An analog compression block is usually placed after the neural amplifier. Therefore, the neural signals have a relatively high voltage amplitude. One of the simplest method to detect and extract the location of the APs is to use a comparator in order to compare the neural signal with a threshold voltage [122, 123]. The APs are detected while the neural signal cross the voltage threshold. The implementation of the detection circuit is simple in both analog and digital circuits. However, in this method, we can only capture the portion of the AP that is above the threshold and the rest is neglected. Besides, accurate detection of this method is only feasible for high SNR [124]. To optimize the detection rate, the threshold voltage is chosen very carefully based on the level of the noise in the channel. This noise consists of background neural noise, flicker noise and thermal noise. The other similar method which is more effective in raising the detection rate is based on exploiting a bilateral threshold to consider both positive and negative signal polarities as presented in [32, 60, 125–127]. Saving a signal in a buffer and transmitting it with a very short delay allows capturing the whole waveshape of the APs without missing any portion of them. Employing an SRAM as a data buffering block and the bilateral threshold technique in [32] improves the accurate detection rate.

In order to improve the SNR and consequently increase the accurate detection rate of the APs, a pre-processor block is used in the designs in [128, 129]. In this method, on one hand the waveform of the neural activities is emphasized and on the other hand the noise is attenuated to increase the SNR. This method also helps to choose the threshold voltage easily to detect the APs accurately. In practice, the implementation methods of the pre-processor in the literature are different. For instance, in [128], a pre-processor detector and spike

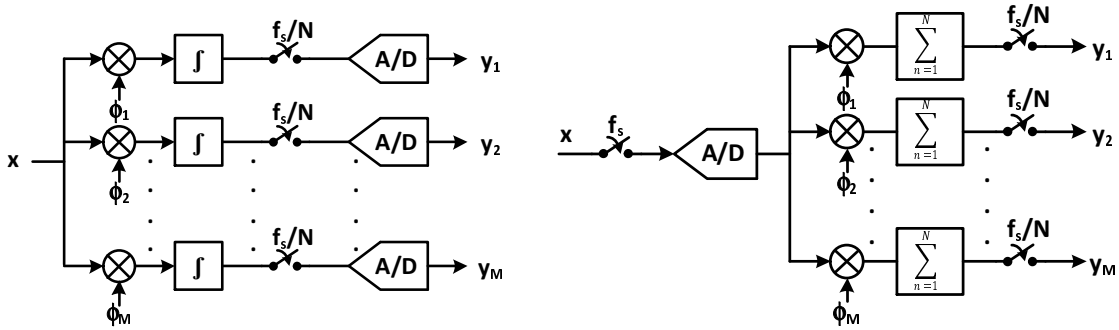


Figure 2.14 Block diagram of CS core of (a) an analog implantation (b) a digital implantation

sorting system is presented which operates based on variance of the neural signal. Another example are presented in [129], where an energy-based pre-processor is implemented utilizing low-power current-mode circuits.

Adaptive threshold is also another method to maximize the detection rate [130–132]. In this method, the threshold of detection is not a constant value and varies dynamically based on the SNR of the neural signal and the background noise. Moreover, Compression and other processes such as spike detection, feature extraction, and spike sorting can be implemented digitally in digital signal processors (DSPs). [133–136] provide samples that exploit DSPs to carry out such processes.

At the end of this section, it is necessary to mention that the application of these methods are based on the input signals. The CS method is the appropriate method for the EEG and iEEG signals due to its simplicity of implementation and no need to high SNR. However, for the extracellular neural activities which is important to extract the spikes for subsequent processes such as spike sorting, the data compression method based on the threshold is suitable and the adaptive threshold method is proposed.

2.7 Conclusion

In this chapter we explain briefly the necessary of the neural recording especially by the invasive method of implanting a chip on the brain in the skull. Also the neural signals and their electrical specifications are discussed. The require of large-scale channel recording as well as utilizing the advanced fabrication process have made new challenges in design of neural recording implants. The two most important parameter that should be considered in these designs are power consumption and chip area. We review the various architecture of

neural recording systems and conclude that the architecture of figure 2.1(c) is the best option for the very large-scale recording. Then we discuss on each blocks of these architectures. The most challenging block of a neural recording implant is neural amplifier. Therefore this block is elaborated in terms of designing a compact, high gain, low power, and low noise amplifier. We demonstrate several typologies for each of the both AC and DC-coupled neural amplifiers. Employing multistage amplifier to obtain high gain and lower the chip area for AC-coupled neural amplifier is described. Also the various techniques to reduce the noise of the neural amplifier is discussed in its subsection. Although SAR ADC as the best choice for the type of ADC in the neural implant application is presented, we demonstrate all the other ADC architectures that exist in the literature. At the end we cover the data compression methods to decrease the output data rate and power consumption.

CHAPTER 3 SYSTEMATIC DESIGN OF ANALOG COMPRESSIVE SENSING ENCODERS

3.1 Introduction

In Chapter 2 the necessity of capturing the brain signals is explained. It is also mentioned that wireless monitoring of brain activity through implantable devices is one of the best solutions that is possible. Increasing the number of recording sites on the brain (also called channels) is desirable and will increase the spatial resolution of the brain signals. Although the bandwidth of neural signals is relatively low, implementing of multi-channel neurorecording will increase dramatically the required bandwidth as well as the power consumption.

On the other hand, to implement a multi-channel implantable neurorecording chip, there are two crucial challenges, the power consumption and chip die area. They should be lowered as much as possible. Considering that most of the power in this system is consumed in the RF power amplifier (PA) of the transmitter [7], decreasing the amount of data that must be transmitted is desirable for diminishing the power consumption. This can be done by a data compression method. An appropriate method for doing this is compressive sensing (CS).

CS encoder decreases the sampling frequency to less than the Nyquist rate and has very simple circuit in both of analog and digital domain in comparison with other compression methods. Therefore, CS addresses both of the mentioned challenges (reducing chip area and power consumption), due to its simple implementation and data transmission rate reduction. The only condition that the signal must satisfy for CS is sparsity in some basis or domains like time, frequency, Wavelet and so on. Most bio-signals such as Extracellular APs, EMG, EKG, EEG, iEEG, LFP, etc., which are produced by bio-sensors, satisfy this condition and have sparse representation in time, Gabor or Wavelet domain [107, 108].

In this chapter we explore CS encoder in the analog domain, which is mainly implemented by switched-capacitor (SC) circuits. Non-ideal specifications of OTA in CS integrator such as finite gain, bandwidth, slew rate and output swing will decrease the total SNR. This chapter investigates the destructive effects of the mentioned non-idealities on the whole system SNR by Matlab and Simulink simulations [137].

3.2 CS Background and Neurorecording System

CS is a signal compressing technique which can reconstruct the original signal efficiently whereas it utilizes simple circuits in its encoder. This is achieved based on the principle that, through optimization, the sparsity of a signal can be exploited to recover it from far fewer samples than required by the Shannon-Nyquist sampling theorem.

The core of CS gets the N-dimensional sampled input signal (X), and encodes them into an M-dimensional sequence of measurements (Y), through a linear transformation by the $M \times N$ measurement matrix, Φ, where $Y = \Phi X$. As explained in section 2.6, the matrix equation is underdetermined and there are an infinite number of feasible solutions for X. If X is sparse, then the sparsest solution with the high probability is often the correct solution.

In some cases, X is not sparse (in time domain) but it can be further expanded by: $X = \Psi\alpha$ where α is the N-sample coefficient sequence and Ψ is the $N \times N$ sampling basis or sparsifying matrix. In the receiver, CS decoder utilizing an optimization method recovers the original signal X. A general method to find the sparse solution is to solve the following optimization problem:

$$\min_{x \in R^n} \|\alpha\|_{\ell_1} \quad subject\ to \quad Y = \Phi\Psi\alpha \tag{3.1}$$

The optimal solution to (3.1) extracts $\hat{\alpha}$ and then the recovered signal will be obtained by $\hat{X} = \Psi\hat{\alpha}$. The data compression factor (CF) is defined as (3.2). Increasing CF will decrease the radio power of the transmitter, but it will increase the error between the reconstructed signal $\hat{X}$ and the original signal X.

$$CF = \frac{N}{M} \tag{3.2}$$

In order to exploit the CS compression technique privilege, there are two crucial requirements. The first is sparsity which is related to input signal and has been mentioned in section 2.6. The second is incoherency between the measurement matrix, Φ, and the sampling matrix, Ψ, which ensures the maximal capturing of the input signal. Coherence between the measurement matrix (Φ), and the sampling matrix (Ψ) is defined as:

$$\mu(\Psi, \Phi) = \sqrt{n}. \max_{0 \leqslant j,k < n} |\langle \psi_j, \phi_k \rangle| \tag{3.3}$$

Where ϕ is a row of Φ and ψ is a column of Ψ. Actually, coherence measures the largest correlation between any row of Φ and any column of Ψ. Fortunately, random sensing matrices

Φ with sufficient samples behave low coherence with any fixed basis [109]. In practice, by exploiting pseudo-random sensing matrix instead of random one, it can be regenerated in the receiver. It is because, the sensing matrix is required in the receiver to recover the original signal. The sparsity is defined as 3.4.

$$Sparsity(\%) = \frac{N-K}{N}(100\%) \tag{3.4}$$

where K is the number of elements of X which are non-zero. In practice the small amounts of X which are near to zero set to zero. For more theoretical information of CS refer to [36–38]. The target is to design of a neurorecording implant for iEEG signal of the brain assuming multichannel implementation. Also, it is considered that the whole system will be implemented in 65 nm standard CMOS process and 1 V supply voltage. Figure 3.1 shows the block diagram of this system for n channels. In this figure, low noise amplifiers (LNA) amplify the captured neural signals by Bio-sensors and CS encoders compress them. An analog multiplexer multiplexes the signal of each channel to an ADC. At the end, these digital signals are transmitted out of the body by a transmitter.

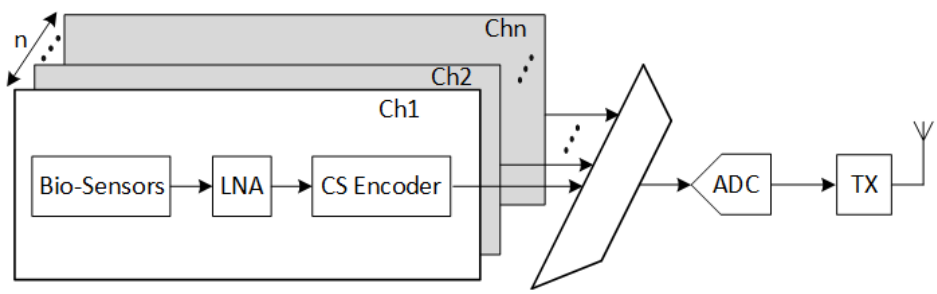


Figure 3.1 Block diagram of neural recording system with the CS in analog domain.

3.3 System Level Simulations

In this section, CS encoder of Figure 2.14(a) has been simulated by Matlab. It has been focused on the non-ideal effects of OTA such as finite gain, bandwidth, slew rate and output swing on the total SNR of reconstructing signal at the receiver. Figure 3.2 shows the Simulink simulation of Figure 2.14(a) in a vector of M elements. It is necessary to assume all the other elements of this system are ideal. Also, to minimize the SNR reduction of reconstruction signal due to lower sparsity of neural signal, we apply samples of a single tone (for instance 13 Hz) which is highly sparse in frequency domain. In this case, just the non-ideal effects of OTA will be considered.

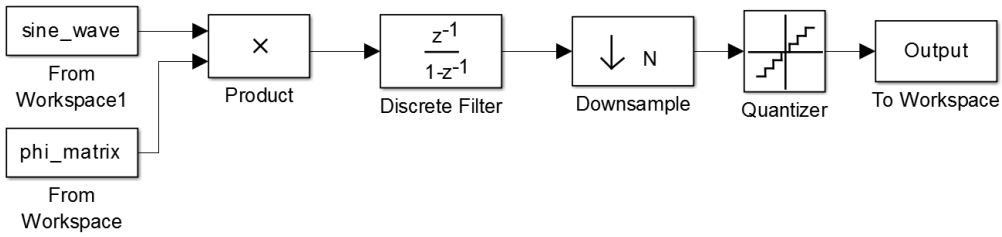


Figure 3.2 Simulink simulation of the CS Encoder.

In Figure 3.2, input signal, (X) is a discrete sine wave signal and the measurement matrix, Φ, has been generated by Matlab built in random function and has Bernoulli probability density function (PDF) with values of ± 1. They are multiplied in vector mode. A discrete ideal integrator accumulates the signal and a down sampler releases the last value of N sample integration. The Number of the input samples, N, and CS output sequences, M, are considered 1000 and 40, respectively.

The CF which is defined as (3.2) is 25. Taking fast fourier transform (FFT) from 1000 input samples and applying the result in (3.4) calculates the sparsity as 99.8 % in frequency domain. Finally, a 15-bit quantizer has been utilized at the end as an ADC. The simulation results of this system achieved 89.7 dB for reconstructed signal SNR which has been restricted by the quantizer.

3.3.1 Finite Gain Effect

Implementing CS encoder in discrete-time domain exploits switched-capacitor (SC) integrator circuits. The finite gain of OTA is one of the non-ideality effects of SC integrators which has destructive effects on the total SNR of system. In this case, it is considered that the integrator has leakage and the transfer function of this leaky integrator changes to [138]:

$$H(z) = \frac{V_o(z)}{V_i(z)} \approx \frac{z^{-1}}{1-(1-\varepsilon)z^{-1}} \tag{3.5}$$

where ε is referred to the leakage factor. In the conventional SC integrator, the leakage factor is:

$$\varepsilon = \frac{C_I/C_F}{A} \tag{3.6}$$

where C_I and C_F are input and feedback capacitors of integrator, respectively, and A is the finite gain of the OTA. C_I/C_F is the gain of the integrator and has been normalized to 1 in

our simulations. (3.5) and (3.6) which show the effect of finite amplifier dc gain will shift the pole of H(z) slightly off of the unit circle. In practice, this leakage adds error in each integration phase and decreases the system SNR.

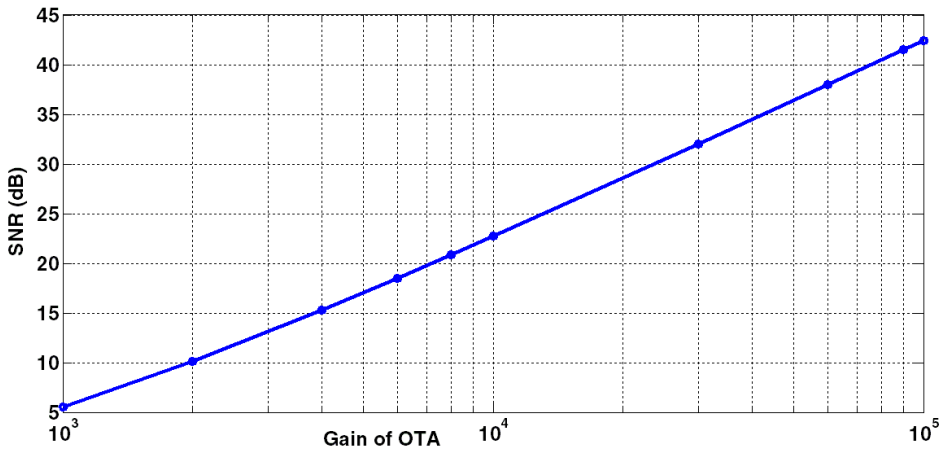


Figure 3.3 SNR of reconstructed signal versus the gain of the OTA.

Inserting leaky integrator instead of the ideal one in the system of Figure 3.2 and iterating the simulation with various gains lead to the curve of Figure 3.3. In order to ignoring the quantization noise, the output sequence was extracted before quantizer. Figure 3.3 illustrates the SNR of reconstructed signal versus the gain of the OTA. The curve has a logarithmic form, so for simplicity of demonstration, X axis was scaled logarithmic. The slope of the curve is almost 18 dB/dec. The reconstruction SNR has almost ± 1 dB tolerance. This figure shows the performance of CS encoder is very sensitive to the gain of the OTA. For instance, for CS encoder with OTA gain of 10000 (80 dB), the SNR of reconstructed signal is 23 dB. This 80 dB gain for OTA is relatively high and is very hard to implement in 65 nm technology.

3.3.2 Finite Bandwidth and Slew Rate Effect

In this section, the non-ideality effect of finite speed in the SC integrator is investigated. The small signal and large signal speed performance of an integrator is translated to Linear settling and slew rate respectively. In other words, if the integrator was not rapid enough, it could not settle to the final expected voltage in each integration phase. Therefore, it injects error in each integration phase and accumulation of these errors leads to SNR reduction. In order to explore the effect of this error solely, all the other error sources in the system of Figure 3.2 should be eliminated. For this purpose, a function block is inserted in the Simulink simulation of Figure 3.2. This function calculates the final expected voltage of the integrator based on input voltage, assumed slew rate and single pole time constant of OTA [139]. The

function equation is:

$$V_o = \begin{cases} V_i(1 - e^{\frac{-1}{2f_s\tau}}) & |V_i| \leqslant \tau.\zeta \\ V_i - sign(V_i).\tau.\zeta.e^{[\frac{|V_i|}{\tau.\zeta} - \frac{1}{2f_s\tau} - 1]} & \tau.\zeta < V_i \leqslant (\tau + \frac{1}{2f_s}).\zeta \\ sign(V_i).\zeta.\frac{1}{2f_s} & (\tau + \frac{1}{2f_s}).\zeta < |V_i| \end{cases} \tag{3.7}$$

where ζ is slew rate, τ is single-pole time constant, f_s is sampling frequency and V_i is input samples voltages. The function is inserted before the ideal integrator of Figure 3.2 and simulated with various amount of time constant and slew rate. For simplicity of demonstration and analysis of the results, time constant τ and slew rate ζ is normalized to:

$$n_\tau = \frac{1}{2f_s\tau} \tag{3.8}$$

$$\zeta_n = \frac{\zeta}{2f_s V_{Maxi}} \tag{3.9}$$

where V_{Maxi} is an assumed maximum input voltage and n_τ and ζ_n are number of single-pole time constants and normalized slew rate respectively. n_τ means the integration phase of the integrator (half of sampling time) consists of n_τ number of time constants. In this simulation, the output sequence has been extracted before quantizer again.

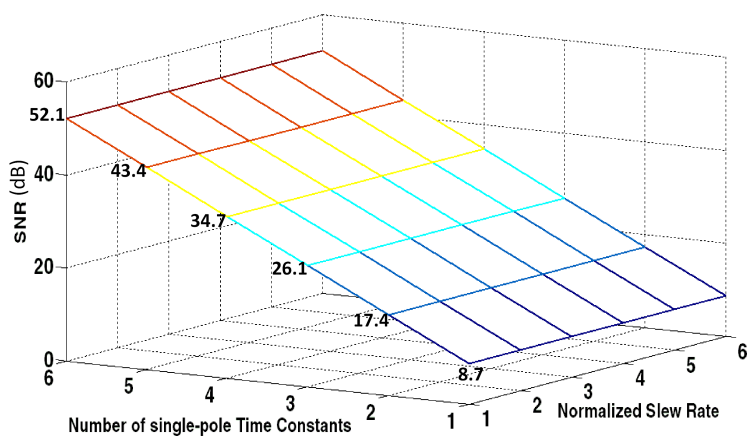


Figure 3.4 Total SNR as a function of slew rate and settling time.

Figure 3.4 shows the result of Matlab simulation considering $V_{Maxi} = 0.5$ V. It illustrates the total SNR versus normalized slew rate and number of time constants. The figure demonstrates the CS encoder performance is sensitive to the number of time constants not to slew rate of the OTA. It is because the small signal effect is dominant. As the SNR does not

Table 3.1 FINITE BANDWIDTH SIMULATION RESULT

n_τ	1	2	3	4	5	6
SNR(dB)	8.7	17.4	26.1	34.7	43.4	52.1

change by changing slew rate in Figure 3.4, Table 3.1 has tabulated the Figure 3.4 for further analysis.

3.3.3 Finite Output Swing Effect

In the simulation of this section, all the error sources are eliminated again, except the output swing of OTA in the integrator. For this purpose, a saturation block is inserted in the integrator of the Simulink simulation of Figure 2.14(a) to limit the output voltage between -1 and +1 V. It should be mentioned that the signals in Simulink schematic of Figure 3.2 are a vector of M (=40) elements which we call each element of the vector as a branch. Therefore, in practice, there are 40 parallel integrators which work concurrently.

Simulations show that as long as the outputs of integrators are not saturated, the total SNR is 106 dB, but as soon as the outputs are saturated, SNR is decreased dramatically. The amount of SNR depends on the number of branches out of M that are saturated and the saturation duration in each branch. The total SNR can be varied between 8 dB and 32 dB. In the best case, when the saturation is happened in just one integrator out of M, and the amount of saturation is 0.8 %, the output SNR is achieved as 32 dB.

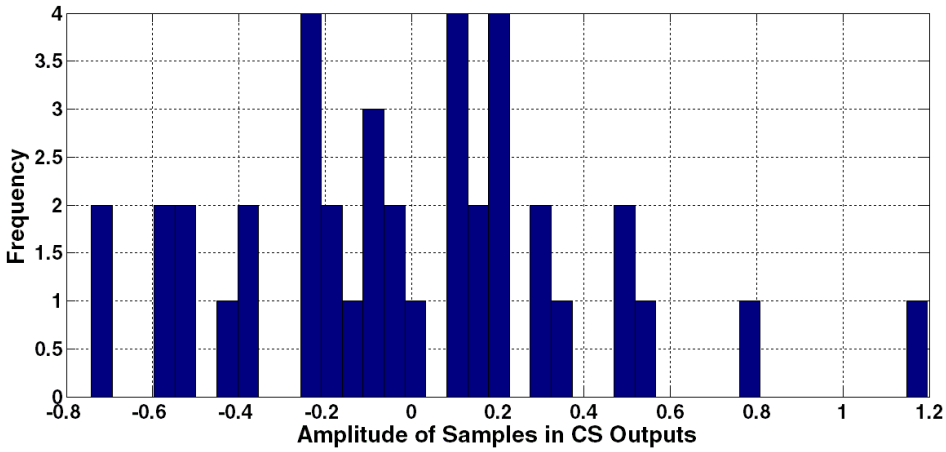


Figure 3.5 Histogram of M samples of CS outputs (M=40).

Like the output swing of integrators, the dynamic range of ADC is limited between -1 and +1 V. It means the input voltage of ADC should be in the range $\pm V_{ref}$. The same saturation

block is inserted before the ADC block and the simulation is iterated assuming that the other blocks are ideal. The result is very similar to the effects of integrator outputs finite swing. Figure 3.5 shows the histogram of 40 output samples of 40 parallel branches (M) of CS encoder. Increasing the amplitude of one sample out of 40 samples from the boundaries ($\pm 1\ V$), decreases the reconstruction SNR to 23 dB.

The results of these simulations show the CS encoder is very sensitive to finite output swing of integrators and they should be avoided from saturating. For this purpose, the gain of the integrators should be controlled to avoid saturation or use other methods such as "smart saturation checking" presented in [101].

3.4 CS Core with iEEG Input

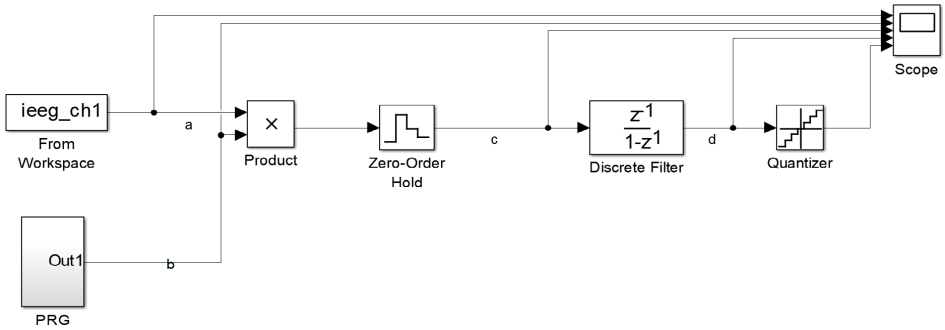


Figure 3.6 Simulink simulation of CS core block diagram in MATLAB.

