UNIVERSITY OF SOUTHAMPTON

Faculty of Engineering and Physical Sciences

Electronics and Computer Science

Ultra-Fine Signal Classification Using Memristor-Enabled Hardware

by

Jiaqi Wang

https://orcid.org/0000-0002-2503-1001

A thesis for the degree of
Doctor of Philosophy

February 2023

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES
ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

ULTRA-FINE SIGNAL CLASSIFICATION USING MEMRISTOR-ENABLE HARDWARE

by Jiaqi Wang

Neural activity recording system promotes the development of diagnostic and therapeutic programs and neuroscience research. Direct recordings of neural signals from the brain have helped scientists access to study and unlock the secrets of neural coding gradually. This can be realised by applying implantable neural recording systems to monitor and record neural signals. Then, the neural information can be transmitted to the external device for processing, storage or application. However, the power consumption of the neural recording system is the primary constraint to monitoring large groups of neurons. It leads the development of neural recording systems in two directions: 'high-channel-count but wired' and 'wireless but low-channel-count'. To address the power issue, we proposed a neural front-end that aims to detect neural spikes by thresholding and output as one-bit digital data so that the afterwards processing can only work on spikes rather than processing all the data points. The most significant feature is that we induce memristors as trimming devices to tune the threshold voltage for spike detection. Meanwhile, it contributes to rejecting up to $50mV$ DC offset from electrodes. The measurement presents that the memristor-based pre-amplifier is capable of achieving above 95% spike detection accuracy with hundreds of nanowatt power consumption per channel. This design indicates a promising approach to conduct spike-detection on-chip with low power consumption and demonstrates the potential of a hybrid memristor/CMOS circuit for power-efficient large-scale neural interfacing application.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always give. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as journal papers and conference contributions in the list of publications.

Signed:..... Date:.....

]

Acknowledgements

I am grateful to those who have accompanied me on the path to my PhD.

My supervisors, Professor Themis Prodromakis and Alexantrou Serb, provided me with the invaluable opportunity to pursue a PhD. I would like to express my sincere gratitude to them. Over the years, I have appreciated your consistent guidance, encouragement, and patience. I would also like to thank my internal examiners, Professors Geoff Merrett and Koushik Maharatna, for their insightful and helpful comments on my work.

I would express appreciation to my colleagues at the University of Southampton: Shiwei Wang, Sachin Maheshwari, Spyros Stathopoulos, Yihan Pan and Patrick Foster. It has been a pleasure collaborating with you. I cherish the time we collaborated, brainstormed, discussed and shared our ideas, working on a single project. And I would like to thank my colleagues at Imperial College London: Christos Papavassiliou, Peilong Feng and Lijie Xie, who provided me with excellent assistance and convenience during my chip testing at Imperial College London.

I would like to express my heartfelt gratitude to my grandparents, parents, aunts and cousin. Their enduring affection and encouragement have been my primary source of support during my years abroad. I would like to express my most profound appreciation to my friends Franco and Vivian, as well as my cats Tyson and Oscar. Thanks for keep accompany with me.

Lastly, I express my gratitude towards The Royal Society for funding the industry fellow PhD student scholarship.

Chapter 1

Introduction

1.1 The Topic Area, Challenges and Motivation

Recording and monitoring neural signals using implantable microsystems is essential to understanding biological activity and functions. These can be utilised to advance diagnostic and therapeutic solutions [1] and overall neuroscience research [2]. From an application point of view, the ability to record and process extracellular neural signals can lead to novel therapeutic neuromodulation solutions for treating disease [3] and new opportunities in prosthesis [4]. For these purposes, high-density micro-electrode arrays (MEA) and recording systems are required to acquire and process a large number of neuronal signals simultaneously in real-time.

With the development of techniques in neuroscience, the micro-electrode arrays can access a more significant number of neurons chronically in the order of 10,000 [5; 6]. Scientists are developing higher-density implantable recording systems [7; 8; 9]. Besides, there is a strong commercial interest in such approaches, with Neuralink [10] proposing a package with 3072 electrodes and a small recording system that delivers amplification and digitisation capabilities. However, we are supposed to overcome several technical challenges to construct an extracellular neural signal recording with a high density of up to 10,000 channels. To give a better perspective, the prior arts, followed by the challenges and motivation of this research field given below.

Neural signal recording can be realised through the approaches shown in Figure 1.1 (in black). Such signals are transduced by a physical interface consisting of probes and electrodes [11; 12]. These raw neural signals generally range around $50 - 500\mu V$ in amplitude with $5 - 10\mu V$ background noise added on top and span around $0.1Hz$ to $10kHz$ in frequency [13]. The raw signals are passed to a front-end, which amplifies the signal to the milli-volt range. The amplified waveform is then filtered in a filter module, removing interference (e.g. power line interference) whilst passing Action Potentials (AP). After that, the signal is passed into the back-end, which refines the signal and generally attempts to

extract the computationally important information. One approach is to process the signal via analogue processing methods [9]. Alternatively, it can be digitised by an ADC and subsequently post-processed in the digital domain [7; 8]. After that, the neural signal can be transmitted off-chip for further processing and/or storage.