In this section, system level simulation is investigated to find out the required total voltage gain. The question is how much the CS core increases the amplitude of the input signal. This increased value of the amplitude is called CS gain in this thesis. Figure 3.6 shows the Simulink simulation of Matlab. The input is a real iEEG signal sampled from a dog captured in 16 channels. The sample data is available freely at https://www.ieeg.org/ web site. The sampling frequency of the iEEG signal is 399.6098 Hz. This input is like a random signal which is integrated after multiplication of this signal by ± 1 randomly in order to create a linear combination from input samples. Matlab simulation will help us to calculate the gain of CS core.

Figure 3.7 shows the various signal waveforms on the scope of Figure 3.6 (The figure shows almost 0.75 sec of the waveform). These signals are input iEEG, pseudo-random generator (PRG), multiplication and integration of these signals. We reset the integrator after a long sequence of signal integration. In other words, increasing the number of samples before resetting will increase the amplitude of the CS output voltage. The important point is

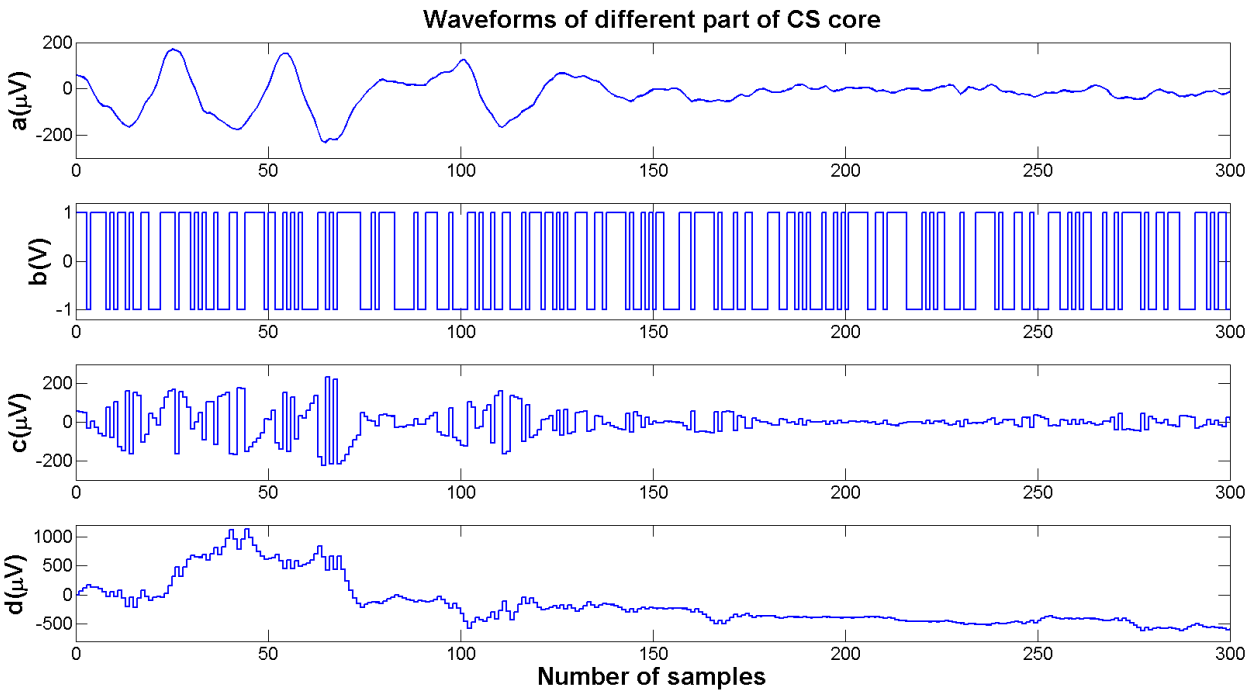


Figure 3.7 Waveforms of Simulink Scope of Figure 3.6: (a) The original iEEG signal, (b) Random ±1 generated by Pseudo-Random Generator (PRG), (c) Multiplication result of signal a and b, and (d) Integration of signal c.

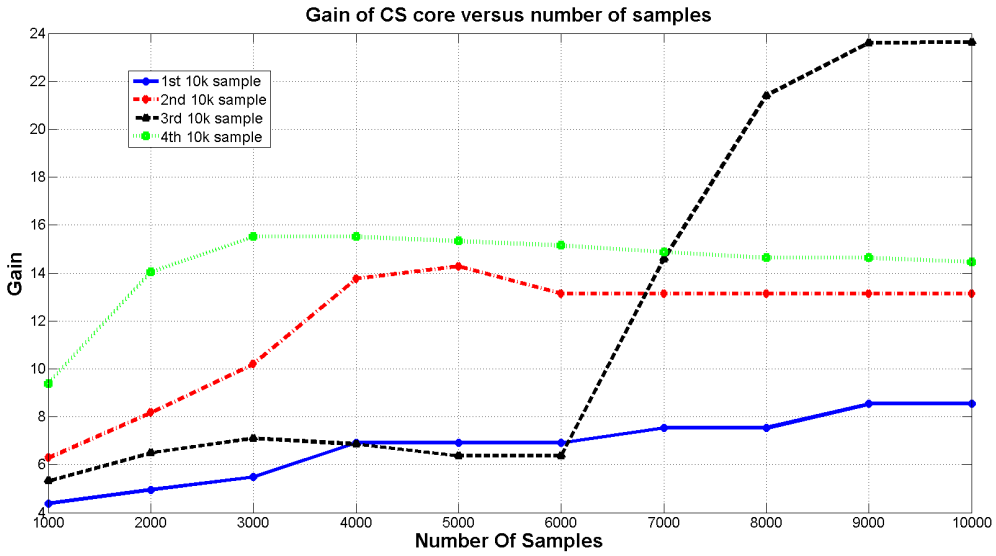


Figure 3.8 Gain of CS core versus the number of sequences.

that the amplitude increase varies a lot so that it could saturate the integrator. Figure 3.8 illustrates the gain of CS core versus the number of samples for 4 sets of 10000 samples, before

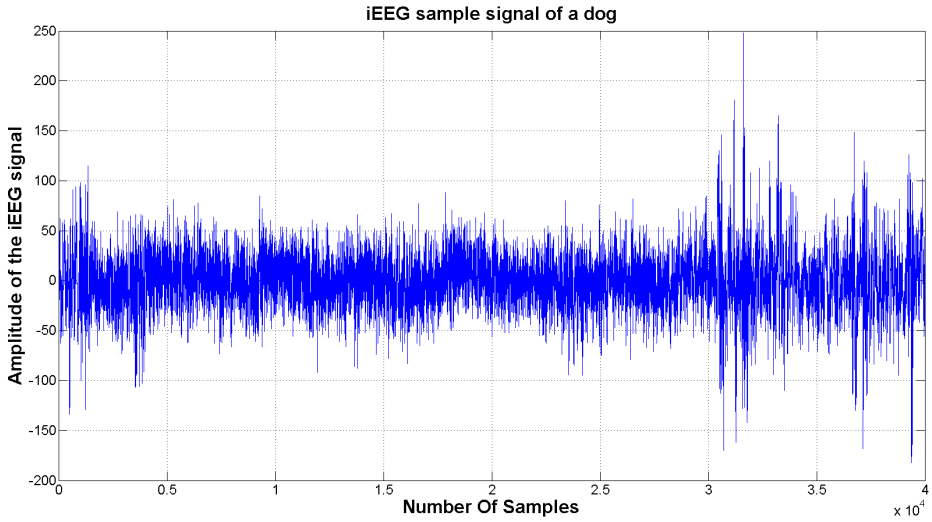


Figure 3.9 iEEG signal of a dog.

resetting the integrator. The whole 40000 samples of iEEG signal is shown in Figure 3.9. Based on the Figure 3.8, the gain of CS stage for 1000 sequence varies almost from 4 up to 10 and for 2000 samples varies between 5 and 14. Therefore resetting the integrator of CS after 1000 up to 2000 sequences is suitable. Utilizing a PGA can configure the total gain to avoid saturating the amplifier and achieve the maximum of dynamic for ADC.

3.5 Variable Compression

CS core implementing with series sequences of Figure 3.10 is another solution for analog implementation [110, 140]. In this architecture, M output sequences are down sampled sequentially from N input samples. The integrator is reset at the end of N samples integration. The measurement matrix, Φ, in this case is similar to the lower triangular matrix. The quality of reconstruction is lower than the CS core of Figure 2.14, due to the higher coherency between Φ and Ψ matrix. In other words, the reconstruction SNR of series-sequences CS core in Figure 3.10 is lower than the parallel-sequences CS core in Figure 2.14, especially when the sparsity is low. However, it requires less hardware and consumes lower power than the parallel-sequences CS core.

Altering the CF in this case is also easier due to changing M and consequently down sampling frequency. To exploit the privileges of both CS core, we propose to combine these architectures to achieve the CS core of Figure 3.11. In this architecture, we have two kinds of parallel and sequential output and the number of these sequences are based on the value of M_P and M_S. In this case the CF is defined as (3.10). The value of M_P is fixed and determine the

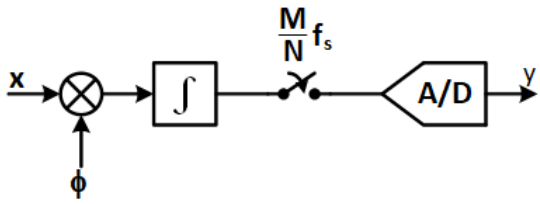


Figure 3.10 The CS core with series sequences.

number of the branches of integration in Figure 3.11. However, M_S is configurable to alter the CF and the output SNR.

$$CF = \frac{N}{M_P.M_S} \tag{3.10}$$

This architecture consumes lower power and silicon area compared to the parallel-sequences CS core and improve the output SNR in comparison with the series-sequences CS core. As future works, it is necessary to do some Matlab simulations to find the optimum value of N, M_P and M_S for the neural input signal in different CF. For instance, to compress the input samples with CF as 5, the typical value of N, M_P and M_S can be 1000, 5 and 40 respectively.

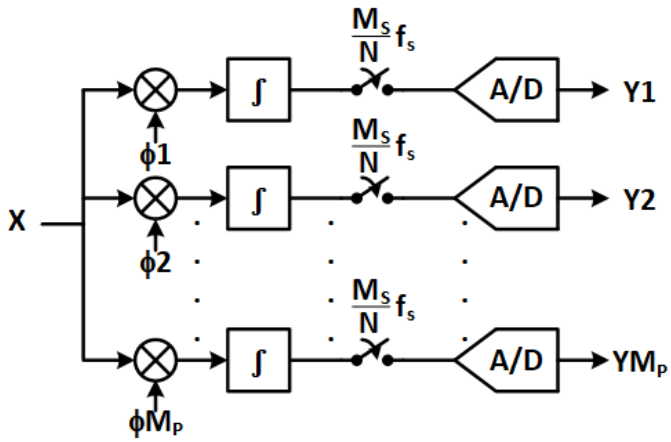


Figure 3.11 The CS core with parallel-series sequences.

3.6 Programable Gain CS Core

According to the Table 2.1, the maximum amplitude of iEEG signal is 1 mV. Considering that power supply and the maximum fully differential dynamic range at the input of ADC

is 1 V, the whole voltage gain of the LNA and CS core in Figure 3.1 should be at least 1000 (60 dB). Assuming the gain of the LNA is 50, the gain of CS core should be 20. Also, the Matalb simulations in section 3.4 show that the gain-normalized CS core has the gain of 4-14 for 1000-2000 samples of iEEG input signal. Therefore, the CS core which is implemented by SC-integrator should be designed conservatively to have the gain between 1.25 and 10 to cover wide range of input samples, N.

3.7 Discussion and Conclusion

In the literature, reconstructed signal in comparison with original signal has typical SNR value between 8 dB and 24 dB for neurorecording implant application. For instance, authors in [114] have reported CS implementations in the digital domain with total SNDR of 10 dB and [79] have implemented CS encoder in analog with SNR of 21.8 dB.

The simulation results of this chapter show that the CS encoder has the most important role in the SNR value compared to the other blocks shown in Figure 3.1. Based on the curve shown in Figure 3.3, it is not possible to achieve more than 23 dB SNR with maximum OTA gain of 10000, which is challenging to implement with a 65 nm CMOS process.

According to this conclusion, to achieve the SNR of 8-24 dB, it is proposed to design a relaxed LNA and ADC. LNAs are often large amplifiers that consume a lot of power and chip area to overcome the flicker noise. In addition, it is obvious that lowering the resolution of SAR ADC (which is the best choice for this application) has the same effects. As a sample design specification, an ADC of 6-bit resolution and CS encoder with OTA gain of 80 dB and $n_\tau = 5$ (from Figure 3.4), accompany with a much relaxed LNA with input referred noise (IRN) voltage of 16 μV_{rms}, will decrease the SNR to 21dB.

CHAPTER 4 NEURAL AMPLIFIER AND REFERENCE CIRCUIT DESIGN

This chapter consists of two sections. In the first section, we investigate the low-cutoff frequency reduction of neural amplifiers in advanced processes such as 65 nm, its causes and solutions. In the second section, we explain the design of a voltage reference for neural recording implant applications.

4.1 Neural Amplifier Design

4.1.1 Introduction

Neural signal acquisition has a crucial role in understanding the function of the different parts of the brain as well as exploring and treating its various disorders [141]. In addition, this data is used in developing the neural prostheses [142] and brain machine interfaces (BMI) [143]. This is why the demand for new techniques that enable monitoring brain activity wirelessly through implantable devices is increasing every day [2–4].

Based on Table 2.1, brain signals are very small and have very low bandwidth. For instance, the maximum amplitude of local field potentials (LFP) is typically 5 mV and the frequency range is less than 1 Hz up to 200 Hz [144]. On the other hand, the amplitude of the spikes or the neural action potentials (AP) are typically as high as 500 μV and their operational frequency is up to 10 kHz [145].

Increasing the number of the neural recording sites, which are called channels, is required in some applications, as the spatial resolution of the capturing signals increases. As an example, the total number of channels reported in [21] is 3072. The electrochemical reaction at the electrode-tissue interface in each channel generates different DC offset voltages across the various electrodes. These voltages vary typically between 1 mV and 10 mV and in some cases up to 50 mV [33]. As the offset voltages of the channels have high value, they can saturate the neural amplifier. Therefore, they should be eliminated. The most common approach to block this DC input offset is to utilize large AC-coupling capacitors [39, 45]. On the other hand, as explained in section 2.4.1, there is an alternative method that blocks these DC offset voltages by using a low-pass filter in the feedback path, which is called DC-coupled input offset rejection. The authors in [16, 57, 58, 146] use this method, however, it requires a huge capacitor or high power consumption amplifier in the feedback path.

To design multichannel neural amplifiers, the following factors should be considered and

diminished as much as possible.

1. Power consumption: the brain tissues that are surrounded by implantable neuro-amplifiers must be protected from heat damage. For this purpose, the power dissipation of these amplifiers must be lowered.

2. Chip area: The neural amplifiers are generally huge. This is because they usually utilize large AC-coupled input capacitors. Also, to decrease the flicker-noise power of amplifiers, the size of the MOS transistors is designed to be very large especially in the differential pairs. Therefore, for a specific chip area, to maximize the number of the channels, the amplifiers should be designed in their minimum area.

3. Noise: the neural signals have very low amplitude and bandwidth. The flicker and thermal noise of the neural amplifier circuit is the main source of the noise, which can decrease the signal to noise ratio (SNR) in the output of the amplifiers. This is why they are designed as a low noise amplifier (LNA). In the low frequency, the power of the flicker noise is dominant. To decrease the flicker-noise power, in addition to increasing the size of the transistors and utilizing a PMOS differential pair, the chopper-stabilization technique is used [66, 82, 83, 147]. The chopper-stabilization technique modulates the low-frequency noise of the OTA (flicker noise), as well as the offset voltage to a higher frequency by the chopper switches. These higher frequencies are eliminated with a low pass filter (LPF).

The 65 nm CMOS and finer technologies introduce new challenges as a result of the short channel effects for analog circuits. One of these challenges is decreasing the transconductance (g_m) of MOS transistors, which diminishes the voltage gain of the whole amplifier. This can be resolved by designing the neural amplifier in 2 or 3 gain stages [42,69]. The other destructive effect of short channel effects is increasing the low-cutoff frequency of the AC-coupled neural amplifiers. In this chapter, we analyze the parameters that affect the low-cutoff frequency and propose two solutions. The first solution utilizes a standard CMOS and improves the low-cutoff frequency by increasing the input resistance. The second method utilizes thick-oxide transistors to increase the input resistance.

4.1.2 Low-Cutoff Frequency Analysis

Figure 4.1 shows the schematic of a neural amplifier with conventional capacitive feedback network (CFN) architecture. As explained in [39] and section 2.4.1, this architecture is one of the most popular architectures of AC-coupled neural amplifiers in terms of low power

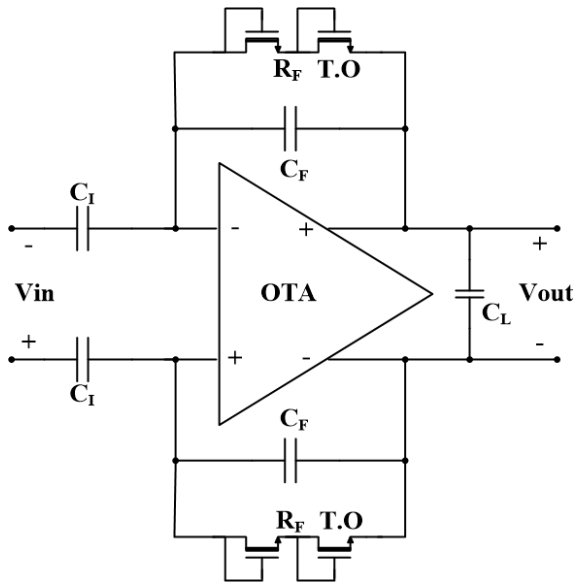


Figure 4.1 Fully differential capacitive feedback network neural amplifier

consumption, low noise, and compact area. Also, utilizing thick-oxide NMOS pseudoresistors instead of PMOS pseudoresistors, proivdes a better total harmonic distortion (THD) [72].

Figure 4.2 shows the frequency response of this CFN neural amplifier as a bandpass amplifier. Assuming that the voltage gain of the operational transconductance amplifier (OTA) is significantly high, the voltage gain of the amplifier in the midband (A_M) can be approximately calculated by

$$A_M = \frac{C_I}{C_F} \tag{4.1}$$

where C_I and C_F are input and feedback capacitance of the amplifier, respectively. Also, the low-cutoff frequency (f_L) of the amplifier can be approximated as

$$f_L = \frac{1}{2\pi R_F C_F} \tag{4.2}$$

where R_F is the dynamic resistance of NMOS pseudoresistors of the amplifier.

As presented in (4.2), in order to reduce f_L, C_F and R_F should be increased. However, by increasing C_F, it is required to increase C_I to maintain the same gain which results in huge area loss for each channel of a multi channel device. In addition, this results in the reduction of the input impedance of the neural amplifier.

MOS pseudoresistors can be utilized as a feedback resistance (R_F) for their compactness and

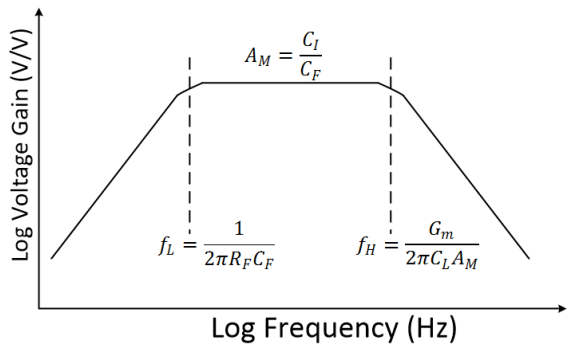


Figure 4.2 Frequency response of the amplifier.

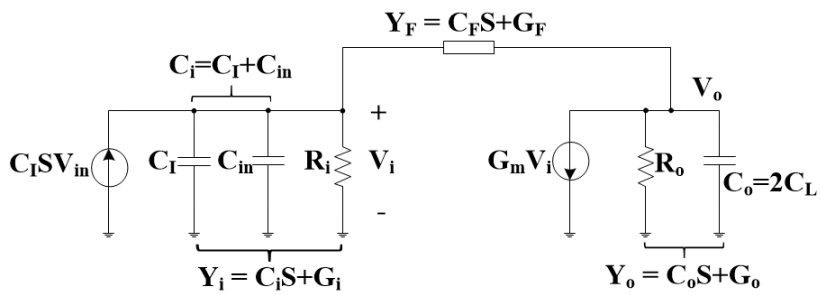


Figure 4.3 Small signal equivalent of the half-circuit of the neural amplifier.

high resistance. However, the drawback of this technique is that the MOS pseudoresistors provide much less resistance in advanced technology. For example, in an old technology such as 1.5 μm CMOS technology, by utilizing a MOS pseudoresistor for the R_F, a C_F of only 200 fF is enough to achieve a f_L of 0.025 Hz [39]. However, with the same technique and the same value for C_F, a f_L of 39 Hz is reported in the 180 nm CMOS technology [79]. Moreover, in the 130 nm CMOS technology [71], a higher C_F of 300 fF is used to compensate for the low R_F to provide a f_L of 0.1 Hz. Moreover, in the 65 nm CMOS technology, our simulation results show that when a C_F of 200 fF is used, the f_L is achieved at 472 Hz. To better understand the effects that increase the f_L value in the advanced CMOS technologies, we provide a small signal analysis of the amplifier in the following.

The equivalent small signal half-circuit of a neural amplifier of Figure 4.1 is depicted in Figure 4.3. The OTA can be modelled as a single pole amplifier with a pole at the output node. In this figure, G_m is the transcunductance of the OTA and C_{in}, R_i, and R_o are OTA's input terminal capacitance, resistance, and the output terminal resistance, respectively. We extract the time constant of the first pole as

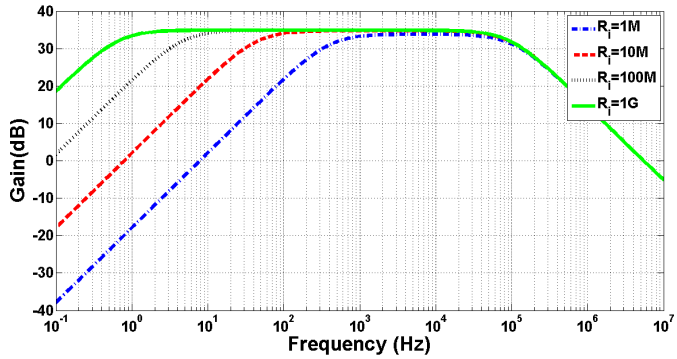


Figure 4.4 Frequency response of a neural amplifier with various amounts of R_i.

$$\tau_1 = \frac{1}{p_1} = \tag{4.3}$$

$$= \frac{C_F(G_o + G_m) + C_o G_F + C_i(G_o + G_F) + G_i(C_o + C_F)}{G_F(G_m + G_o) + G_i(G_F + G_o)}$$

Reduction of the oxide thickness in advanced technologies translates to lower input resistance (i.e., higher G_i) due to higher gate leakage current. By increasing G_i, the denominator in (4.3) grows much faster than the numerator. Therefore, the time constant (τ_1) increases resulting in lower f_L.

However, for older technologies, we can simplify (4.3) to (4.4) with the assumption that OTA's input resistance (R_i) is infinity (i.e., G_i is approximately zero) [148].

$$\tau_1 = \frac{1}{p_1} = R_F C_F + \frac{C_o R_o}{1 + G_m R_o} + \frac{C_i(R_F + R_o)}{1 + G_m R_o} \tag{4.4}$$

If the gain of the OTA ($G_m R_o$) is high, the second and third terms of this equation can be considered negligible resulting in (4.5) where the corresponding frequency to τ_1 is the same as (4.2). In other words, (4.5) is a special case of (4.3) where the gain of the OTA is high and the input resistance of the OTA is infinity.

$$\tau_1 = \frac{1}{p_1} = R_F C_F \tag{4.5}$$

Figure 4.4 illustrates the frequency response of the small signal model of the amplifier shown in Figure 4.3 for different values of R_i. The DC voltage of the outputs is biased at 0.5 V and thick-oxide NMOS pseudoresistors are utilized for feedback resistors. The values of G_m, R_o,

C_I, C_F, C_{in}, and C_o are chosen as 22.4 $\mu\mho$, 157 $M\Omega$, 11.5 pF, 200 fF, 3 pF, and 200 fF, respectively. As shown in this figure, f_L decreases by increasing R_i.

4.1.3 Proposed Solutions

In this section, we propose two solutions to decrease the low-cutoff frequency down to 1 Hz of OTA's in advanced CMOS technologies without increasing the feedback capacitance (C_F).

Cross-Coupled Positive Feedback

Figure 4.5 shows the architecture of the neural amplifier with cross-coupled positive feedback (CCPF) connections in which multiple (n+2) numbers of pseudoresistors are utilized [148]. Figure 4.6 shows two implementations of the CCPF connections (far and close connections) in which each pseudoresistor is implemented with a standard PMOS transistor. By knowing the fact that the CCPF provides a negative resistance ($-|R_N|$), the equivalent input resistance of the OTA can be presented by

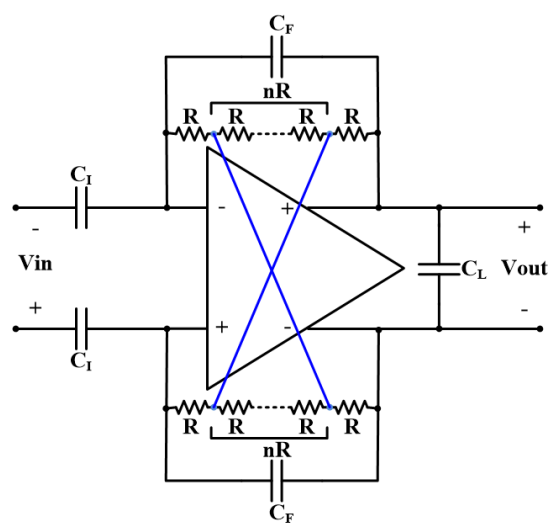


Figure 4.5 The neural amplifier with cross-coupled positive feedback architecture.

$$R_{ieq} = R_i || -|R_N| = \frac{R_i|R_N|}{|R_N| - R_i} \tag{4.6}$$

As presented in (4.6), to maximize R_{ieq}, ($|R_N| - R_i$) must be minimized. In other words, to achieve a very high positive equivalent input resistance, the amount of $|R_N|$ must be slightly higher than R_i, and ($|R_N| - R_i$) should approach zero. However, since this negative

resistance is created by positive feedback, the stability of the amplifier limits the lower bound of $(|R_N| - R_i)$.

To verify (4.6), we calculate the negative resistance of the CCPF. Figure 4.7 shows the small signal equivalent circuit of the neural amplifier with a far CCPF connections. For simplicity of calculation, we assume all the pseudoresistors are identical and have the same value.

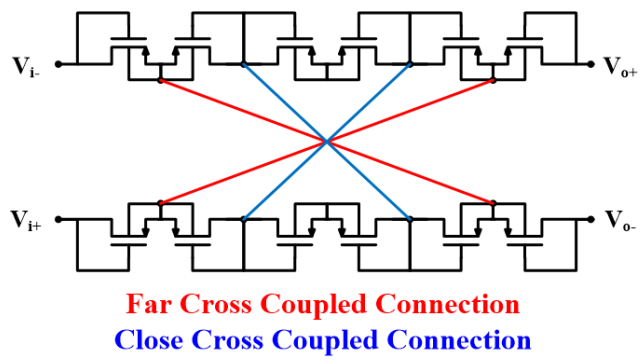


Figure 4.6 Cross-coupled positive feedback connections.

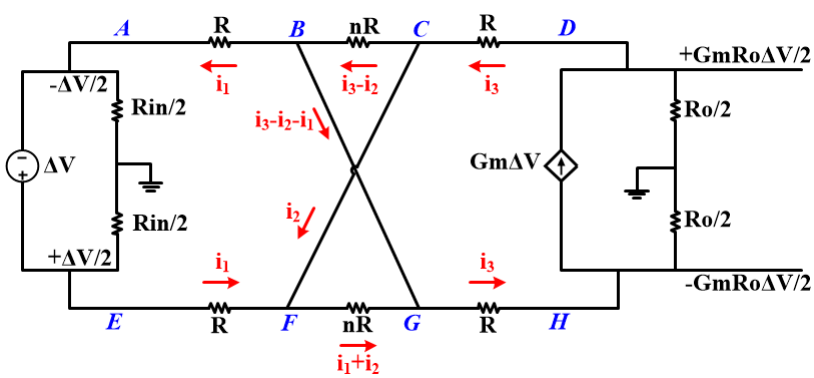


Figure 4.7 Small signal equivalent circuit of the neural amplifier with a CCPF connection.

Performing a KVL in the loops DCBGHD and DCFGHD results in

$$i_3 = i_1 + 2i_2 \tag{4.7}$$

Also Performing KVL on the loops of ABCFEA and DCBGHD and considering (4.7) results in the following two equations

$$(n+2)Ri_1 + nRi_2 = \Delta V \tag{4.8}$$

$$(n+2)Ri_1 + (n+4)Ri_2 = G_m R_o \Delta V \tag{4.9}$$

After solving these equations, the value of i_1 will be

$$i_1 = \frac{(n+4) - G_m R_o n}{4(n+2)R} \Delta V \tag{4.10}$$

As shown in Figure 4.7, $R_N = \frac{\Delta V}{i_1}$ is the equivalent resistance of the whole circuit connected to input terminals of the OTA (nodes A and E), which is parallel to R_{in}. By considering (4.10), R_N can be presented as

$$R_N = \frac{4(n+2)R}{(n+4) - G_m R_o n} \tag{4.11}$$

By knowing that the gain of the OTA ($G_m R_o$) is very high, the dominator of R_N is negative. In practice, the values of the pseudoresistors are not equal and vary based on their currents (or their voltages). Therefore (4.11) is not accurate and simulation results are required to calculate the exact value of R_N.

The value of the low-cutoff frequency of the amplifier depends on the number and size (W/L) of the pseudoresistors as well as the position of the CCPF connections (far or close). For example, assuming $C_I = 10\ pF$, $C_F = 200\ fF$, $C_L = 1.7\ pF$ and $n = 4$ for a far CCPF connection in the amplifier shown in Figure 4.5 achieves a f_L of 0.27 Hz with the midband gain of 31.67 dB, while the total capacitance value of this amplifier is 22 pF. In order to decrease the total capacitance, we exploited a T-capacitor feedback network shown in Figure 4.8 [149]. The pseudoresistors and CCPF connections in this figure are implemented similar to Figure 4.6 with 6 PMOS transistors.

The midband gain of the amplifier in Figure 4.8 is calculated as

$$A_M = \left(\frac{C_I}{C_{F1}}\right)\left(\frac{C_{F1} + C_{F2} + 2C_{F12}}{C_{F12}}\right) \tag{4.12}$$

We can adjust the capacitances in (4.12) to keep the total capacitance of the OTA low while maintaining the same gain. For example, in Figure 4.8, by choosing the value of the capacitors as $C_I = 1.4\ pF$, $C_{F1} = C_{F2} = 200\ fF$, $C_{F12} = 400\ fF$, and $C_L = 200\ fF$, the total capacitor value of the amplifier decreases to 4.2 pF, and the low-cutoff frequency increases from 0.27 Hz to 1.5 Hz, which is still in the acceptable range.

Figure 4.9 illustrates the frequency response of the amplifier in terms of gain and phase, in far, close, and no CCPF connections. The amount of the low-cutoff frequency for far, close, and no CCPF connections are 1.5 Hz, 143 Hz, and 320 Hz, respectively.

The positive feedback in the CCPF architecture of the amplifier can result in instability.

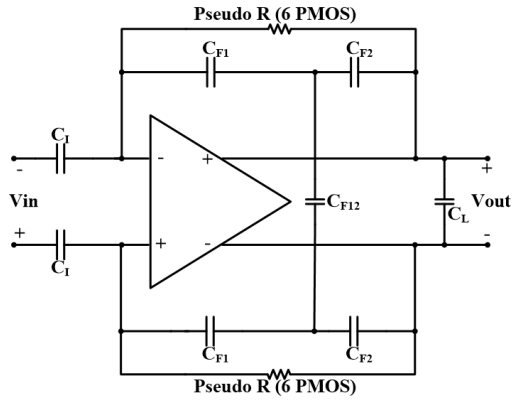
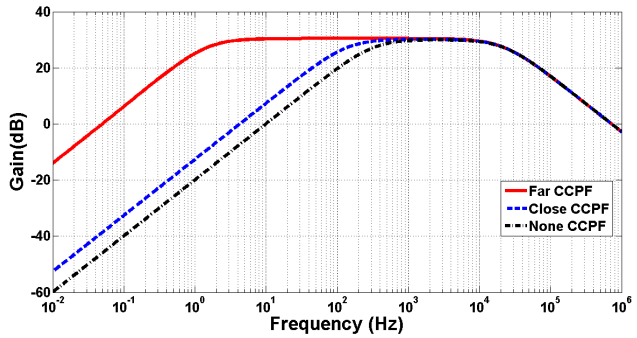
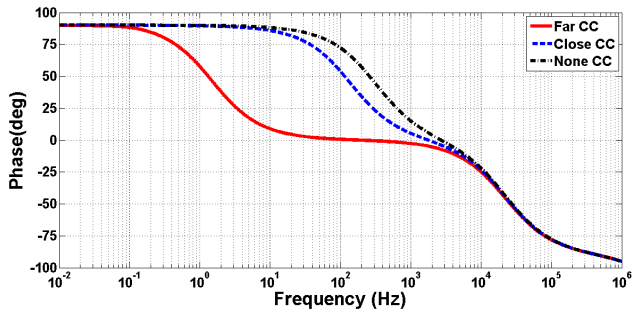


Figure 4.8 T-capacitor feedback network architecture with CCPF.



(a)



(b)

Figure 4.9 Frequency response of the amplifier of Figure 4.8 with far, close and no CCPF connection. (a) gain, and (b) phase.

However, by carefully designing the number of psudoresistors, transistor sizes, and the position of the CCPF connection we can make sure that the negative feedback is dominant and the whole architecture is stable and satisfies at least a 60 degree phase margin.

By adding switches to the CCPF connection we can program (i.e., turn on or off) the connections in the post-fabrication process. In case of multiple pseudoresistors (e.g., 18), the

farther CCPF connections might observe instability due to process variation. Therefore, by programming the connections and choosing closer connections, we can avoid instability. In addition, programmability can also give us control over the value of f_L. The closer connections have higher value of f_L and are more stable. On the other hand, the farther connections have lower value of f_L at the cost of less stability.

4.1.4 Thick Oxide Differential Pair

The second method to increase the input resistance of the OTA without increasing the feedback capacitance is to utilize thick-oxide MOS transistors in the input differential pair.

Figure 4.10 shows the transistor level implementation of the OTA of Figure 4.1 with thick-oxide PMOS input differential pair and Figure 4.11 illustrates the layout of neural amplifier drown by Cadence.

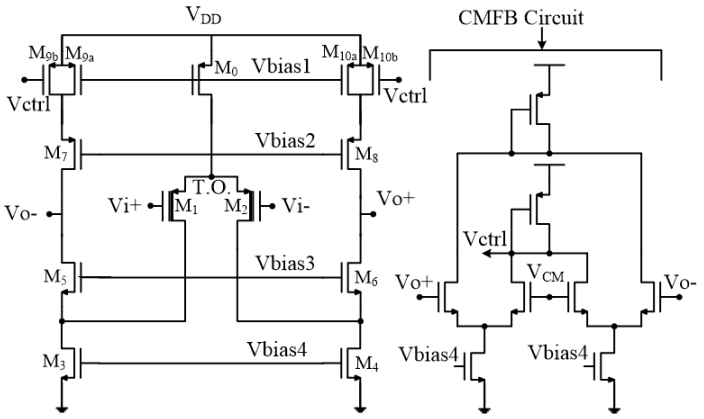


Figure 4.10 Fully differential folded cascode OTA utilized in neural amplifier.

Figure 4.12 shows the simulation results of the designed neural amplifier utilizing the OTA of Figure 4.10 and the OTA with standard PMOS input differential pair. The gain of the OTA and the whole neural amplifier are 68.2 dB and 34.6 dB, respectively. As shown in this figure, applying a thick-oxide PMOS in the input differential pair improved the low-cutoff frequency from 360 Hz to 0.19 Hz. These simulation results confirm that increasing the input resistance of the OTA by utilizing thick-oxide PMOS in the differential pair decreases the low-cutoff frequency dramatically.

In order to increase the SNR of the neural amplifier, the first stage of a neural amplifier is designed as an LNA. To reduce the flicker noise of the OTA of Figure 4.10, we optimize the size of the PMOS transistors in the input differential pair (i.e., M_1 and M_2). Also, as mentioned in [39], to minimize the thermal noise, the transistors M_1 and M_2 are biased

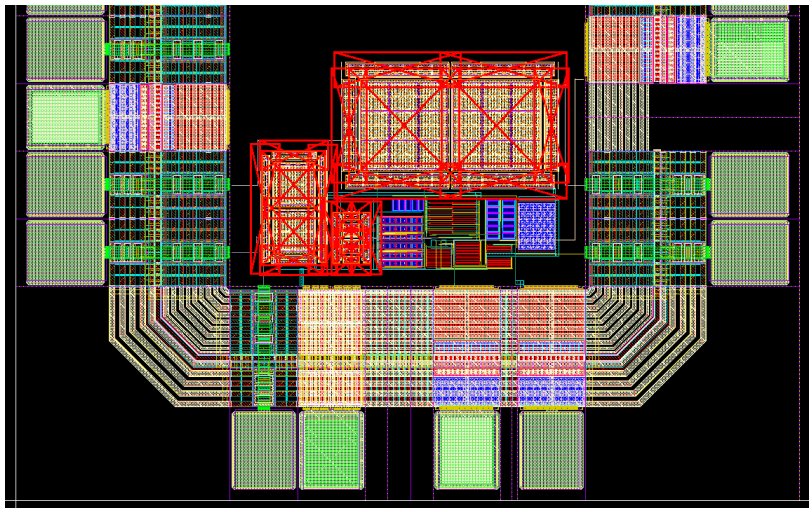


Figure 4.11 The layout of the designed neural amplifier.

in the sub-threshold region to maximize their transconductance over drain current called transconductance efficiency (g_m/I_D), and the transistors M_3, M_4, M_{9a}, M_{9b}, M_{10a} and M_{10_b} are biased in the saturation region to minimize their g_m/I_D.

As mentioned earlier, the bandwidth and operating frequency of neural amplifiers are very low, therefore the dominant noise power is the flicker noise. Also, in the OTA of Figure 4.10, the differential pair transistors are the main source of the flicker noise in comparison with other transistors [78]. Therefore, to analyze the noise of the proposed neural amplifier, we only investigate the effect of the thick-oxide PMOS differential pair. Utilizing thick-oxide PMOS transistors in the differential pair of the OTA decreases the gate-oxide capacitance per unit area (C_{ox}) due to the increased gate oxide thickness (t_{ox}).

Utilizing the thick-oxide PMOS in the input differential pair increases the flicker noise power due to decreasing C_{ox}. The relation between the input-referred noise of the whole neural amplifier ($\overline{V_{ni,amp}^2}$) and the OTA input-referred noise ($\overline{V_{ni}^2}$) is presented as

$$\overline{V_{ni,amp}^2} = \left(\frac{C_I + C_F + C_{in}}{C_I}\right)^2 . \overline{V_{ni}^2} \tag{4.13}$$

Decreasing the C_{ox} due to utilizing the thick-oxide PMOS differential pair, increases $\overline{V_{ni}^2}$ and decreases the C_{in} in (4.13). Since the increase in $\overline{V_{ni}^2}$ is much higher than the reduction of C_{in}, the $\overline{V_{ni,amp}^2}$ increases by decreasing the C_{ox}. To compensate this drawback, we can increase the gain of the LNA (C_I/C_F) by increasing C_I to reduce the $\overline{V_{ni,amp}^2}$ in (4.13). Simulation results show that the minimum input-referred noise voltage of the neural amplifier is 5.9 μV_{rms} in the frequency range between 1 Hz and 5.6 kHz (bandwidth).

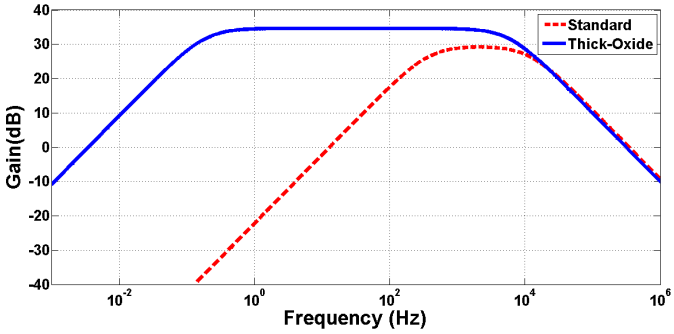


Figure 4.12 Frequency response of neural amplifier with thick-oxide and standard PMOS differential pair.

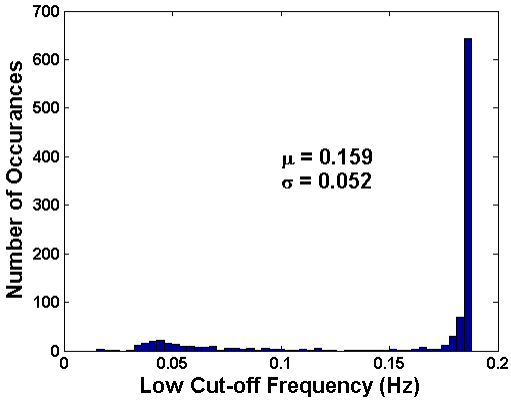


Figure 4.13 Monte Carlo simulation of low-cutoff frequency of the neural amplifier.

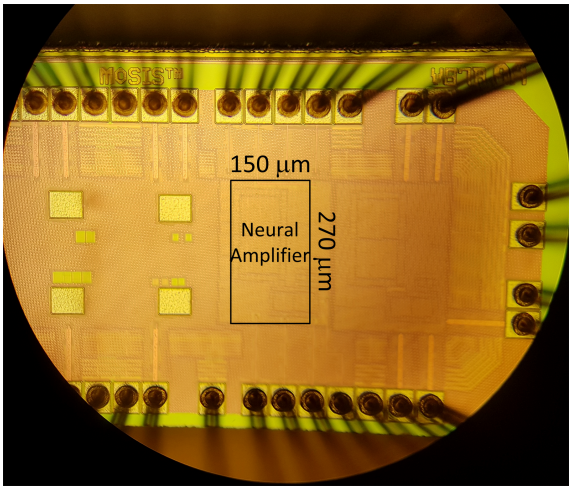


Figure 4.14 Micrograph of chip containing the neural amplifier with 270 μm $\times$ 150 μm die area.

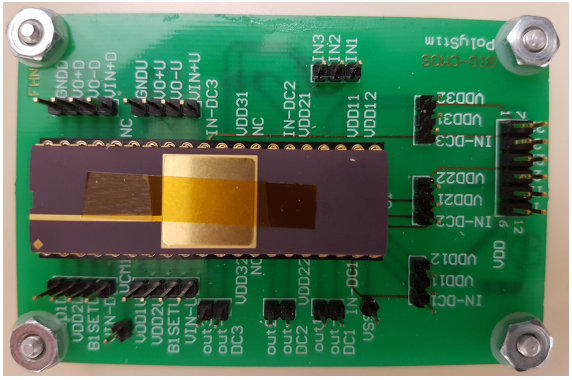


Figure 4.15 The test PCB and the prototype IC.

Note that to further reduce the noise of the OTA, it is required to apply noise reduction techniques which has been explained in section 2.4.3.

Figure 4.13 shows the Monte Carlo simulation results (N=1000) of the low-cutoff frequency. As shown in this figure, the μ is equal to 0.159 Hz and the σ is equal to 0.052, resulting $\frac{3\sigma}{\mu}$ of 0.983.

4.1.5 Measurement and In Vitro Results

Measured Performance

The prototype is implemented in the TSMC 65 nm CMOS process. The C_I and C_F are set to 11.5 pF and 208 fF, respectively, in the layout to achieve a gain of 55 V/V (or 34.8 dB) ($A_M = \frac{C_I}{C_F}$). The prototype uses 0.04 mm^2 (270 μm $\times$ 150 μm) of silicon area. The micrograph of the die containing the amplifier is shown in Figure 4.14 and Figure 4.15 illustrates the test printed circuit board (PCB) and the prototype IC.

The measured frequency response from 0.1 Hz to 1 MHz is performed through saline medium to mimic the brain environment and is illustrated in Figure 4.16. The midband gain is 34.3 dB and the low and high-cutoff frequencies are 2 Hz and 5.6 kHz, respectively. The simulated low-cutoff frequency is 0.19 Hz which is less than that achieved in the measurement result. This deviation is expected as the MOS pseudoresistors are nonlinear and significantly sensitive to their operating point [39].

Figure 4.17 shows the measured input-referred noise voltage spectral density of the neural amplifier. The RMS value of the input referred noise is achieved as 6.1 μV_{rms} by integrating the area under the curve from 1 Hz to 5.6 kHz (amplifier bandwidth) in Figure 4.17. This value is slightly higher than the simulated result (5.9 μV_{rms}).

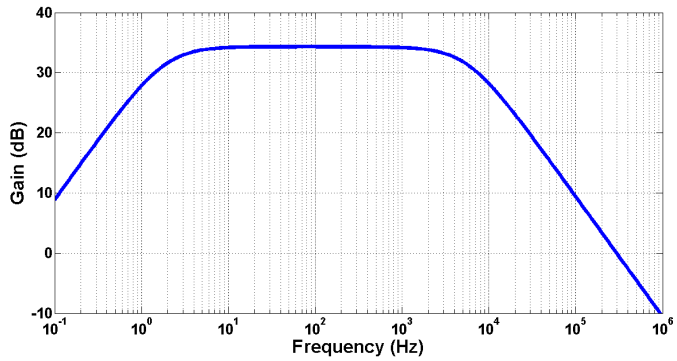


Figure 4.16 Measured frequency response of the amplifier. Midband gain is 34.3 dB, and the low and high-cutoff frequencies occur at 2 Hz and 5.6 kHz, respectively.

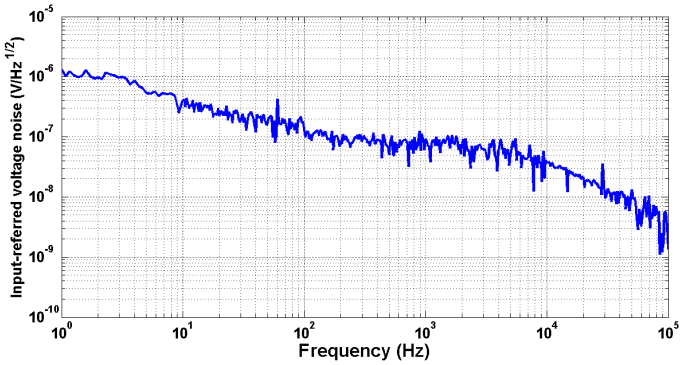


Figure 4.17 Measured input-referred noise voltage spectrum.