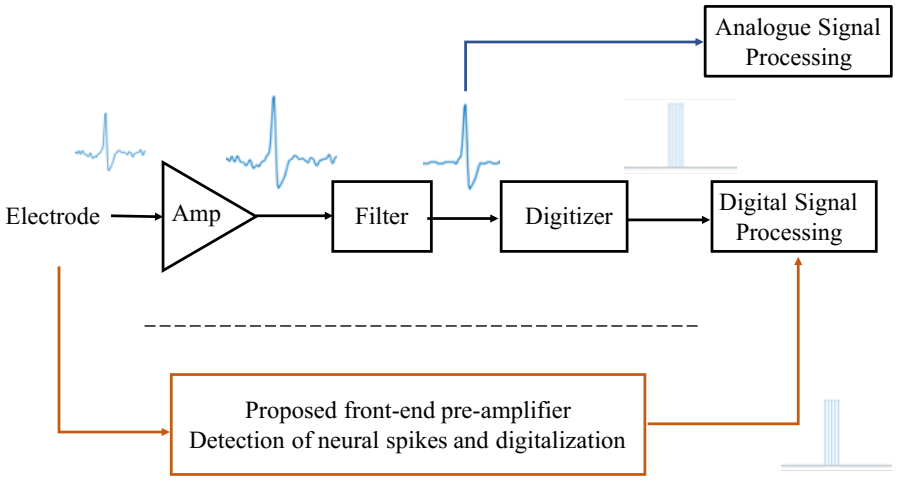


FIGURE 1.1: Structure comparison among the previous works in digital and analogue forms and our proposed work.

The state-of-the-art in-vivo systems consist of two schemes: 'high-channel count but wired' [7; 8; 9] and 'wireless but low-channel-count' recording system [14; 15; 16]. As for the wired system, the front-end channels are on the implanted chip. With tethered cables connected, the data processing module for high-channel-count is on the external device, which is able to process further computation. To the best of our knowledge, Neuralink [10] published a wired recording system that can achieve 3072 recording channels. However, the wire connection not only limits the mobility of observed animals, but also induces the risk of infection and device breakage [17]. The alternative solution, a fully wireless system, can overcome the above challenges. However, the high power/area budget of data transmission constrain scaling up the system. As for the wireless system, the number of recording channels is limited to around 100 [15]. It is because the transmission module occupies $20\% - 50\%$ of the power budget depending on the real-time transmission speed. To reduce power consumption, scientists focus on these three approaches: resource sharing, power scheduling and supply voltage reduction [3]. However, following the conventional path will limit them to enhance an order of magnitude. In this research, we target both wireless and high-channel-count recording systems.

The systems are required to be low-power to prevent damaging the surrounding tissues and extend the battery life [18]. The international guideline stated that the power consumption of the implantable neural interface should be confined within $15 - 40mW$ depending on the depth that the interface is implanted [19]. Chen et al. [20] have studied the state-of-the-art wireless neural systems and summarised that the transmission module occupies $20\% - 50\%$ of total power consumption depending on the volume of signals to be transmitted and

the system functionalities. The power constraint poses critical challenges to developing high-channel-count recording solutions, limiting the monitored neurons in the order of 100-1000 simultaneously [21].

As mentioned above, we set the maximum power consumption of the interface to $15mW$. To improve the power efficiency, we proposed a hybrid Complementary Metal–Oxide–Semiconductor (CMOS)/Memristor design that directly detects and classifies neural signals. It bypasses several steps in the conventional approach that brings in the possibility of building both wireless and high-channel-count neural interfaces (Figure 1.1). It is because we do not need to record the complete waveform or with high resolution of digitisation by taking advantage of the memristor-based neural signal processing, which is evidenced in the prior work [22]. Only 1-bit digital signals need to be transmitted off-chip; however, we target 10,000-channel data. Therefore, we estimated that data transmission account for 50% of the power budget which is $7.5mW$. Additionally, Fassio et al. [23] published an ultra-low-power biasing circuit that consumes picowatt-level power; thus, we omitted the power consideration of the biasing circuit temporarily. Therefore, the rest modules of our system consume $7.5mW$. In addition, we target a 10,000-channel system which indicates we need to constrain the power consumption of each channel to $750nW$. The detailed solution is given in the following section.

1.2 The Proposed Solution

This research presents an alternative paradigm by developing a front-end pre-amplifier concept that can operate in a start-stop scheme to achieve nanowatt-level power consumption per channel. This processing scheme conducts the amplification from microvolts to millivolts, spike detection and digitisation at the first stage (Figure 1.1 in orange). Thus, the requirement on afterwards modules is relaxed to processing digital signals representing neural spikes rather than the entire signal. This approach led to a design concept translating a train of analogue signals into digital, reducing the critical data points from the first stage.

The proposed front-end pre-amplifier consists of an integrating amplifier and a Dynamic Latch Comparator (DLC), conducting amplification and digitisation. As opposed to traditional practice where an 'expensive' amplifier is used to deliver high gain and low noise, here, we opt to integrate the microvolt-level signal to millivolts. The integration process not only filters out noise but also provides amplification; thus, we refer to this as an integrating amplifier. This fully differential system approach allows for integrating the difference between the input neural signal and a reference to a voltage level which triggers a threshold detector. Three operations (amplification, filtering and digitisation) can thus be realised within a single module. Further back-end processing is possible on the data emitted by our system in principle.