Table 4.1 EXPERIMENTAL AND SIMULATION CHARACTERISTICS OF NEURAL AMPLIFIER

Parameter	**Simulation**	**Measured**
Supply voltage [V]	1	1
Supply current [μA]	3.63	3.63
Gain [dB]	34.6	34.3
Band width [kHz]	5.8	5.6
Low-cutoff frequency [Hz]	0.19	2
Input-Referred Noise [μV_{rms}]	5.9	6.1
Noise efficiency factor	5.8	6.1
THD (2 mV_{pp} at 1 kHz) [%]	0.18	< 1

Table 4.1 shows a summary of the simulated and measured parameters of the prototype. A comparison of our work and the other published works is presented in Table 4.2. All of the chosen neural amplifiers are AC-coupled. To fairly compare these amplifiers with different gain values, number of stages and technology, we only consider the first stage of

each amplifiers.

Measurement results show that the achieved gain is the highest among all in Table 4.2. Note that the gain for [151] is reported for two stages. Also, the area of the fabricated chip is less than others. However, we should note that comparing the chip area itself without considering the midband gain is not a fair comparison. The midband gain (A_m) of the amplifier is equal to $\frac{C_I}{C_F}$. The low-cutoff frequency (f_L) is determined by C_F, and C_I is determined by the gain and C_F. Also, note that the main contributor to the chip area is C_I. In other words, for a normalized gain, lower C_F results in less chip area. Therefore, comparing C_F is a better figure of merit for comparing the chip area while the amplifiers have different gains. In this case, the values of C_F of the proposed amplifier and [68] are 208 fF and 350 fF, respectively. Note that the gain reported in our work is 34.3 dB, while the gain in [68] is 26.4 dB. This is why the total area of our work is almost the same as that of [68].

The amplifier of [150] has been implemented in the 0.18 μm technology with a gain of 26 dB. Its high pass pole is 80 Hz. The value of C_F is not reported, however, the total area of the amplifier is 0.16 mm^2 which is significantly large. In [71], neural amplifiers with a gain of 54-60 dB in two gain stages have been implemented in the 0.13 μm process. The first stage (LNA) with the estimated gain of 31.8 dB has 300 fF feedback capacitors with 0.1 Hz low-cutoff frequency. Our analysis shows that the C_F in [71] could be reduced to 200 fF if

Table 4.2 COMPARISON OF FULLY INTEGRATED NEURAL AMPLIFIERS

Parameter	**[150]**	**[71]**	**[151]**	**[67]**	**[68]**	**This Work**
Technology [CMOS]	0.18 μm	0.13 μm	0.13 μm	65 nm	65 nm	65 nm
Area [mm^2]	0.16	N/A	0.4*	N/A	0.042	0.04
Supply [V]	1.2	1.2	0.8	1	1	1
Power Consumption [μW]	0.43	4.5	0.64	1.2	3.28	3.63
Gain [dB]	26	31.8	49 **	26	26.4	34.3
BW [Hz]	80-15k	0.1-5k	100*-6.2k	10-8k	1-8.2k	2-5.6k
C_F [fF]	N/A	300	N/A	500	360	208
Input-Referred Noise [μV_{rms}]	8.1	6.5	14	7.5	4.13	6.1
Noise BW [Hz]	80-15k	10-5k	100*-6.2k	100-10k	1-8.2k	1-5.6k
NEF	1.52	7.2	6.5	3.6	3.19	6.1

* Estimated.
** This gain is reported for two stages. All other gains are reported for the first stage.

the thick-oxide differential pair is used.

The neural amplifier of [151] has employed two gain stages to obtain 49 dB in the 0.13 μm process. The value of the C_F is not reported . However, the estimated amplifier area and f_L are 0.4 mm^2 and 100 Hz, respectively. This amplifier occupies a very large area and has a high low-cutoff frequency. The designs in [67] utilize LNA with a gain of 26 dB fabricated in the 65 nm CMOS Technology. It employs a 500 fF feedback capacitor parallel to a pseudoresistor in a conventional CFN architecture similar to our work. The low-cutoff frequency f_L is adjustable, with the minimum value of 10 Hz. The neural amplifier consists of a variable gain amplifier (VGA) and buffer to achieve a gain of 45-60 dB. The amplifier in [68] has been implemented with two gain stages with 52.1 dB midband gain in the 65 nm technology. The gain in the first stage, LNA, is 26.4 dB and the f_L is reported as 1 Hz. The LNA exploits a CMOS-inverter-based OTA with 360 fF as C_F.

In Vitro Neural Recording

We used this neural amplifier for neural recordings in an *in vitro* experiment on the slices of a mouse brain at the faculty of Dentistry at University of Montreal. A micropipette is used to capture the electrical activity of the brain. The micropipette is filled with NaCl (0.5 mol) without bubbles. This micropipette contains a metal electrode of AgCl which records the extracellular APs of the brainstem of the mouse brain slice. The brain slice is inserted and fixed in a chamber which contains artificial cerebrospinal fluid (ACSF) which is continuously oxygenated and kept humid to mimic a real brain environment and to keep the neurons alive for a few hours. The micropipette is gradually penetrated into the brainstem tissue by means of a microscope and its peripheral tools.

To complete the test setup, the AgCl electrode of the micropipette is connected to the non-inverting input of the prototype amplifier. The connection of the chamber, including the ACSF, is connected to the inverting port of the amplifier as a Vref. It should be noted that shielded wires are utilized to perform these connections. A commercial setup of a neural recording system containing an instrumentation amplifier (A-M systems, Inc.), rack mounted data acquisition equipment and a PC with a spike2 Windows-based software (version 5.19, Cambridge Electronic design) was utilized. The output of the proposed amplifier is connected to the commercial amplifier. The commercial amplifier is a band pass amplifier with a midband gain of 100 (V/V) and with low and high cutoff frequencies of 300 Hz and 5 kHz, respectively. Setting the low-cutoff frequency at 300Hz allows us to eliminate the LFP and extract the extra cellular APs from the output signal. By using the commercial amplifier as the second stage amplifier, the total gain is achieved at 5300 V/V. During the

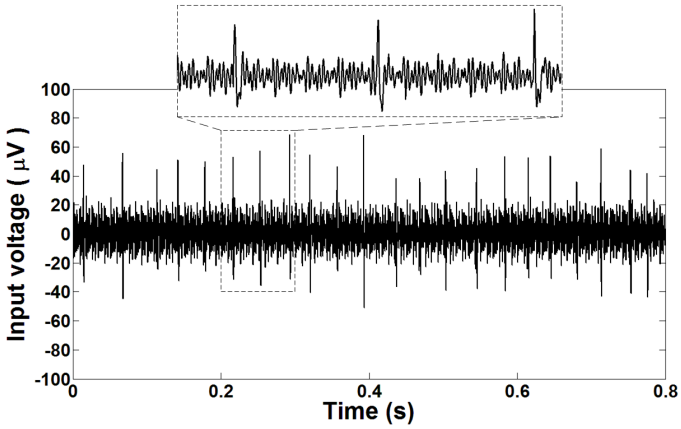


Figure 4.18 Recorded extracellular APs extracted from the brainstem of a mouse with the fabricated neural amplifier.

test procedure, the amplified signal is sampled with a frequency of 10 kS/s and digitized by the mentioned data acquisition equipment and transferred to the PC. Spike2 was used to observe the captured data in the PC. Figure 4.18, illustrates the recorded spontaneous extra cellular APs from the brainstem of the mouse with the proposed neural amplifier.

4.1.6 Conclusion

Scaling down technology introduces new challenges in neural amplifier design. One main challenge is the increased low-cutoff frequency (f_L) of the AC-coupled amplifiers, assuming the same feedback capacitance value is used. The simplest solution is to increase the feedback capacitors. However, this comes at the cost of increased input capacitors for the same gain of the amplifier, which increases the silicon area and decreases the input impedance of the amplifier. Assuming a neural recording implant requires a large array of these amplifiers, the total consumption of the silicon area increases dramatically.

In design of the neural amplifier, we focus on this challenge, find its roots, and propose solutions to improve it. Scaling down the technology increases the leakage current of the differential pair of the OTA due to decreasing the gate oxide thickness (short channel effects). This is translated to decreasing the input resistance (R_i) of OTA. We show, through simulations backed by an analytical analysis, that decreasing R_i is the fundamental reason for the increase in f_L. Two different solutions are presented in this section to increase R_i: applying a cross-coupled positive feedback architecture and utilizing thick-oxide PMOS transistors in a differential pair of the OTA. The simulations confirm that both of the solutions decrease the f_L. We designed and fabricated the latter solution in the 65 nm TSMC process.

The experimental results show that the low-cutoff frequency decreases to 2 Hz with 208 fF feedback capacitor (C_F). The neural amplifier is verified by *in vitro* experiment on mouse brainstem slices.

4.2 Voltage Reference Design

4.2.1 Introduction

Telemetry powering is commonly used in biomedical applications for its many advantages [152]. These advantages are, but not limited to no battery requirement, ease of movement, not restricted to wire lines, and reduction in the risk of infection. The energy required for these systems is transferred through waves from the outside world to the implanted device. Since the distance between the receiver and the transmitter varies, resulting in supply voltage variation, these waves have to first get rectified and get regulated for correct operation. To provide a clean supply voltage to the implanted device, the voltage regulators should operate steadily over a range of supply voltage and temperature [78, 153] . The area and power consumption are important factors in designing voltage references when they are implanted in human body [63].

The output voltage produced by conventional voltage references is limited by the bandgap of silicon which is near 1.2 V [78]. This limitation imposes a significant challenge for advanced technologies where the nominal supply voltage is less than 1 V.

The first attempt for sub 1 V output voltage-reference design is presented in [154]. we produce the voltage reference as the sum of two currents, unlike the conventional designs where the voltage reference is the sum of two voltages. The main drawback of this circuit is the BJTs used for implementing the diodes which significantly increases the area. In addition, these diodes require at least 0.7 V for the forward-biased voltage of PN junctions that avoids further reduction of the supply voltage.

The authors in [155] provide two voltage reference designs operating at the supply voltage of below 0.7 V in the 90 nm CMOS technology. These designs suffer from very high power consumption.

A subthreshold voltage reference is proposed in [156]. By replacing the amplifier in the design of [154], with a low-voltage comparator, a charge pump circuit, and a digital control circuit, the power consumption is significantly reduced. This design still illustrates high power consumption (5.35 μW at 250 mV) for implanted biomedical applications. The design in [157] utilizes an optimized operational transconductance amplifier (OTA) where the supply voltage can be reduced down to 0.4 V. This design consumes high power consumption and

its high temperature coefficient (T_C) is not suitable for many implanted applications.

The authors in [158] and [159] propose two voltage references in 12 nm and 7 nm FinFET technology, respectively. The design in [159] provides high precision and programmable temperature coefficient structure. The drawback of both of these designs is their high power consumption.

In this section, we explore two low power voltage references suitable for implanted applications [160]. We avoid BJTs to potentially reduce the area and force the transistors to operate in the subthreshold region to reduce the power consumption. In addition, to achieve improved PSRR and proper line sensitivity, in the first approach, we utilize two levels of voltage references. In this technique, the first voltage reference provides the supply voltage to the second voltage reference. In the second approach, we utilize feedback and voltage regulated technique to provide a local supply voltage (V_{DDL}) to the main voltage reference. Both of these techniques illustrate low power consumption. The fundamentals of voltage-reference design are presented in Section 4.2.2. We present the proposed designs and the corresponding simulations in Section 4.2.3. Section 4.2.4 concludes this voltage reference.

4.2.2 Fundamentals of Voltage Reference Design

A voltage reference circuit is composed of two subcircuits. The first subcircuit is complementary to absolute temperature (CTAT) and the second subcircuit is proportional to absolute temperature (PTAT). The gate-source voltage (V_{GS}) of a CMOS transistor is used as the CTAT component [161]. On the other hand, the difference of two gate-source voltages (ΔV_{GS}) in the subthreshold region is used to produce the PTAT component which is proportional to the thermal voltage (V_T). V_T is equal to $\frac{kT}{q}$ where k is the Boltzmann constant, and q is the electric charge on the electron [78]. By increasing the temperature, the V_{GS} decreases while ΔV_{GS} increases. The slope value of PTAT is much lower than the CTAT. Therefore, to make the values of these slopes equal, we amplify the slope of PTAT by a coefficient (K). V_{REF} is produced by summing the V_{GS} with $K\Delta V_{GS}$ which is represented in (4.14). Therefore, the variation of V_{REF} with respect to temperature is almost zero.

$$V_{REF} = V_{GS} + K\Delta V_{GS} \tag{4.14}$$

4.2.3 Proposed Design

Figure 4.19 shows the main block voltage reference used in [162,163] which satisfies (4.14) and is appropriate for implant applications. To make sure that this block operates correctly, the resistor R_1, and transistors M1 and M2 should be sized in such a way that these transistors operate in the subthreshold region. The C_S is used for startup operation. In addition, the current mirror constructed by M4 and M5 dictates that the currents of both branches to be equal (i.e., $I_{D1} = I_{D2}$), assuming that channel-length modulation is negligible for M4 and M5 transistors. We also assume that the body effect and channel-length modulation are negligible for M1 and M2 transistors. Since I_{D1} is equal to I_{D2} and M2 is sized larger than M1, V_{GS2} is lower than V_{GS1}. Therefore V_{R1}, equal to the difference of the V_{GS1} and V_{GS2}, is represented as

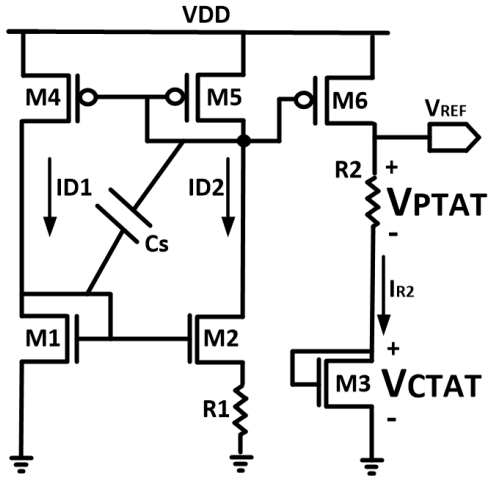


Figure 4.19 The main block voltage reference.

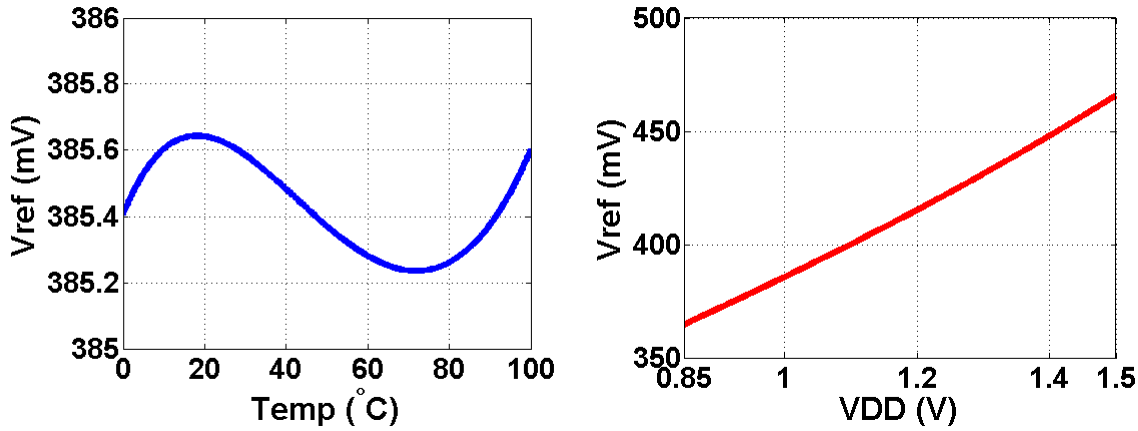
$$V_{R1} = \Delta V_{GS} = V_{GS1} - V_{GS2} \tag{4.15}$$

and the current in R_1 is calculated by (4.16) which is mirrored by M5 and M6 and is equal to I_{R2}.

$$I_{R1} = I_{R2} = I_{D1} = I_{D2} = \frac{\Delta V_{GS}}{R_1} \tag{4.16}$$

By performing a KVL on the branch including R_2 and M3, V_{REF} can be calculated as

$$V_{REF} = V_{GS3} + R_2 I_{R2} = V_{GS3} + \frac{R_2}{R_1}\Delta V_{GS} \tag{4.17}$$



(a) (b)

Figure 4.20 (a) V_{REF} produced by the main block versus temperature at 1 V. (b) V_{REF} produced by the main block versus V_{DD}.

Figures 4.20(a) and (b) show the behavior of the V_{REF} generated by the main voltage reference (shown in Figure 4.19) versus temperature and the supply voltage (V_{DD}), respectively. The T_C is calculated as

$$T_C = \frac{1}{V_{REF(27^\circ C)}} . \frac{V_{REF_{max}} - V_{REF_{min}}}{T_{max} - T_{min}} \times 10^6 \tag{4.18}$$

where $V_{REF(27^\circ C)}$ is the V_{REF} at 27°C, $V_{REF_{max}}$ and $V_{REF_{min}}$ are the maximum and minimum of the V_{REF}, respectively, and T_{max} and T_{min} are the maximum and minimum of the temperature, respectively. The T_C calculated from Figure 4.20(a) is equal to 10.6 $ppm/^\circ C$ which is in the acceptable range. Figure 4.20(b) shows that V_{REF} is very sensitive to the supply voltage and the sensitivity is 14.8%. In this case, the PSRR is -16.97dB (from DC to 10 kHz) which is not suitable for many biomedical applications. The high sensitivity of V_{REF} to power supply voltage is due to short channel effects in this technology.

In the following, we propose two solutions to reduce the effect of the supply voltage on V_{REF} (power supply rejection ratio and line sensitivity). The first solution is shown in Figure 4.21 as the two-stage cascode voltage reference (TSC-VR). As shown in this figure, the output of

the auxiliary voltage reference produces a supply voltage to the main voltage reference V_{DDL}.

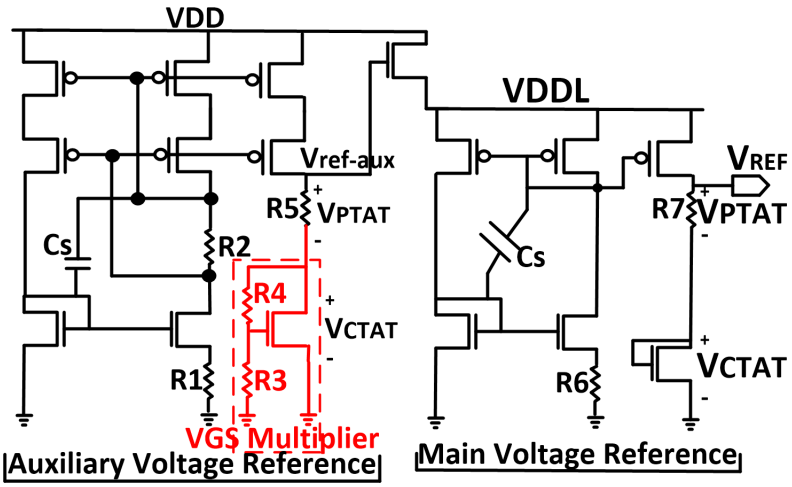


Figure 4.21 Two-stage cascode voltage reference (TSC-VR).

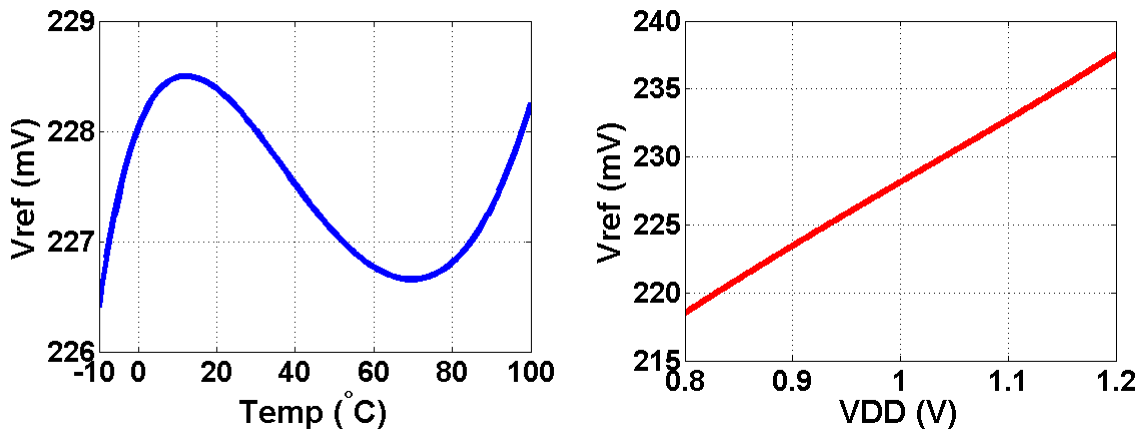
The main voltage reference produces a 228 mV of V_{REF} at 27°C. Figure 4.22(a) and (b) shows the V_{REF} variation versus temperature and the supply voltage, respectively. Note that to reduce the effect of the supply voltage variation, we utilize a self-biased wide swing cascode current mirror in the auxiliary voltage reference. In addition, to sufficiently increase the voltage at the $V_{ref-aux}$, a V_{GS} multiplier is used to create a proper V_{CTAT}. The temperature coefficient is observed as 83.5 $ppm/°C$ and the sensitivity of the V_{REF} to the supply voltage is equal to 4.75% (PSRR = -26.5 dB). Although this approach can be used for many applications, the sensitivity to the supply voltage, and the PSRR is still high.

As a second solution, to reduce the effect of the supply voltage variation on V_{REF}, we utilize a voltage regulator technique as presented in [164]. The voltage reference with regulated supply voltage (VRRS) circuit implemented based on this technique is shown in Figure 4.23. The output voltage of the voltage regulator, V_{DDL}, is calculated by (4.19).

$$V_{DDL} = V_{REF}(1 + \frac{R3}{R4}) \tag{4.19}$$

As shown in (4.19), V_{DDL} is proportional to V_{REF} and independent to V_{DD}. By connecting V_{DDL} to the supply voltage of the main block, we are separating the main supply voltage (V_{DD}) from this block. In other words, we are avoiding the effect of V_{DD} variation on V_{REF}.

The voltage regulator is composed of a two-stage OTA to increase the gain of the OTA. The biased currents of the OTA are generated by V_{REF} to avoid any change in the operating



(a) (b)

Figure 4.22 (a) V_{REF} produced by TSC-VR versus temperature at 1 V. (b) V_{REF} produced by TSC-VR reference versus V_{DD}.

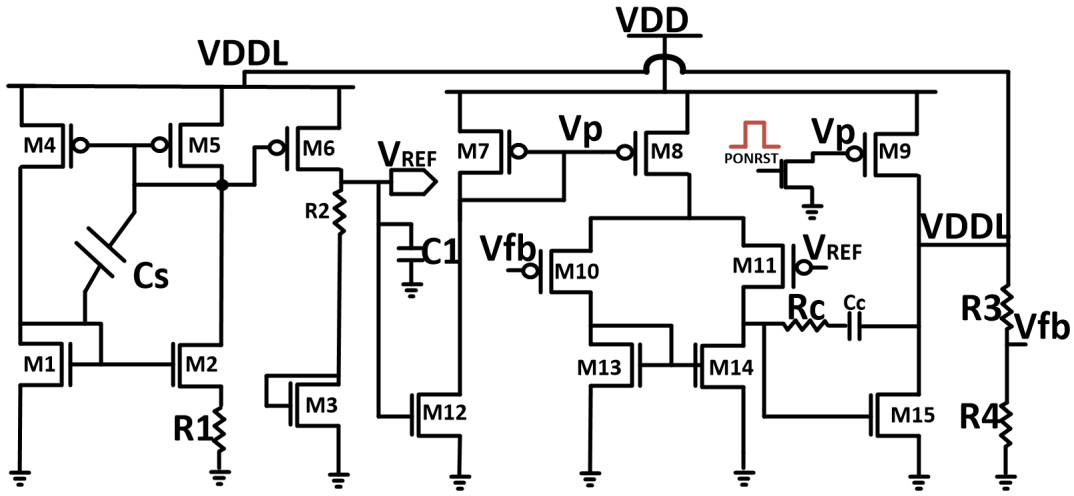
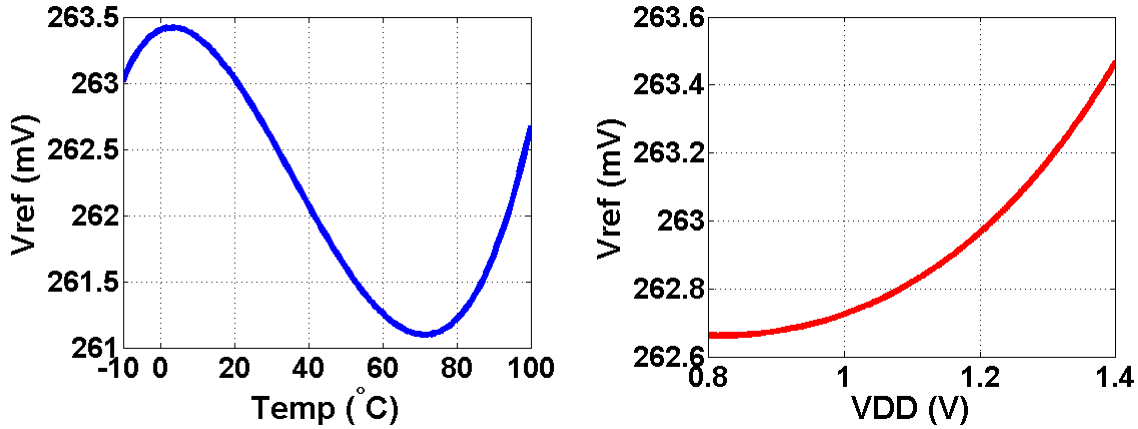


Figure 4.23 The transistor level of the voltage reference with regulated supply voltage (VRRS).

points of the transistors with respect to the V_{DD}. To provide stability for the OTA, it is required to carefully size the compensation resistor and capacitor (R_c and C_c). The OTA is stabilized with 70° phase margin for unity gain feedback. Also, C_1 helps to stabilize V_{REF} as well as V_{DDL} while the process is turning on. Moreover, we provide the power-on-reset (PONRST) pulse to initially power up the V_{DDL} output.

Figures 4.24(a) and (b) show the variation of V_{REF} of VRRS design versus temperature and V_{DD}, respectively. As shown in these figures, T_C is 80 $ppm/°C$ and the sensitivity to the supply voltage variation is 0.13%. The PSRR is -63 dB (from DC to 10 kHz). Simulation



(a) (b)

Figure 4.24 (a) V_{REF} produced by VRRS versus temperature at 1 V. (b) V_{REF} produced by VRRS versus V_{DD}.

results show that the power consumption of the proposed voltage reference for V_{DD} equal to 1 and 0.8 V are 2.77 and 2.08 μW, respectively.

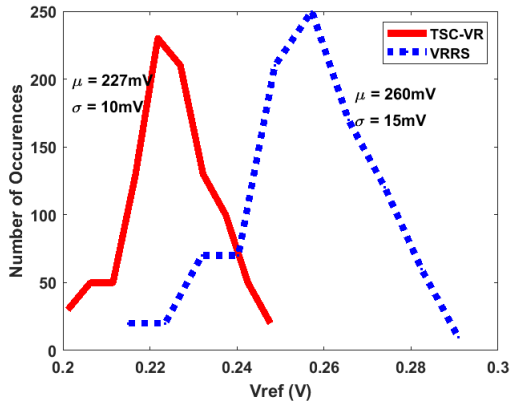


Figure 4.25 Monte Carlo simulations of V_{Ref} (N=1000) of both proposed voltage references.

Figure 4.25 shows the Monte Carlo simulations (N=1000) of both proposed voltage references. This figure shows that the design of TSC-VR and VRRS have $\frac{3\sigma}{\mu}$ of 0.132 and 0.173, respectively.

A comparison of the main parameters of the proposed designs with that of previous designs is provided in Table 4.3. The output voltage produced by all the designs in this table is around of 200 mV except for the design in [159]. We simulated both of the proposed designs in the range of -10 to 100°C. The TC of both TSC-VR and VRRS designs are less than the designs of [155], [156], and [157]. The design in [159] provides the minimum T_C among all. This is due

Table 4.3 COMPARISON OF MAIN PARAMETERS OF CHOSEN VOLTAGE REFERENCES

Ref.	**Tech**	**Nominal Voltage (V)**	**Minimum Voltage (V)**	V_{REF} **(mV)**	**Temp Range** ($^{\circ}C$)	**TC (ppm** /$^{\circ}C$)	PSRR (dB)	ΔV_{Ref} / ΔV_{DD} (mV / 100 mV)	**Power** (μ**W**)
[155]	90 nm	0.6	0.55	241	5 to 100	150	N/A	5	482
[156]	110 nm	0.4	0.242	195.6	-20 to 120	134	N/A	0.8	5.35
[157]	65 nm	0.6	0.4	275	-50 to 80	176	-36	N/A	62
[158]	12 nm	0.8	0.7	207	-20 to 125	40.51	N/A	N/A	6.88
[159]	7 nm	1.375	1.2	1000	-45 to 125	6	N/A	N/A	9.47
TSC-VR	65 nm	1	0.8	228	-10 to 100	83.5	-26.5	4.75	1.72
VRRS	65 nm	1	0.8	262.7	-10 to 100	80	-63	0.13	2.77

to the fact that V_{REF} in this design is significantly high (1 V). To provide a fair comparison, we compared the variation of the output voltage within 100 mV of supply voltage variation ($\Delta V_{Ref}/\Delta V_{DD}$). This comparison shows that the VRRS design achieves the minimum value. In addition, both of the proposed designs show the minimum power consumption where the power consumption of VRRS shows one-third of that of the design in [158]. The TSC-VR design shows almost half of the power consumption of that of the VRRS design, however, it suffers from the relatively higher PSRR. On the other hand, the VRRS design achieves a significantly low PSRR at the cost of a slightly increased power consumption.

4.2.4 Conclusion

Reducing the power consumption and chip area are crucial factors in designing voltage references for implanted devices. In addition, due to the impact of noise sources on the supply voltage, the effect of supply voltage variation on V_{REF} has to be reduced. Reducing the effect of supply voltage on V_{REF} is challenging in scaled technologies due to short channel effects. We propose two voltage references, TSC-VR and VRRS, in the 65 nm CMOS technology optimized for minimum power consumption with reduced PSRR. To reduce the PSRR, the design of TSC-VR utilizes an auxiliary supply voltage generator for the main voltage reference. On the other hand, in the VRRS design, the supply voltage of the main block is directly disconnected from V_{DD}. By utilizing a voltage regulator that is composed of a two-stage OTA, we provide a local supply voltage to the main block. Due to the feedback characteristic of the voltage regulator, the local supply voltage produced by this circuit re-

sults in lower variation with respect to the V_{DD}. Simulation results show that the sensitivity of the proposed voltage reference with respect to the supply voltage is -63 dB. Moreover, the achieved temperature coefficient is 80 $ppm/^{\circ}C$ and the power consumption is 2.77 μW.

CHAPTER 5 CS ENCODER AND ADC DESIGN

In this chapter, we explain the design of the CS encoder and SAR ADC. CS encoder consists of two circuits: pseudo-random number generator (PNG) and CS core. SAR ADC is the appropriate ADC for neural recording implants. We explain the design of a SAR ADC with configurable resolution between 7 and 10 bit. The sampling clock is also can be increased up to 50 MS/s to maximize the number of the channels for digitizing.

5.1 CS Encoder Design

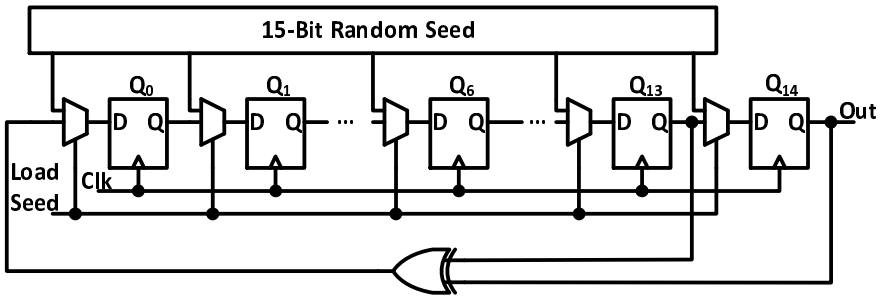


Figure 5.1 A 15-bit maximal-length Fibonacci LFSR.

Implementation of CS encoder can be done in digital or analog domain, but we propose to implement it in analog building blocks. The latter allows to reduce power consumption of subsequent circuits altogether with much compact circuit. It is because, power consumption is proportional to frequency of clock or in this case sampling frequency. As we explained in section 3.5, the CS core with parallel-series sequences of Figure 3.11 is appropriate for neural recording implants. In this architecture, changing the CF is possible easily by altering M_S, and subsequently the down sampling frequency of each branch ($\frac{M_S}{N} f_S$). This can be done by control unit of Figure 1.2.

The measurement matrix, Φ is needed in the receiver to decode the recovered data and reconstruct the original signal. It is also mentioned in section 3.2, the measurement matrix, Φ and the sampling matrix Ψ must be incoherent to capture the maximum independent linear combination of input samples. It has been proved that a random matrix Φ which is large enough in size is incoherent with any sampling matrix Ψ [109]. One of the simple methods to implement Φ is, utilizing of a PNG.

To achieve CS core implementation in analog circuits, most of available PNGs are generated

by Bernoulli distribution function. It means the value of measurement matrix (Φ) is 0 or 1 which can switch on or off an analog mixer. Therefore, the input signal samples are multiplied by ± 1 randomly. Implementing these random numbers is provided by digital circuits easily.

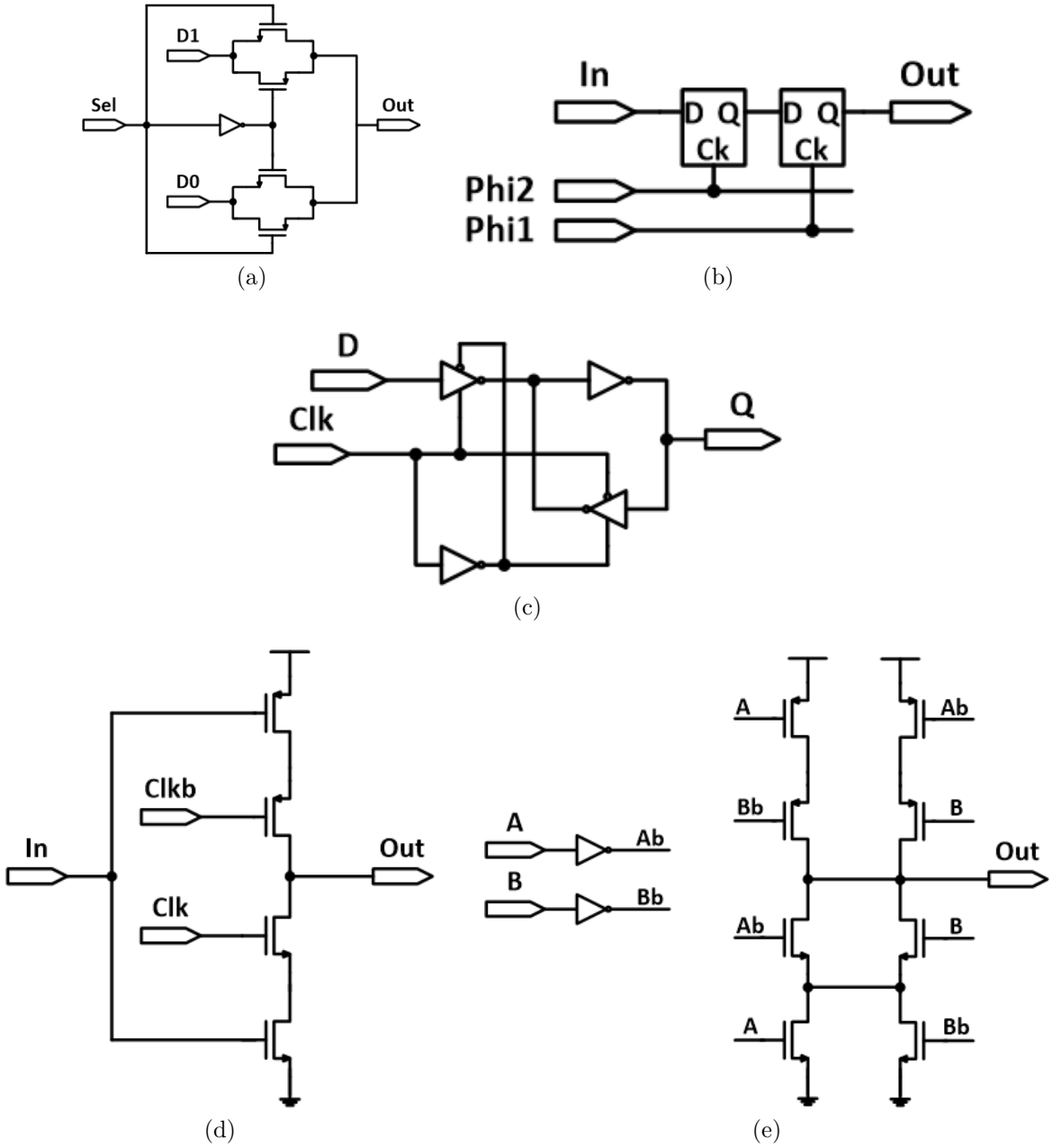


Figure 5.2 The utilized logic and digital circuits in the LFSR of Figure 5.1. (a) a 2×1 decoder. (b) Implement of a DFF activated by rising-edge of $\Phi 1$. (c) A clocked DFF. (d) A tristate inverter. (e) An XOR circuit.

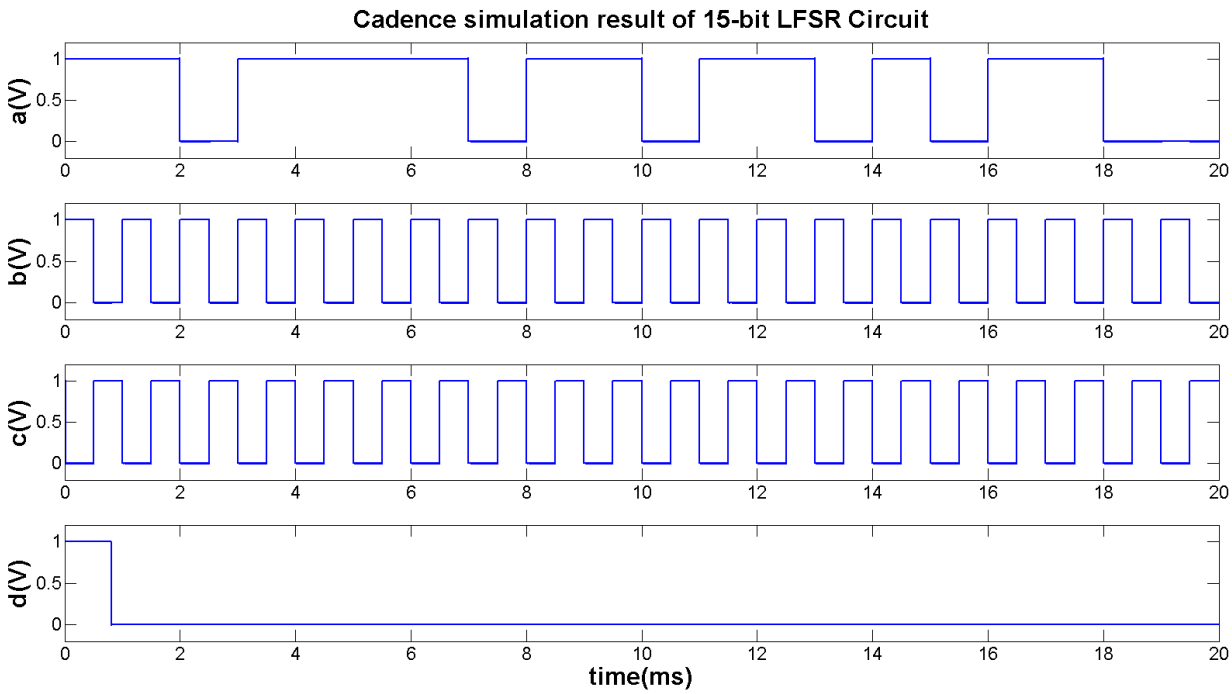


Figure 5.3 Simulation result of LFSR circuit: (a) Pseudo-Random output, (b) Φ1 (Non-overlapping clock), (c) Φ2 (Non-overlapping clock), and (d) Load Seed to start up the circuit.

5.1.1 Pseudorandom Number Generator

In practice generating these random numbers can be realized in various methods. Linear-feedback shift register (LFSR) is one of these methods. However, these numbers are not truly random due to their sequence iteration. Figure 5.1 shows a 15-bit maximal-length Fibonacci LFSR which is suitable for the proposed CS core. It is a string of 15 D-flip-flops (DFF) with an external XOR gate feedback. The initial value of the LFSR is called the seed and it is loaded for the first iteration. The value of seed is a random value and it should never be all zero as well as other states, otherwise the function of LFSR is stopped.

A maximal-length n-bit Fibonacci LFSR can generate $2^n - 1$ different sequences, where n is the number of bits in the LFSR. The LFSR of Figure 5.1 generates $2^{15} - 1$ unique random sequences. The bit positions that affect the next state are called the taps. In Figure 5.1 the taps are $\{14, 15\}$ which XOR's the taps 14 and 15 [165]. To implement the LSFR of Figure 5.1 and achieve the reliable expected signals, we exploit two non-overlapping clock Φ1 and Φ2 and DFF which capture the input data in rising edge of Φ1. Figure 5.2 shows the designed logics and circuits which have been utilized in the LFSR of Figure 5.1. The word "110,1111,0110,1101", (0X6F6D), is applied in the *15-bit random seed* block of Figure 5.1 and the circuit is simulated in Cadence. The result is illustrated in Figure 5.3. The clock frequency of Φ1 and Φ2 is 1 kHz.

5.1.2 Implementation of CS Core and PGA

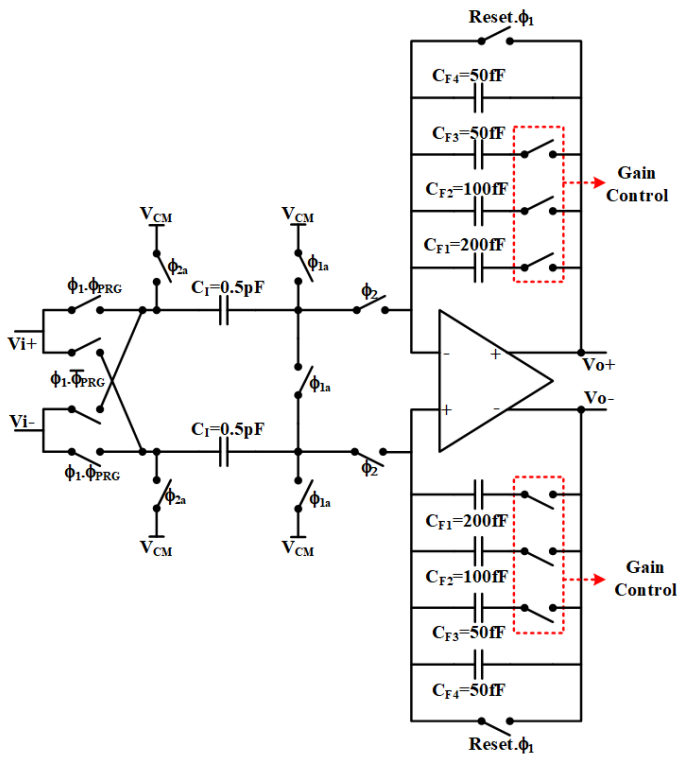


Figure 5.4 CS core circuit. (a) Switched-capacitor implementation of CS core and PGA. (b) Non-overlapping clock circuit.

In this section, we design the CS core circuit with programable integrator gain. Figure 5.4 shows this circuit, which is implemented by switched-capacitor (SC) technique. Multiplication of the input signal with pseudo-random sequence of Φ_{PRG} (± 1) is realized by the front-end passive double-balanced mixer, where its output signal is integrated in the following block. Moreover, input signal is amplified with the gain of $\frac{C_I}{C_{F,eq}}$. Changing the equivalent value of C_F is performed by means of gain switches. Gain control command is issued from control unit (CU) to the gain switches. The transfer function of non-inverting SC integrator is calculated as:

$$\frac{V_o(z)}{V_i(z)} = \frac{C_I}{C_{F,eq}} \cdot \frac{z^{-1}}{1 - z^{-1}} \tag{5.1}$$

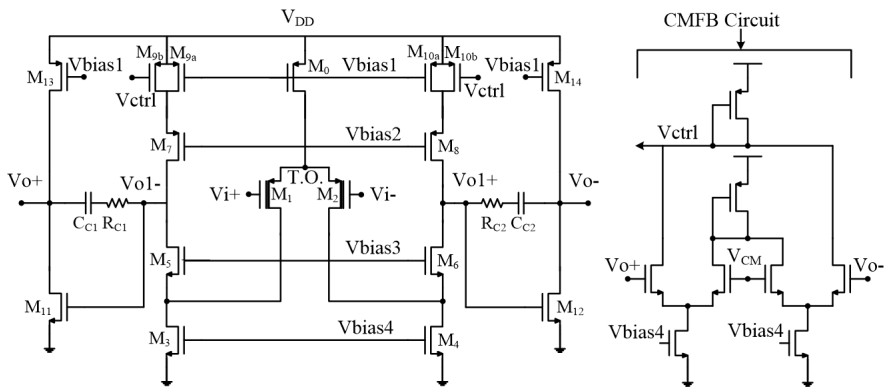


Figure 5.5 The OTA circuit used in the CS core.

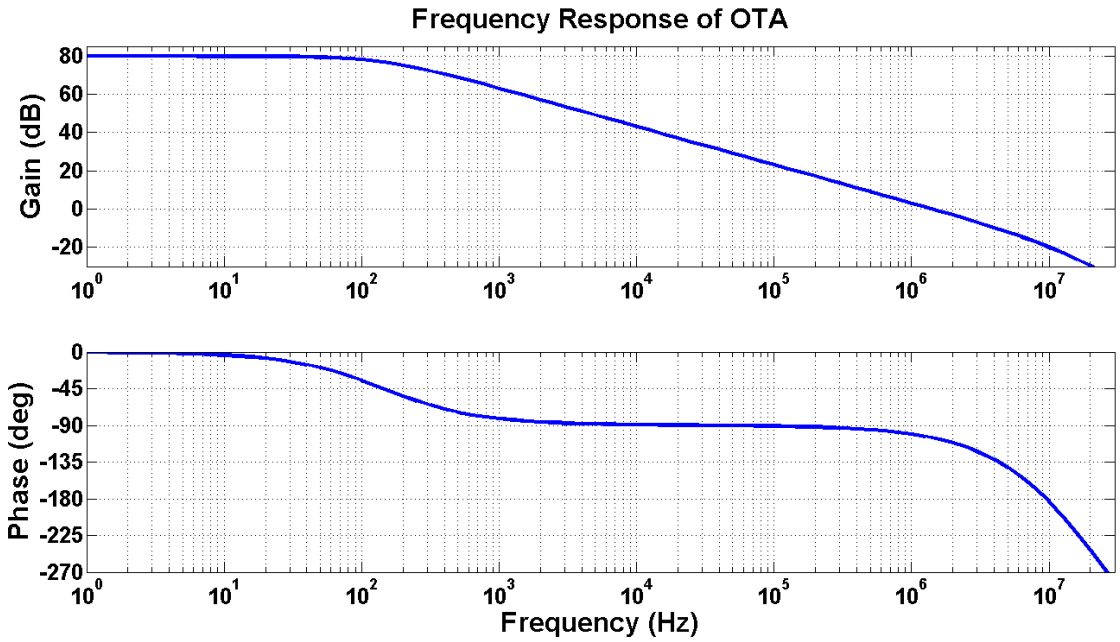


Figure 5.6 Frequency response of the OTA used in CS core.

Connecting the feedback switches by gain control unit changes the equivalent feedback capacitor. Altering the $C_{F,eq}$ from 50 fF up to 400 fF with the step of 50 fF in Figure 5.4 changes the gain of the integrator between 1.25 and 10 as it is suggested in section 3.6.

To design of the OTA utilized in CS core circuit of Figure 5.4, We should consider the non-ideality effects of OTA explained in section 3.3. To achieve 23 dB for the SNR of CS core

based on the Figure 3.3, the gain of OTA should be not less than 10000 (80 dB). At the same time, to achieve the maximum output SNR of the ADC we need a differential 1 V_{pp} in the output of the OTA (0.5 V_{pp} on each output terminal). One of the best topologies that satisfies these constrains in 65 nm CMOS process is 2-stage OTA with folded cascode in the first stage as illustrated in Figure 5.5. The frequency response of the designed OTA is shown in Figure 5.6. In this figure, the DC open loop gain of the OTA is 82 dB and the OTA is stabilized with 60° phase margin for the unity gain. The output swing is 530 mV_{pp} on each output terminal where the output common mode voltage is 0.5 V.

We simulate the CS encoder by a sine wave at the input by Cadence. The amplitude and frequency of the input sine wave are 15 mV and 40 Hz, respectively. The sampling clock frequency is 2 kHz and the gain of integrator is normalized to one. The PRG is the designed maximal-length Fibonacci LFSR of Figure 5.1. The simulation result and the input signal are illustrated in Figure 5.7 .

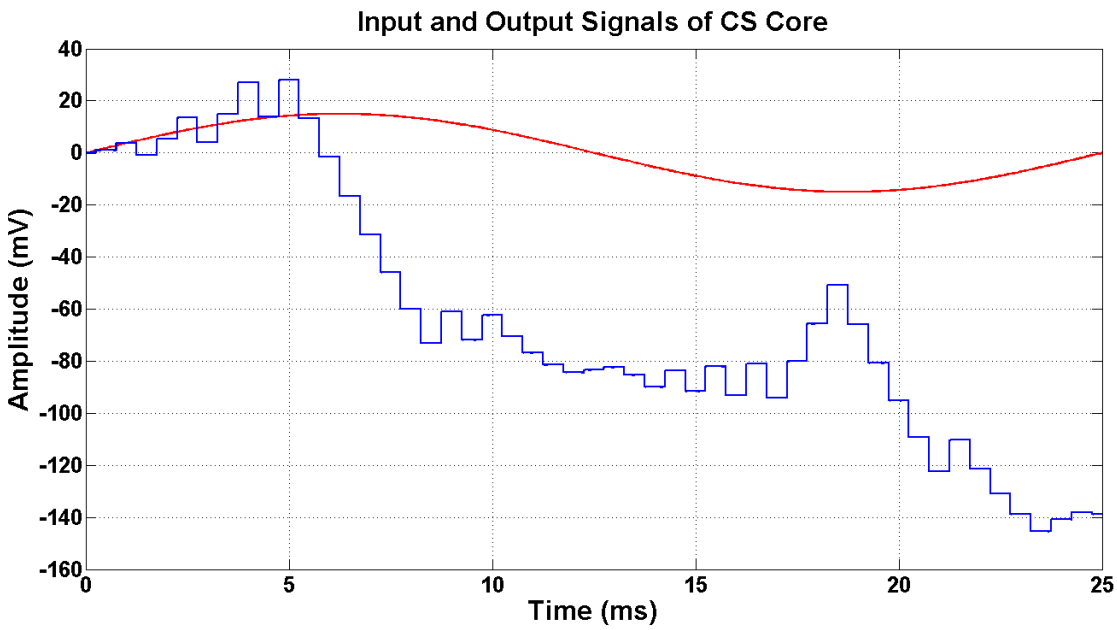


Figure 5.7 The input and output signals of CS encoder.

5.2 ADC Design

As we explained in section 2.5, the ADCs for neural recording application has not very tight specifications due to the low-frequency and narrow bandwidth of the neural signals. We also explained that SAR ADCs are more popular in the literature, due to their lower power consumption, simplicity of implementation and better specifications. Assuming maximum bandwidth of 10 kHz for extracellular neural actives, a 10-bit 30-MS/s SAR ADC can digitize 1000 channels through multiplexing the signals of each channel by an analog multiplexer. For other neural signals with lower bandwidth a SAR ADC with much relaxed specifications is required. Although the design of such SAR ADC is not challenging and some papers have even designed a conventional SAR ADC [33, 71], we prefer to design a 10-bit resolution SAR ADC with a monotonic capacitor switching procedure [166], to decrease the dynamic average switching power and total capacitance by around 81% and 50%, respectively, compared to conventional SAR ADC. We design this ADC so that we can alter the sampling frequency up to 50 MS/s and configure its resolution between 7 and 10 bit. The rest of this section has been organized to explain the SAR ADC architecture as well as designing its blocks.

5.2.1 SAR ADC Architecture

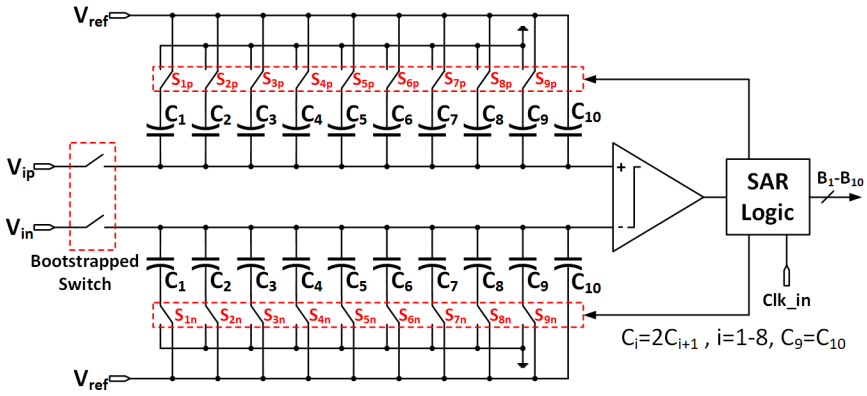


Figure 5.8 The utilized SAR ADC architecture.

We utilize a fully differential architecture to achieve 10-bit accuracy. Also, fully differential architecture eliminates supply and substrate noise as well as possessing appropriate common-mode noise rejection. Binary-weighted capacitor array are usually used for SAR ADCs, compared to C-2C capacitor due to better linearity. Considering the mentioned points, we chose the SAR ADC architecture of Figure 5.8, where the switching procedure can be downward or upward [166]. In this figure, the input neural signal are sampled on the top

plates of the capacitors by bootstrapped switches to achieve the appropriate linearity as well as increasing the settling speed. The bottom plates of the capacitors, at the same time, are reset to V_{ref}. Then, the bootstrapped switches are turned off and the first comparison is performed by comparator without switching any capacitor. Based on the comparator output, SAR logic block switches the largest capacitor, C_1, to ground on the higher voltage side and does not change the other one. This procedure is repeated by ADC until the LSB is extracted. As there is just one capacitor switch for each bit cycle, the charge transfer in the capacitor DAC network and the transitions of the SAR logic circuit are reduced. This results in lower power consumption. The calculation presented in [166] shows, the SAR ADC architecture of Figure 5.8 for 10-bit accuracy, decreases the switching energy and total capacitance 81% and 50%, respectively, in comparison with the conventional architecture.

5.2.2 S/H Circuit Design

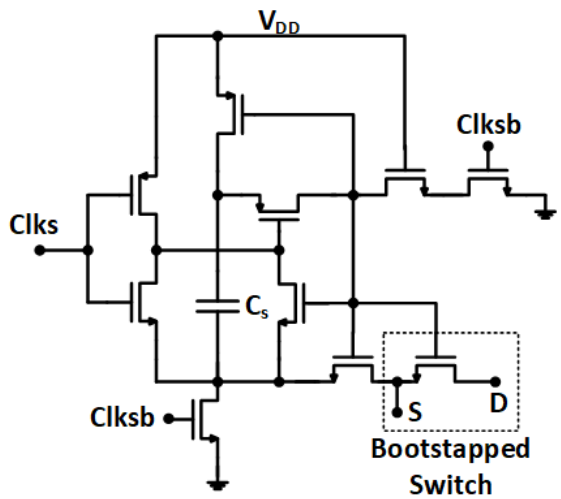
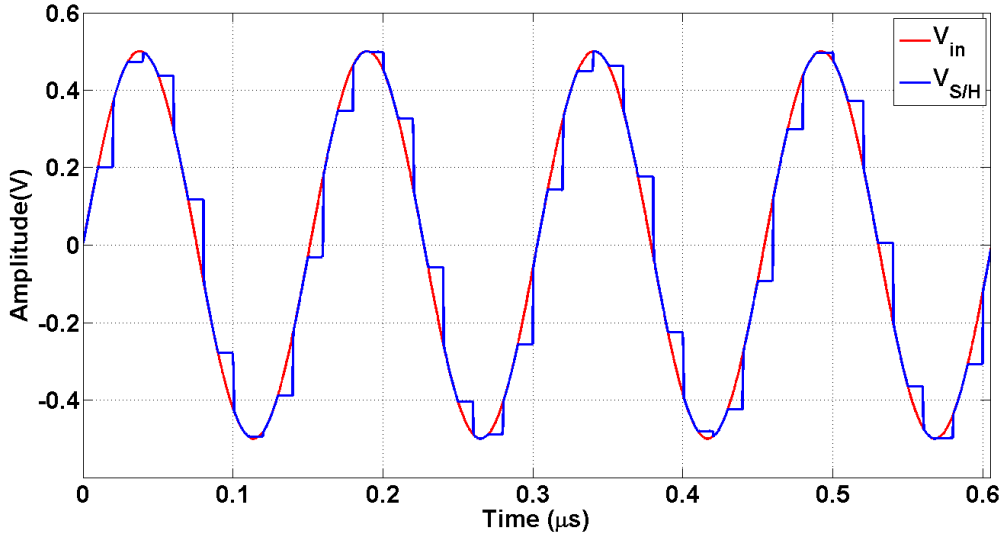
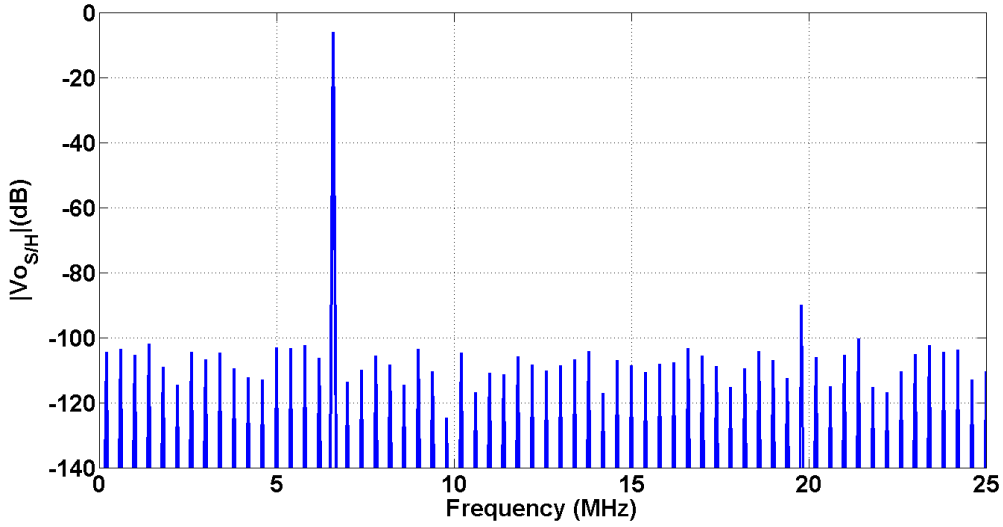


Figure 5.9 Bootstrapped switch.

A single NMOS and capacitor can operate as a sample and hold circuit, where the V_{DD} is high enough and the high linearity is not required. The NMOS switch is on by applying V_{DD} to the gate of this transistor. The drain-source resistance of this switch is calculated as (5.2). High input voltage swing varies the V_{GS} amount significantly and causes a large non-linearity. To overcome this limited linearity, a bootstrapped switch is utilized. The V_{GS} value of this sampling transistor is fixed at V_{DD} by using the bootstrapped switch. In this case, the on-resistance is a small constant value based on (5.2) and therefore, the switch linearity is improved.



(a)



(b)

Figure 5.10 Time and frequency analysis of the S/H signals. (a) Input and output signal of a S/H in time domain. (b) FFT result of the sampled sine wave signal.

$$r_{ds} = \frac{1}{\mu c_{ox}\frac{W}{L}(V_{GS} - V_{thn}) - V_{DS}} \tag{5.2}$$

Figure 5.9 shows the bootstrapped switch [166,167]. It operates on a single phase Clks, that turns the bootstrapped switch on and off. During the sampling phase, Clks is high and the capacitor C_s which is charged to V_{DD} is connected to the gate-source of the bootstrapped switch. In the hold phase, Clks is low and Clksb is high and the capacitor C_s is disconnected from bootstrapped switch and connected to V_{DD} and ground. Meanwhile the gate of the bootstrapped switch is connected to ground and the bootstrapped switch is off.

Figure 5.10 shows the time and frequency analysis of the designed S/H signals, simulated by Cadence for sampling capacitor, C_s, of 2.5 pF connected to each terminal. The input signal is a sine wave with the frequency of 6.6 MHz and the sampling clock frequency, Clks, is 50 MS/s. Figure 5.10(a) illustrates the input and sampled signals of the bootstrapped S/H circuit. The result of taking a 256-point FFT from the sampled input signal is shown in Figure 5.10(b). The calculating SNR amount of this figure in Matlab is 80 dB.

5.2.3 Dynamic Comparator

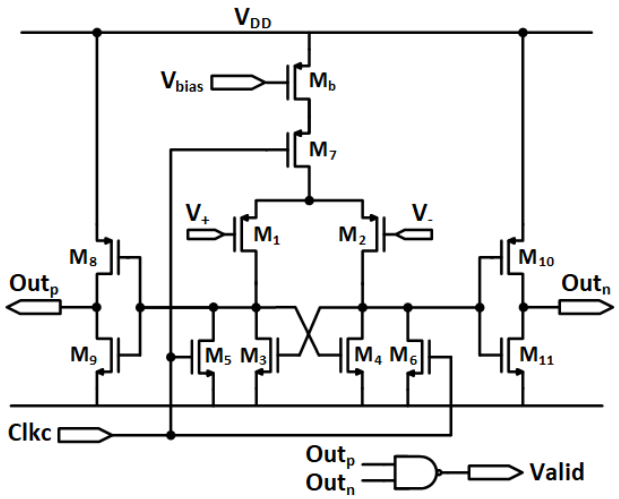


Figure 5.11 Dynamic Comparator.

High-speed comparators have one or two stages of preamplifier followed by a track-and-latch [168]. To minimize the power consumption, we avoid pre-amplification phase which consumes static current. Figure 5.11 shows schematic of dynamic comparator [166]. The input common-mode voltage of the comparator approaches from half V_{ref} to ground. Therefore, the PMOS differential pair is suitable and utilized in the comparator. In Figure 5.11, when Clkc, the comparator clock, is high Outp and Outn, the comparator outputs, go to high and the Valid signal is pulled down to ground. When Clkc changes and goes to low, M_5 and M_6 get off and M_7 becomes on as a switch and the comparator compares the two input voltages. The latch switches of M_3 and M_4 force one output to low and the other to high. The Valid signal which is nand of Outp and Outn signals goes to high to enable the digital block of asynchronous control logic.

In the hold phase of the S/H circuit, to extract 10 bits from sampled voltage, 10 successive comparison is performed by this comparator. Assuming the hold phase time is 80 % of

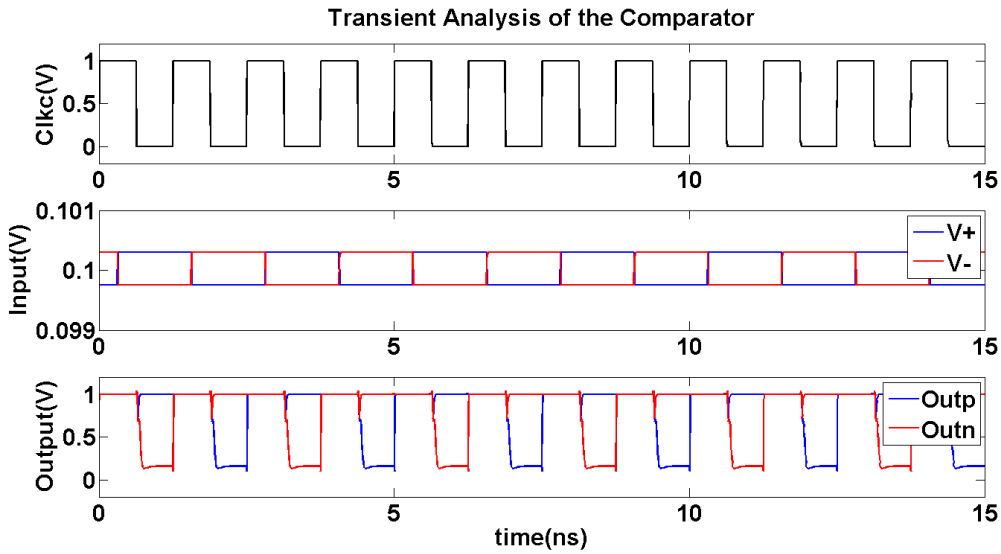


Figure 5.12 Transient analysis of comparator for an alternating input signal of ±0.25 mV.

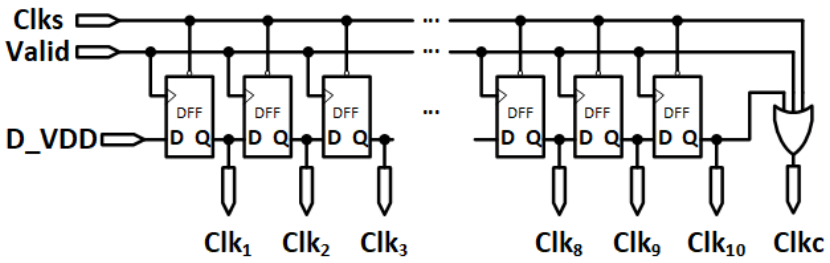
sampling period and the maximum sampling frequency is 50 MS/s, the maximum comparator clock, f_{Clkc}, is calculated as 625 MS/s. Figure 5.12 shows the transient analysis results of the comparator simulated by Cadence. In this figure, the input signal of the comparator alters between ±0.25 mV (less than 1 LSB) and the f_{Clkc} is 800 MS/s.

5.2.4 SAR Asynchronous Control Logic

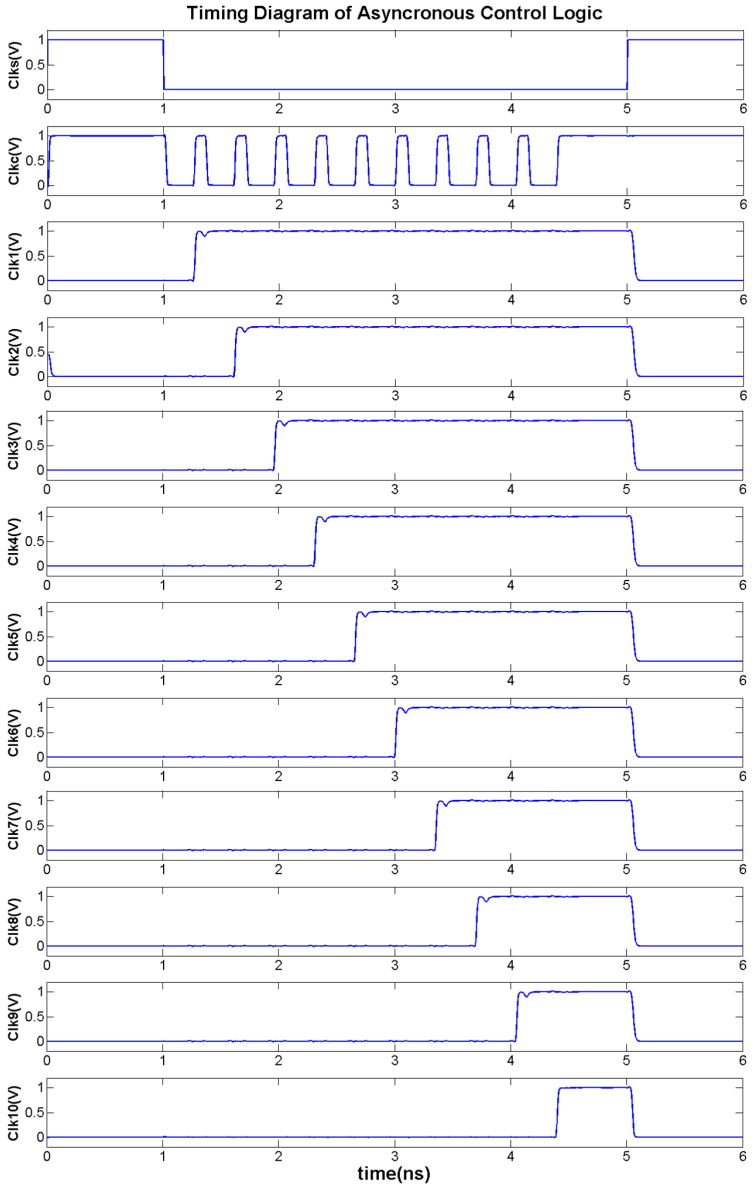
We exploit the SAR control logic of [166] to control the DAC switches. The ADC uses an asynchronous control circuit to avoid utilizing a high-frequency clock generator and to decrease the power consumption. This control logic block generates internally the necessary clock signal and by slightly modification of this circuit, we can configure the resolution of this ADC.

Figure 5.13 shows a schematic and the simulated timing diagram of the asynchronous control logic by Cadence. Figure 5.14 illustrates the DFF circuit with asynchronous reset used in Figure 5.13(a). As explained in section 5.2.3, the comparator generates the Valid signal. Clks is the sampling signal and turns on the bootstrapped switch when it goes high and turns off the sampling switches when it goes low. The sampling phase time is about 20% of the clock period. The control signal of the comparator is Clkc. The Valid signal enables the asynchronous control clock in rising edge. Clk1 to Clk10 are generated to sample the corresponding digital output codes of the comparator as well as controlling the DAC switches.

Schematic and timing diagram of the DAC control logic is shown in Figure 5.15. A static DFF samples the Outp signal at the rising edge of Clki. If the Outp is low, the corresponding



(a)



(b)

Figure 5.13 Asynchronous control logic for 10 bit SAR ADC. (a) Schematic. (b) Timing diagram simulated by Cadence.

capacitor switch is remained connected to V_{ref}. If the Outp is high, the corresponding capacitor switch is connected from V_{ref} to ground. All the capacitors are switched to V_{ref}, at the falling edge of Clki. There is a delay buffer in Figure 5.15 to insure Clki triggers the AND gate not before the output of the DFF. An inverter is used as a switch buffer. To minimize the delay of the buffer switches, we try to keep the RC value of them the same.

To do this, we design the switch buffers of the last three capacitors as the unit size and the first six switch buffers are sized based on the corresponding capacitances.

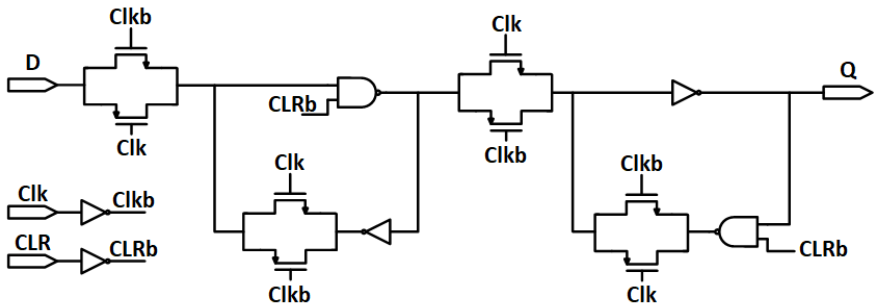


Figure 5.14 DFF circuit with asynchronous reset used in Figure 5.13(a).

To be able to configure the resolution of the ADC from 7 to 10 bits, we modify the asynchronous control logic of Figure 5.13 to Figure 5.16. In this figure the last Clki is multiplexed in the OR gate to generate Clkc. Input signals of S_0 and S_1 configure the resolution of the ADC. Table 5.1 shows the various logical values of S_0 and S_1 for relevant resolutions.

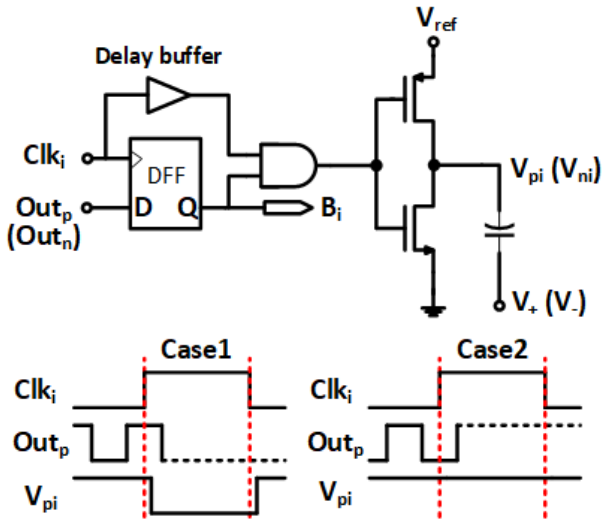


Figure 5.15 DAC control logic.

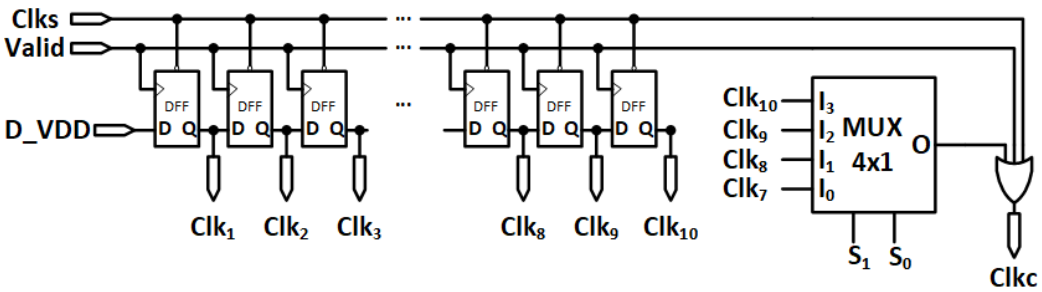


Figure 5.16 Modified asynchronous control logic to configure the resolution of the ADC.

To construct the capacitor array of Figure 5.8, we can use metal-oxide-metal (MOM) capacitors. Utilizing the multi-layer sandwich capacitor doubles the effective capacitor area [166]. Assuming the capacitance of the unit multi-layer sandwich capacitor as 5 fF, the total sampling capacitance of one capacitor array is about 2.5 pF.

Table 5.1 CONFIGURATION OF ADC RESOLUTION

S_1	S_0	**Mux output**	**Resolution(bit)**
0	0	Clk7	7
0	1	Clk8	8
1	0	Clk9	9
1	1	Clk10	10

CHAPTER 6 CONCLUSION

6.1 Contributions

Here, we review the contributions of this thesis briefly and link them to our corresponding published papers as follows:

- One of the critical problem in design of AC-coupled neural amplifiers in advanced technologies is the increase of the low-cutoff frequency. We investigated the problem comprehensively and found that the decrease in input resistance of the OTA is the origin of the increase in low-cutoff frequency. We proposed two different solutions to increase the input resistance of the OTA. Our proposed solutions were validated by fabricating a prototype exploiting the second method using the 65 nm TSMC CMOS Process.

The above contribution is detailed in the following published articles:

F. Hashemi Noshahr, M. Nabavi, and M. Sawan, "Low-cutoff Frequency Reduction Analysis of Neural Amplifiers in 65 nm CMOS", IEEE Transactions on Biomedical Circuits and Systems, (submitted to IEEE Symposium on Integrated Circuits and Systems, ISICAS 2020).

F. Hashemi Noshahr, and M. Sawan, "A compact and low power bandpass amplifier for low bandwidth signal applications in 65-nm CMOS", IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1-4, 2017.

- Modeling the CS encoder before designing it in circuit level helps designers minimizing the power consumption and silicon area of the CS encoder. We did this modeling in Matlab and Simulink to extract the required specifications of the OTA of the CS encoder. The simulation results showed that the SNR of the reconstructed signal in the receiver is very sensitive to the gain, bandwidth and output swing of OTAs but not to the slew rate. The extracted results were utilized in circuit design of the OTA.

The above contribution is reported in the following article:

F. Hashemi Noshahr, and M. Sawan, "Analog-based Compressive Sensing of Multichannel Neural Signals: Systematic Design Approaches", IEEE International Conference on Electronics, Circuits and Systems (ICECS), pp. 537-540, 2018.

- Design of voltage references with low sensitivity to power supply variations is challenging in advanced technologies due to short channel effects. We proposed and designed two new low power and compact voltage reference circuits suitable for neural implant applications in the 65 nm CMOS technology. The first design utilized an open-loop architecture while the second design exploited the feedback architecture. The feedback architecture was shown to yield the best results.

The details of this contribution as well as a comprehensive review on multi-channel neural recording implants can be found in the following articles:

F. Hashemi Noshahr, M. Nabavi, and M. Sawan, "A 2.77μW, 80 ppm/°C Temperature Coefficient Voltage Reference for Biomedical Implants", IEEE International Conference on Electronics, Circuits and Systems (ICECS), pp. 362-365, 2019.

F. Hashemi Noshahr, M. Nabavi, and M. Sawan, "Multi-Channel Neural Recording Implants: A Review", Sensors, Vol. 20, Issue 3,904, pp. 1-29, Feb. 2020.

6.2 Summary of Works

In this thesis we have investigated the design of a neural recording implant with analog compressive sensing. We explored the design of this system with systematic and circuit approach. In Chapter 1, we have expressed the importance of the neural recording implants in diagnosing and treatment of some brain diseases and disorders. Also, we discussed about the necessity of them in BMIs and controlling the artificial limbs and prosthesis. We have presented an interactive simplified block diagram which allows some parameters of the system to be configured.

We have reviewed the literature of the neural recording implants in Chapter 2. The neural signals and the various architectures of neural recording systems were presented first. Then we have discussed all blocks of the architecture. We have investigated the neural amplifiers thoroughly as one of the challenging blocks, in terms of topologies, techniques of achieving the necessary gain, IRN, and so on. At the end of this chapter, the ADCs and compression methods which have been utilized in the neural recording implants were stated.

The analog CS encoder design in system level has been focused in Chapter 3. To design an analog SC-integrator for implementing the CS core, an OTA is required. To extract the necessary gain, bandwidth, slew rate and output swing of this OTA to achieve the required output reconstructed SNR, Matlab simulations were done. Also, some simulations were completed to extract the required gain of the PGA. The simulation results of this chapter helped to optimize the specifications of the neural amplifier, CS core, PGA and the ADC.

As we explained in Chapter 2, conventional CFN architecture is the most appropriate topology for implementing neural amplifier. In advanced technologies, low-cutoff frequency in this architecture increases. In Chapter 4, we have proved the main cause of this problem is the decrease of input resistance of the OTA. Also we proposed two different solutions to increase it. The first solution which is called it cross-coupled positive feedback is suitable for the standard CMOS process and the second solution is to utilize thick-oxide MOS transistors in the input differential pair of the OTA to decrease the input leakage current and increase the input resistance of the OTA. We also presented a BGR without bipolar transistor in 65 nm process in Chapter 4. This BGR is compact and low power with low sensitivity to power supply variation.

We have designed a CS encoder and SAR ADC which have been presented in Chapter 5. CS encoder consists of CS core and PRG circuits. The system level design of Chapter 3 helped us to extract the specifications of the OTA utilized in CS core for output reconstruction SNR of around 20 dB. At the end, we have designed a SAR ADC with configurable resolution between 7 and 10 bits. The sampling frequency of the ADC is configurable as well, up to 50 MHz to maximize the number of the channels for digitizing.

6.3 Future Research

The defined project in Chapter 1 is a big system on a chip (SOC). Here, in this thesis, we have focused to solve the issues of neural recording which were explained in the previous chapters. Therefore, to complete this SOC, a consistent work still is required to be done. Brief list of future works are as follows.

As explained in section 3.5, to extract the values of M_S and M_P in CS core of Figure 3.11 for specific reconstruction SNR, Matlab simulations are required. The optimized value of M_S and M_P for various amount of SNR can be tabulated. Also based on the corresponding value of M_S and M_P for SNR of 20 dB (for instance), CS encoder is implemented. The blocks of CU, PMU and analog multiplexer are also necessary to be designed and implemented. At the end, all of the designed blocks and circuits should be assembled as a SOC. The whole SOC should be simulated and fabricated and finally, the prototype chip is verified by experimental tests.

BIBLIOGRAPHY

[1] A. M. R. Dixon, "Understanding the Practical Limitations of Applying Analog Compressed Sensing Systems to ECG Signals," Ph.D. dissertation, 2012.

[2] M. Mollazadeh, K. Murari, G. Cauwenberghs, and N. Thakor, "Wireless micropower instrumentation for multimodal acquisition of electrical and chemical neural activity," *IEEE transactions on biomedical circuits and systems*, vol. 3, no. 6, pp. 388–397, 2009.

[3] M. J. Cook, T. J. O'Brien, S. F. Berkovic, M. Murphy, A. Morokoff, G. Fabinyi, W. D'Souza, R. Yerra, J. Archer, L. Litewka *et al.*, "Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study," *The Lancet Neurology*, vol. 12, no. 6, pp. 563–571, 2013.

[4] A. B. Schwartz, X. T. Cui, D. J. Weber, and D. W. Moran, "Brain-controlled interfaces: movement restoration with neural prosthetics," *Neuron*, vol. 52, no. 1, pp. 205–220, 2006.

[5] E. Jovanov, A. Milenkovic, C. Otto, and P. C. De Groen, "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation," *Journal of NeuroEngineering and rehabilitation*, vol. 2, no. 1, p. 6, 2005.

[6] R. S. Istepanian, E. Jovanov, and Y. Zhang, "Guest editorial introduction to the special section on m-health: Beyond seamless mobility and global wireless health-care connectivity," *IEEE Transactions on information technology in biomedicine*, vol. 8, no. 4, pp. 405–414, 2004.

[7] J. Pandey and B. P. Otis, "A Sub-100 μW MICS/ISM Band Transmitter Based on Injection-Locking and Frequency Multiplication," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 5, pp. 1049–1058, 2011.

[8] F. Hashemi Noshahr, M. Nabavi, and M. Sawan, "Multi-Channel Neural Recording Implants: A Review," *Sensors*, vol. 20, no. 3, pp. 1–29, 2020.

[9] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. Van Der Smagt *et al.*, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, p. 372, 2012.

[10] J. L. Collinger, B. Wodlinger, J. E. Downey, W. Wang, E. C. Tyler-Kabara, D. J. Weber, A. J. McMorland, M. Velliste, M. L. Boninger, and A. B. Schwartz, "High-performance neuroprosthetic control by an individual with tetraplegia," *The Lancet*, vol. 381, no. 9866, pp. 557–564, 2013.

[11] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, p. 493, 2019.

[12] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, no. 7099, p. 164, 2006.

[13] W. Wang, J. L. Collinger, A. D. Degenhart, E. C. Tyler-Kabara, A. B. Schwartz, D. W. Moran, D. J. Weber, B. Wodlinger, R. K. Vinjamuri, R. C. Ashmore *et al.*, "An electrocorticographic brain interface in an individual with tetraplegia," *PloS one*, vol. 8, no. 2, p. e55344, 2013.

[14] T. Aflalo, S. Kellis, C. Klaes, B. Lee, Y. Shi, K. Pejsa, K. Shanfield, S. Hayes-Jackson, M. Aisen, C. Heck *et al.*, "Decoding motor imagery from the posterior parietal cortex of a tetraplegic human," *Science*, vol. 348, no. 6237, pp. 906–910, 2015.

[15] R. Yuste, "From the neuron doctrine to neural networks," *Nature reviews neuroscience*, vol. 16, no. 8, pp. 487–497, 2015.

[16] R. F. Yazicioglu, P. Merken, R. Puers, and C. Van Hoof, "A 200μW Eight-Channel EEG Acquisition ASIC for Ambulatory EEG Systems," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 3025–3038, 2008.

[17] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents—eeg, ecog, lfp and spikes," *Nature reviews neuroscience*, vol. 13, no. 6, p. 407, 2012.

[18] B. Pesaran, M. Vinck, G. T. Einevoll, A. Sirota, P. Fries, M. Siegel, W. Truccolo, C. E. Schroeder, and R. Srinivasan, "Investigating large-scale brain dynamics using field potential recordings: analysis and interpretation," *Nature neuroscience*, vol. 21, no. 7, pp. 903–919, 2018.

[19] T. Kaiju, M. Yokota, K. Watanabe, M. Inoue, H. Ando, K. Takahashi, F. Yoshida, M. Hirata, T. Suzuki *et al.*, "High spatiotemporal resolution ECoG recording of so-

matosensory evoked potentials with flexible micro-electrode arrays," *Frontiers in neural circuits*, vol. 11, p. 20, 2017.

[20] A. M. Obaid, M.-E. S. Hanna, Y.-W. Wu, M. Kollo, R. R. Racz, M. R. Angle, J. Muller, N. Brackbill, W. Wray, F. Franke *et al.*, "Massively parallel microwire arrays integrated with CMOS chips for neural recording," *bioRxiv*, p. 573295, 2019.

[21] E. Musk *et al.*, "An integrated brain-machine interface platform with thousands of channels," *Journal of medical Internet research*, vol. 21, no. 10, p. e16194, 2019.

[22] J. W. Clark, M. R. Neuman, W. H. Olson, R. Peura, F. Primiano, M. Siedband, J. Webster, and L. Wheeler, *Medical instrumentation: application and design.* Wiley, 1998.

[23] A. V. Nurmikko, J. P. Donoghue, L. R. Hochberg, W. R. Patterson, Y.-K. Song, C. W. Bull, D. A. Borton, F. Laiwalla, S. Park, Y. Ming *et al.*, "Listening to brain microcircuits for interfacing with external world—progress in wireless implantable microelectronic neuroengineering devices," *Proceedings of the IEEE*, vol. 98, no. 3, pp. 375–388, 2010.

[24] D. H. Hubel *et al.*, "Tungsten microelectrode for recording from single units," *Science*, vol. 125, no. 3247, pp. 549–550, 1957.

[25] R. Q. Quiroga and S. Panzeri, "Extracting information from neuronal populations: information theory and decoding approaches," *Nature Reviews Neuroscience*, vol. 10, no. 3, p. 173, 2009.

[26] P. G. Patil and D. A. Turner, "The development of brain-machine interface neuroprosthetic devices," *Neurotherapeutics*, vol. 5, no. 1, pp. 137–146, 2008.

[27] J. R. White, T. Levy, W. Bishop, and J. D. Beaty, "Real-time decision fusion for multimodal neural prosthetic devices," *PloS one*, vol. 5, no. 3, p. e9493, 2010.

[28] G. Buzsáki, "Large-scale recording of neuronal ensembles," *Nature neuroscience*, vol. 7, no. 5, p. 446, 2004.

[29] G. Gerstein and W. Clark, "Simultaneous studies of firing patterns in several neurons," *Science*, vol. 143, no. 3612, pp. 1325–1327, 1964.

[30] E. R. Kandel, J. H. Schwartz, T. M. Jessell, D. of Biochemistry, M. B. T. Jessell, S. Siegelbaum, and A. Hudspeth, *Principles of neural science.* McGraw-hill New York, 2000, vol. 4.

[31] A. M. Sodagar, K. D. Wise, and K. Najafi, "A wireless implantable microsystem for multichannel neural recording," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 10, pp. 2565–2573, 2009.

[32] B. Gosselin, A. E. Ayoub, J.-F. Roy, M. Sawan, F. Lepore, A. Chaudhuri, and D. Guitton, "A mixed-signal multichip neural recording interface with bandwidth reduction," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 3, no. 3, pp. 129–141, 2009.

[33] A. Bagheri, M. T. Salam, J. L. P. Velazquez, and R. Genov, "Low-frequency noise and offset rejection in DC-coupled neural amplifiers: A review and digitally-assisted design tutorial," *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 1, pp. 161–176, 2016.

[34] S. B. Lee, H.-M. Lee, M. Kiani, U.-M. Jow, and M. Ghovanloo, "An inductively powered scalable 32-channel wireless neural recording system-on-a-chip for neuroscience applications," *IEEE transactions on biomedical circuits and systems*, vol. 4, no. 6, pp. 360–371, 2010.

[35] B. Gosselin, M. Sawan, and E. Kerherve, "Linear-phase delay filters for ultra-low-power signal processing in neural recording implants," *IEEE transactions on Biomedical Circuits and Systems*, vol. 4, no. 3, pp. 171–180, 2010.

[36] D. L. Donoho *et al.*, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[37] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[38] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, 2007.

[39] R. R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE Journal of solid-state circuits*, vol. 38, no. 6, pp. 958–965, 2003.

[40] W. Wattanapanitch, M. Fee, and R. Sarpeshkar, "An energy-efficient micropower neural recording amplifier," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 1, no. 2, pp. 136–147, 2007.

[41] M. Azin, D. J. Guggenmos, S. Barbay, R. J. Nudo, and P. Mohseni, "A battery-powered activity-dependent intracortical microstimulation IC for brain-machine-brain interface," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 4, pp. 731–745, 2011.

[42] X. Zou, X. Xu, L. Yao, and Y. Lian, "A 1-V 450-nW fully integrated programmable biomedical sensor interface chip," *IEEE journal of solid-state circuits*, vol. 44, no. 4, pp. 1067–1077, 2009.

[43] R. Harrison, P. Watkins, R. Kier, R. Lovejoy, D. Black, R. Normann, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," in *2006 IEEE International Solid State Circuits Conference-Digest of Technical Papers*. IEEE, 2006, pp. 2258–2267.

[44] J. Lee, H.-G. Rhew, D. R. Kipke, and M. P. Flynn, "A 64 channel programmable closed-loop neurostimulator with 8 channel neural amplifier and logarithmic ADC," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 9, pp. 1935–1945, 2010.

[45] K. A. Ng and Y. P. Xu, "A compact, low input capacitance neural recording amplifier with C in/gain of 20fF. V/V," in *2012 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2012, pp. 328–331.

[46] P. Mohseni and K. Najafi, "A fully integrated neural recording amplifier with DC input stabilization," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 5, pp. 832–837, 2004.

[47] I. Obeid, M. A. Nicolelis, and P. D. Wolf, "A low power multichannel analog front end for portable neural signal recordings," *Journal of Neuroscience Methods*, vol. 133, no. 1-2, pp. 27–32, 2004.

[48] V. Chaturvedi and B. Amrutur, "An area-efficient noise-adaptive neural amplifier in 130 nm CMOS technology," *IEEE Journal on emerging and selected topics in circuits and systems*, vol. 1, no. 4, pp. 536–545, 2011.

[49] W. Zhao, H. Li, and Y. Zhang, "A low-noise integrated bioamplifier with active DC offset suppression," in *2009 IEEE Biomedical Circuits and Systems Conference*. IEEE, 2009, pp. 5–8.

[50] J. Ruiz-Amaya, A. Rodriguez-Perez, and M. Delgado-Restituto, "A low noise amplifier for neural spike recording interfaces," *Sensors*, vol. 15, no. 10, pp. 25 313–25 335, 2015.

[51] W.-S. Liew, X. Zou, L. Yao, and Y. Lian, "A 1-V 60-μW 16-channel interface chip for implantable neural recording," in *2009 IEEE Custom Integrated Circuits Conference*. IEEE, 2009, pp. 507–510.

[52] X. Zou, W.-S. Liew, L. Yao, and Y. Lian, "A 1V 22μW 32-channel implantable EEG recording IC," in *2010 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 2010, pp. 126–127.

[53] F. Zhang, J. Holleman, and B. P. Otis, "Design of ultra-low power biopotential amplifiers for biosignal acquisition applications," *IEEE transactions on biomedical circuits and systems*, vol. 6, no. 4, pp. 344–355, 2012.

[54] B. Gosselin, M. Sawan, and C. A. Chapman, "A low-power integrated bioamplifier with active low-frequency suppression," *IEEE transactions on biomedical circuits and systems*, vol. 1, no. 3, pp. 184–192, 2007.

[55] K. A. Ng and P. K. Chan, "A CMOS analog front-end IC for portable EEG/ECG monitoring applications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 11, pp. 2335–2347, 2005.

[56] R. F. Yazicioglu, P. Merken, R. Puers, and C. Van Hoof, "A 60 μw 60 nv/$\sqrt{}$hz readout front-end for portable biopotential acquisition systems," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 5, pp. 1100–1110, 2007.

[57] R. Muller, S. Gambini, and J. M. Rabaey, "A 0.013 mm^2, 5μW, DC-Coupled Neural Signal Acquisition IC With 0.5 V Supply," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 1, pp. 232–243, 2011.

[58] W. Biederman, D. J. Yeager, N. Narevsky, A. C. Koralek, J. M. Carmena, E. Alon, and J. M. Rabaey, "A Fully-Integrated, Miniaturized (0.125 mm^2) 10.5 μW Wireless Neural Sensor," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 4, pp. 960–970, 2013.

[59] R. R. Harrison, "The design of integrated circuits to observe brain activity," *Proceedings of the IEEE*, vol. 96, no. 7, pp. 1203–1216, 2008.

[60] R. H. Olsson, D. L. Buhl, A. M. Sirota, G. Buzsáki, and K. D. Wise, "Band-tunable and multiplexed integrated circuits for simultaneous recording and stimulation with microelectrode arrays," *IEEE transactions on biomedical engineering*, vol. 52, no. 7, pp. 1303–1311, 2005.

[61] S. Yuan, L. G. Johnson, C. C. Liu, C. Hutchens, and R. L. Rennaker, "Current biased pseudo-resistor for implantable neural signal recording applications," in *2008 51st Midwest Symposium on Circuits and Systems*. IEEE, 2008, pp. 658–661.

[62] M. Yin and M. Ghovanloo, "A low-noise preamplifier with adjustable gain and bandwidth for biopotential recording applications," in *2007 IEEE International Symposium on Circuits and Systems*. IEEE, 2007, pp. 321–324.

[63] A. M. Sodagar, G. E. Perlin, Y. Yao, K. Najafi, and K. D. Wise, "An implantable 64-channel wireless microsystem for single-unit neural recording," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2591–2604, 2009.

[64] M. S. Chae, W. Liu, and M. Sivaprakasam, "Design optimization for integrated neural recording systems," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 9, pp. 1931–1939, 2008.

[65] A. Rodríguez-Pérez, J. Ruiz-Amaya, M. Delgado-Restituto, and Á. Rodríguez-Vázquez, "An auto-calibrated neural spike recording channel with feature extraction capabilities," in *Bioelectronics, Biomedical, and Bioinspired Systems V; and Nanotechnology V*, vol. 8068. International Society for Optics and Photonics, 2011, p. 80680N.

[66] T. Denison, K. Consoer, W. Santa, A.-T. Avestruz, J. Cooley, and A. Kelly, "A 2.2 μW 94nV/$\sqrt{Hz}$ Chopper-Stabilized Instrumentation Amplifier for Chronic Measurement of Neural Field Potentials," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 12, pp. 2934–2945, 2007.

[67] W. Biederman, D. J. Yeager, N. Narevsky, J. Leverett, R. Neely, J. M. Carmena, E. Alon, and J. M. Rabaey, "A 4.78 mm 2 fully-integrated neuromodulation SoC combining 64 acquisition channels with digital compression and simultaneous dual stimulation," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 4, pp. 1038–1047, 2015.

[68] K. A. Ng and Y. P. Xu, "A low-power, high CMRR neural amplifier system employing CMOS inverter-based OTAs with CMFB through supply rails," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 3, pp. 724–737, 2016.

[69] H. Rezaee-Dehsorkh, N. Ravanshad, R. Lotfi, K. Mafinezhad, and A. M. Sodagar, "Analysis and design of tunable amplifiers for implantable neural recording applications," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 4, pp. 546–556, 2011.

[70] K. Abdelhalim and R. Genov, "Compact chopper-stabilized neural amplifier with low-distortion high-pass filter in 0.13 μm CMOS," in *2012 IEEE International Symposium on Circuits and Systems*. IEEE, 2012, pp. 1075–1078.

[71] K. Abdelhalim, L. Kokarovtseva, J. L. P. Velazquez, and R. Genov, "915-MHz FSK/OOK wireless neural recording SoC with 64 mixed-signal FIR filters," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 10, pp. 2478–2493, 2013.

[72] H. Kassiri, K. Abdelhalim, and R. Genov, "Low-distortion super-GOhm subthreshold-MOS resistors for CMOS neural amplifiers," in *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2013, pp. 270–273.

[73] K. Abdelhalim, H. M. Jafari, L. Kokarovtseva, J. L. P. Velazquez, and R. Genov, "64-channel UWB wireless neural vector analyzer SOC with a closed-loop phase synchrony-triggered neurostimulator," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 10, pp. 2494–2510, 2013.

[74] C. Qian, J. Parramon, and E. Sanchez-Sinencio, "A micropower low-noise neural recording front-end circuit for epileptic seizure detection," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 6, pp. 1392–1405, 2011.

[75] T. C. Carusone, D. A. Johns, and K. W. Martin, "Integrated-Circuit Devices and Modeling," *Analog Integrated Circuit Design, 2nd ed., United States: John Wiley & Sons*, pp. 42–44, 2012.

[76] M. S. Steyaert and W. M. Sansen, "A micropower low-noise monolithic instrumentation amplifier for medical purposes," *IEEE Journal of Solid-State Circuits*, vol. 22, no. 6, pp. 1163–1168, 1987.

[77] B. Gosselin, "Recent advances in neural recording microsystems," *Sensors*, vol. 11, no. 5, pp. 4572–4597, 2011.

[78] B. Razavi, *Design of analog CMOS integrated circuits*, 2005.

[79] M. Shoaran, M. H. Kamal, C. Pollo, P. Vandergheynst, and A. Schmid, "Compact low-power cortical recording architecture for compressive multichannel data acquisition," *IEEE transactions on biomedical circuits and systems*, vol. 8, no. 6, pp. 857–870, 2014.

[80] C. C. Enz and G. C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: autozeroing, correlated double sampling, and chopper stabilization," *Proceedings of the IEEE*, vol. 84, no. 11, pp. 1584–1614, 1996.

[81] K. Makinwa, "Dynamic-offset cancellation techniques in CMOS," *IEEE ISSCC Tutorial Session*, 2007.

[82] J. Xu, R. F. Yazicioglu, B. Grundlehner, P. Harpe, K. A. Makinwa, and C. Van Hoof, "A 160 μW 8 Channel Active Electrode System for EEG Monitoring," *IEEE Transactions on Biomedical circuits and systems*, vol. 5, no. 6, pp. 555–567, 2011.

[83] N. Verma, A. Shoeb, J. Bohorquez, J. Dawson, J. Guttag, and A. P. Chandrakasan, "A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system," 2010.

[84] T. Jochum, T. Denison, and P. Wolf, "Integrated circuit amplifiers for multi-electrode intracortical recording," *Journal of neural engineering*, vol. 6, no. 1, p. 012001, 2009.

[85] R. Wu, K. A. Makinwa, and J. H. Huijsing, "A chopper current-feedback instrumentation amplifier with a 1 mHz $1/f$ noise corner and an AC-coupled ripple reduction loop," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 12, pp. 3232–3243, 2009.

[86] F. Witte, K. Makinwa, and J. Huijsing, *Dynamic offset compensated CMOS amplifiers*. Springer Science & Business Media, 2009.

[87] A. Bakker, K. Thiele, and J. Huijsing, "A CMOS nested chopper instrumentation amplifier with 100 nV offset," in *2000 IEEE International Solid-State Circuits Conference. Digest of Technical Papers (Cat. No. 00CH37056)*. IEEE, 2000, pp. 156–157.

[88] J. C. Van Der Meer, F. R. Riedijk, E. van Kampen, K. A. Makinwa, and J. H. Huijsing, "A fully integrated CMOS Hall sensor with a 3.65 μT 3σ offset for compass applications," in *ISSCC. 2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference, 2005.* IEEE, 2005, pp. 246–247.

[89] M. A. Pertijs, K. A. Makinwa, and J. H. Huijsing, "A CMOS smart temperature sensor with a 3σ inaccuracy of$\pm$0.1°C from -55°C to 125°C," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2805–2815, 2005.

[90] K. A. Makinwa and M. F. Snoeij, "A CMOS Temperature-to-Frequency Converter With an Inaccuracy of Less Than $\pm 0.5°C$ (3σ) From $-40°C$ to $-105°C$," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 12, pp. 2992–2997, 2006.

[91] C.-C. Enz, "High precision CMOS micropower amplifiers," EPFL, Tech. Rep., 1989.

[92] C. C. Enz, E. A. Vittoz, and F. Krummenacher, "A CMOS chopper amplifier," *IEEE Journal of Solid-State Circuits*, vol. 22, no. 3, pp. 335–342, 1987.

[93] C. Menolfi and Q. Huang, "A fully integrated, untrimmed CMOS instrumentation amplifier with submicrovolt offset," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 3, pp. 415–420, 1999.

[94] ——, "A chopper modulated instrumentation amplifier with first order low-pass filter and delayed modulation scheme," in *Proceedings of the 25th European Solid-State Circuits Conference*. IEEE, 1999, pp. 54–57.

[95] A. Bilotti and G. Monreal, "Chopper-stabilized amplifiers with a track-and-hold signal demodulator," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 46, no. 4, pp. 490–495, 1999.

[96] Q. Huang and C. Menolfi, "A 200 nV offset 6.5 nV/$\sqrt{Hz}$ noise PSD 5.6 kHz chopper instrumentation amplifier in 1μm digital CMOS," in *2001 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC (Cat. No. 01CH37177)*. IEEE, 2001, pp. 362–363.

[97] T. Denison, K. Consoer, A. Kelly, A. Hachenburg, and W. Santa, "A 2.2 μW 94nV/$\sqrt{Hz}$, Chopper-Stabilized Instrumentation Amplifier for EEG Detection in Chronic Implants," in *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*. IEEE, 2007, pp. 162–594.

[98] Q. Fan, F. Sebastiano, J. H. Huijsing, and K. A. Makinwa, "A 1.8μw 60 nv/hz capacitively-coupled chopper instrumentation amplifier in 65 nm cmos for wireless sensor nodes," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 7, pp. 1534–1543, 2011.

[99] Y. Khazaei and A. M. Sodagar, "Multi-Channel ADC with Improved Bit Rate and Power Consumption for ElectroCorticoGraphy Systems," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2019, pp. 1–4.

[100] M. Nasserian, A. Peiravi, and F. Moradi, "An adaptive-resolution signal-specific ADC for sensor-interface applications," *Analog Integrated Circuits and Signal Processing*, vol. 98, no. 1, pp. 125–135, 2019.

[101] F. Pareschi, P. Albertini, G. Frattini, M. Mangia, R. Rovatti, and G. Setti, "Hardware-algorithms co-design and implementation of an analog-to-information converter for biosignals based on compressed sensing," *IEEE transactions on biomedical circuits and systems*, vol. 10, no. 1, pp. 149–162, 2015.

[102] P.-Y. Robert, B. Gosselin, A. E. Ayoub, and M. Sawan, "An ultra-low-power successive-approximation-based ADC for implantable sensing devices," in *2006 49th IEEE International Midwest Symposium on Circuits and Systems*, vol. 1. IEEE, 2006, pp. 7–11.

[103] R. Shulyzki, K. Abdelhalim, A. Bagheri, C. M. Florez, P. L. Carlen, and R. Genov, "256-site active neural probe and 64-channel responsive cortical stimulator," in *2011 IEEE Custom Integrated Circuits Conference (CICC)*. IEEE, 2011, pp. 1–4.

[104] J. Wang, Y. Hua, and Z. Zhu, "A 10-bit reconfigurable ADC with SAR/SS mode for neural recording," *Analog Integrated Circuits and Signal Processing*, vol. 101, no. 2, pp. 297–305, 2019.

[105] S. Barati and M. Yavari, "An adaptive continuous-time incremental ΣΔ ADC for neural recording implants," *International Journal of Circuit Theory and Applications*, vol. 47, no. 2, pp. 187–203, 2019.

[106] C. Qian, J. Shi, J. Parramon, and E. Sánchez-Sinencio, "A low-power configurable neural recording system for epileptic seizure detection," *IEEE transactions on biomedical circuits and systems*, vol. 7, no. 4, pp. 499–512, 2013.

[107] S.-G. Miaou and S.-N. Chao, "Wavelet-based lossy-to-lossless ECG compression in a unified vector quantization framework," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 3, pp. 539–543, 2005.

[108] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE/SP 14th Workshop Stat. Signal Process*, 2007, pp. 181–184.

[109] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[110] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *2007 IEEE International Symposium on Circuits and Systems*. IEEE, 2007, pp. 1959–1962.

[111] X. Chen, Z. Yu, S. Hoyos, B. M. Sadler, and J. Silva-Martinez, "A sub-Nyquist rate sampling receiver exploiting compressive sensing," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 3, pp. 507–520, 2010.

[112] M. Shoaran, M. M. Lopez, V. S. R. Pasupureddi, Y. Leblebici, and A. Schmid, "A low-power area-efficient compressive sensing approach for multi-channel neural recording," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*. IEEE, 2013, pp. 2191–2194.

[113] M. H. Kamal, M. Shoaran, Y. Leblebici, A. Schmid, and P. Vandergheynst, "Compressive multichannel cortical signal recording," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 4305–4309.

[114] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, 2012.

[115] Y. Suo, J. Zhang, T. Xiong, P. S. Chin, R. Etienne-Cummings, and T. D. Tran, "Energy-efficient multi-mode compressed sensing system for implantable neural recordings," *IEEE transactions on biomedical circuits and systems*, vol. 8, no. 5, pp. 0–0, 2014.

[116] X. Liu, H. Zhu, M. Zhang, A. G. Richardson, T. H. Lucas, and J. Van der Spiegel, "Design of a low-noise, high power efficiency neural recording front-end with an integrated real-time compressed sensing unit," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2015, pp. 2996–2999.

[117] J. Zhang, S. Mitra, Y. Suo, A. Cheng, T. Xiong, F. Michon, M. Welkenhuysen, F. Kloosterman, P. S. Chin, S. Hsiao *et al.*, "A closed-loop compressive-sensing-based neural recording system," *Journal of neural engineering*, vol. 12, no. 3, p. 036005, 2015.

[118] F. Chen, A. P. Chandrakasan, and V. Stojanović, "A signal-agnostic compressed sensing acquisition system for wireless and implantable sensors," in *IEEE Custom Integrated Circuits Conference 2010*. IEEE, 2010, pp. 1–4.

[119] X. Liu, M. Zhang, T. Xiong, A. G. Richardson, T. H. Lucas, P. S. Chin, R. Etienne-Cummings, T. D. Tran, and J. Van der Spiegel, "A fully integrated wireless compressed sensing neural signal acquisition system for chronic recording and brain machine interface," *IEEE Transactions on biomedical circuits and systems*, vol. 10, no. 4, pp. 874–883, 2016.

[120] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," *Network: Computation in Neural Systems*, vol. 9, no. 4, pp. R53–R78, 1998.

[121] A. Eftekhar, E. P. Sivylla, and G. C. Timothy, "Towards a next generation neural interface: Optimizing power, bandwidth and data quality," in *2010 Biomedical Circuits and Systems Conference (BioCAS).* IEEE, 2010, pp. 122–125.

[122] R. R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 4, 2003, pp. 3325–3328.

[123] Y. Perelman and R. Ginosar, "An integrated system for multichannel neuronal recording with spike/LFP separation, integrated A/D conversion and threshold detection," *IEEE Transactions on biomedical engineering*, vol. 54, no. 1, pp. 130–137, 2006.

[124] I. Obeid and P. D. Wolf, "Evaluation of spike-detection algorithms for a brain-machine interface application," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 905–911, 2004.

[125] A. Bonfanti, M. Ceravolo, G. Zambra, R. Gusmeroli, A. Spinelli, A. L. Lacaita, G. Angotzi, G. Baranauskas, and L. Fadiga, "A multi-channel low-power system-on-chip for single-unit recording and narrowband wireless transmission of neural signal," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology.* IEEE, 2010, pp. 1555–1560.

[126] M. Rizk, I. Obeid, S. H. Callender, and P. D. Wolf, "A single-chip signal processing and telemetry engine for an implantable 96-channel neural data acquisition system," *Journal of neural engineering*, vol. 4, no. 3, p. 309, 2007.

[127] A. M. Sodagar, K. D. Wise, and K. Najafi, "A fully integrated mixed-signal neural processor for implantable multichannel cortical recording," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 6, pp. 1075–1088, 2007.

[128] A. M. Haas, M. H. Cohen, and P. A. Abshires, "Real-time variance based template matching spike sorting system," in *2007 IEEE/NIH Life Science Systems and Applications Workshop.* IEEE, 2007, pp. 104–107.

[129] B. Gosselin and M. Sawan, "An ultra low-power CMOS automatic action potential detector," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 4, pp. 346–353, 2009.

[130] C. L. Rogers and J. G. Harris, "A low-power analog spike detector for extracellular neural recordings," in *Proceedings of the 2004 11th IEEE International Conference on Electronics, Circuits and Systems, 2004. ICECS 2004*. IEEE, 2004, pp. 290–293.

[131] C.-C. Peng, P. Sabharwal, and R. Bashirullah, "An adaptive neural spike detector with threshold-lock loop," in *2009 IEEE International Symposium on Circuits and Systems*. IEEE, 2009, pp. 2133–2136.

[132] P. T. Watkins, G. Santhanam, K. V. Shenoy, and R. R. Harrison, "Validation of adaptive threshold spike detector for neural recording," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2. IEEE, 2004, pp. 4079–4082.

[133] A. M. Kamboh, M. Raetz, K. G. Oweiss, and A. Mason, "Area-power efficient VLSI implementation of multichannel DWT for data compression in implantable neuroprosthetics," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 1, no. 2, pp. 128–135, 2007.

[134] B. Gosselin, S. S. Hosseini Khayat, A. Quotb, and M. Sawan, "Hardware implementation of wavelet transforms for real-time detection and compression of biopotentials in neural implants," *Current development in theory and applications of wavelets*, vol. 5, 2011.

[135] C.-C. Peng, Z. Xiao, and R. Bashirullah, "Toward energy efficient neural interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2697–2700, 2009.

[136] J. N. Aziz, K. Abdelhalim, R. Shulyzki, R. Genov, B. L. Bardakjian, M. Derchansky, D. Serletis, and P. L. Carlen, "256-channel neural recording and delta compression microsystem with 3D electrodes," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 3, pp. 995–1005, 2009.

[137] F. Hashemi Noshahr and M. Sawan, "Analog-based Compressive Sensing of Multichannel Neural Signals: Systematic Design Approaches," in *2018 25th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. IEEE, 2018, pp. 537–540.

[138] M. José and R. del Rio, *CMOS sigma-delta converters: Practical design guide*. John Wiley & Sons, 2013.

[139] L. A. Williams, "Modeling and design of high-resolution sigma-delta modulators," Ph.D. dissertation, Stanford University, 1993.

[140] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, "Analog-to-information conversion via random demodulation," in *2006 IEEE Dallas/CAS Workshop on Design, Applications, Integration and Software.* IEEE, 2006, pp. 71–74.

[141] I. H. Stevenson and K. P. Kording, "How advances in neural recording affect data analysis," *Nature neuroscience*, vol. 14, no. 2, p. 139, 2011.

[142] F. Sun, M. Morrell, and R. J. Wharen, "Responsive Cortical Stimulation for the Treatment of Epilepsy," *Neurotherapeutics*, vol. 5, no. 1, pp. 68–74, 2008.

[143] M. S. Fifer, S. Acharya, H. L. Benz, M. Mollazadeh, N. E. Crone, and N. V. Thakor, "Toward electrocorticographic control of a dexterous upper limb prosthesis: Building brain-machine interfaces," *IEEE pulse*, vol. 3, no. 1, pp. 38–42, 2012.

[144] A. M. Van Rijn, A. Peper, and C. Grimbergen, "High-quality recording of bioelectric events," *Medical and Biological Engineering and Computing*, vol. 29, no. 4, pp. 433–440, 1991.

[145] K. Najafi and K. D. Wise, "An implantable multielectrode array with on-chip signal processing," *IEEE Journal of Solid-State Circuits*, vol. 21, no. 6, pp. 1035–1044, 1986.

[146] C. C. Enz, F. Krummenacher, and E. A. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog integrated circuits and signal processing*, vol. 8, no. 1, pp. 83–114, 1995.

[147] R. F. Yazicioglu, S. Kim, T. Torfs, H. Kim, and C. Van Hoof, "A 30 μW Analog Signal Processor ASIC for Portable Biopotential Signal Monitoring," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 209–223, 2010.

[148] F. Hashemi Noshahr and M. Sawan, "A compact and low power bandpass amplifier for low bandwidth signal applications in 65-nm CMOS," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS).* IEEE, 2017, pp. 1–4.

[149] K. A. Ng and Y. P. Xu, "A compact, low input capacitance neural recording amplifier," *IEEE Transactions on biomedical circuits and systems*, vol. 7, no. 5, pp. 610–620, 2013.

[150] S. Song, M. J. Rooijakkers, P. Harpe, C. Rabotti, M. Mischi, A. H. van Roermund, and E. Cantatore, "A 430nW 64nV/vHz current-reuse telescopic amplifier for neural

recording applications," in *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2013, pp. 322–325.

[151] Z. Xiao, C.-M. Tang, C. M. Dougherty, and R. Bashirullah, "A 20μW neural recording tag with supply-current-modulated AFE in 0.13 μm CMOS," in *2010 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 2010, pp. 122–123.

[152] A. M. Sodagar and K. Najafi, "Extremely-wide-range supply-independent CMOS voltage references for telemetry-powering applications," *Analog Integrated Circuits and Signal Processing*, vol. 46, no. 3, pp. 253–261, 2006.

[153] M. Nabavi, M. Shams, and M. Sawan, "Temperature Independent Subthreshold Circuits Design," in *2018 International SoC Design Conference (ISOCC)*. IEEE, 2018, pp. 92–93.

[154] H. Banba, H. Shiga, A. Umezawa, T. Miyaba, T. Tanzawa, S. Atsumi, and K. Sakui, "A CMOS bandgap reference circuit with sub-1-V operation," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 670–674, May 1999.

[155] P. Kinget, C. Vezyrtzis, E. Chiang, B. Hung, and T. Li, "Voltage references for ultra-low supply voltages," in *2008 IEEE Custom Integrated Circuits Conference*. IEEE, 2008, pp. 715–720.

[156] B.-D. Yang, "250-mV supply subthreshold CMOS voltage reference using a low-voltage comparator and a charge-pump circuit," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 11, pp. 850–854, 2014.

[157] O. Abdelfattah, I. Shih, G. Roberts, and Y. Shih, "A 0.6V-supply bandgap reference in 65 nm CMOS," in *2015 IEEE 13th International New Circuits and Systems Conference (NEWCAS)*, June 2015, pp. 1–4.

[158] Y. Chen, J. Horng, C. Chang, A. Kundu, Y. Peng, and M. Chen, "18.7 A 0.7V, 2.% 3σ-Accuracy Bandgap Reference in 12nm CMOS," in *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, Feb 2019, pp. 306–307.

[159] U. Kamath, E. Cullen, T. Yu, J. Jennings, S. Wu, P. Lim, B. Farley, and R. B. Staszewski, "A 1 V Bandgap Reference in 7 nm FinFET With a Programmable Temperature Coefficient and Inaccuracy of 0.2% From -45°C to 125°C," *IEEE Journal of Solid-State Circuits*, pp. 1–11, 2019.

[160] F. Hashemi Noshahr, M. Nabavi, and M. Sawan, "A 2.77μW, 80 ppm/°C Temperature Coefficient Voltage Reference for Biomedical Implants," in *2019 26th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. IEEE, 2019, pp. 362–365.

[161] Y. Tsividis, *Operation Modeling MOS Transistor*. 2ed ed, Oxford University Press, 1999.

[162] J. Fellrath, "CMOS Analog Integrated Circuits Based On Weak Inversion Operations," *IEEE Journal of Solid-State Circuits*, vol. 12, no. 3, pp. 224–231, 1977.

[163] L. H. de Carvalho Ferreira and T. C. Pimenta, "A CMOS voltage reference for ultra low-voltage applications," in *2005 12th IEEE International Conference on Electronics, Circuits and Systems*. IEEE, 2005, pp. 1–4.

[164] T. L. Brooks and A. L. Westwick, "A low-power differential CMOS bandgap reference," in *Proceedings of IEEE International Solid-State Circuits Conference-ISSCC'94*. IEEE, 1994, pp. 248–249.

[165] P. Alfke, "Efficient shift registers, LFSR counters, and long pseudo-random sequence generators," *http://www. xilinx. com/bvdocs/appnotes/xapp052. pdf*, 1998.

[166] C.-C. Liu, S.-J. Chang, G.-Y. Huang, and Y.-Z. Lin, "A 10-bit 50-MS/s SAR ADC with a monotonic capacitor switching procedure," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 731–740, 2010.

[167] A. M. Abo and P. R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 599–606, 1999.

[168] D. A. Johns and K. Martin, *Analog integrated circuit design*. John Wiley & Sons, 2008.