One of the most significant features is that the memristor is introduced as a trimming device to compensate for the offset voltage. Memristors are metal-oxide-metal devices whose Resistive States (RS) can be controlled by applying appropriate bias voltages. Memristor has been proven to be compatible with CMOS technique in terms of processes and materials [24; 25]. Memristor is constructed as a stack by the bottom electrode, active layer and top electrode, where both top and bottom electrodes can connect with peripheral and control CMOS circuitry, e.g. the drain/source of the transistor. In addition, memristors can be directly fabricated on the top of the CMOS circuits (above passivation) and connect to the transistor through Tungsten via, which increases area efficiency [26]. Being protected by passivation, the CMOS circuit will not be affected by both fabrication processes and materials of the memristor. We utilised the in-house $Pt/Al_2O_3/TiO_2/Pt$ memristor in our designs since it can achieve 92 distinct resistive states from $20k\Omega$ and $120k\Omega$, which can be programming in the precision of $1.09k\Omega$ [27]. Thus, we can tune the offset with the precision of the microvolt level by operating the memristor as a trimming device.

In addition, it can be used in various applications ranging from non-volatile memory [28], neuromorphic networks [29] to programmable resistances [30]. Its non-volatility, nanometer scale and CMOS-compatible processing [31; 32] add to the benefits of using memristors in our system as trimming devices. Such a passive trimming element adds no power dissipation in principle except when the device needs to be programmed to tune the circuit. Its high non-volatility features the architecture low maintain power consumption, compared with Dynamic Random-Access Memory (DRAM) approach that needs to be refreshed to preserve the data [33]. Besides, compared with the flash memory whose program and erase voltages are typical $15-21V$ [34], the low programming voltage (below $3V$ in [27]) simplifies the actual chip design. Incidentally, memristors have also been demonstrated to discriminate spiking from background activity [22] when used directly as sensing elements (alas on an already amplified waveform). It is a distinct approach that nicely complements this programme of research.

Throughout this work, memristors are operated under a 'write once – read many' regime where they are used to trim differential current paths and, as a result, tune the threshold (for spike detection) voltage of our system. Moreover, to program the device, a pair of single-transistor switches connecting the devices to an appropriate programming voltage and a trans-impedance amplifier (all shared across multiple channels) can perform the programming. In contrast, the main amplifier being programmed is shut down. Our research focuses exclusively on the function of the channel in regular operation, where the memristor resistance does not change. Therefore, this specific programming regime will not issue energy efficiency or complexity in our design.

1.3 Research Objectives

For this research program, we aim to construct a hybrid CMOS/Memristor neural interface that can be used for the real-time processing of neural signals. To achieve the aims of the research, the following objectives are identified:

1. Develop and validate power and area-efficient CMOS/Memristor neural interface circuitry that can directly detect and classify neural signals.
2. Fabricate and demonstrate that the developed circuitry (single channel) can achieve sub-microwatts average power consumption with an implementation in silicon.
3. Prove the proposed approach for large-scale neural interfacing application in the experiment and demonstrate that the developed chip can readout and process neural signals.

1.4 Contribution

The novel contributions presented in this dissertation are listed below.

1. Made the memristive Verilog-A quadratic model available to Cadence [35]. Previously, our group only had the exponential Verilog-A model under the $10 - 17k\Omega$ RS range, which limits the utilisation of memristor in circuit design. In [35], I converted the mathematical model from [36] into the Verilog-A model which owns a wider RS range from $20k\Omega$ to $120k\Omega$. The methodology of building the Verilog-A model refers to [37]. The Verilog-A model and methodology of integrating memristor into Cadence have been included in Chapter 3.

2. Established methodology for evaluating performance metric of dynamic pre-amplifier [38; 39]. The pre-amplifier implements the amplification by integrating load capacitors. Thus, it features a dynamic operating point, rendering traditional analysis (AC/DC) unsuitable. In this work, I analysed the transistor-level pre-amplifier and developed the methodology to complete the performance metric by combining transient simulation and periodic steady-state analysis. This benchmarking approach is finally leveraged for providing useful insights and design trade-offs of the memristor-based integrating amplifier in Chapter 4 and Chapter 5.

3. Studied the impact of memristor IV non-linearity on the pre-amplifier [40]. In this design, I calibrated RS measurements to the operating tail current. The measurement illustrates that the small deviations from tail current during integration do not materially change the extreme offset trimming precision or the overall performance. This was the last conceptual bottleneck identified before practical implementation, which has now been overcome. It can be accessed in Chapter 4.

4. Designed memristor-based DC-coupled pre-amplifier [41]. Instead of using huge capacitors at the input to reject the DC offset (AC-coupled solution), in this work, we utilised the memristive device to tune the offset in the wider range that can cover up to 50mV (DC-coupled solution). The full performance metric is elaborated in Chapter 6. The simulation results present that the DC-coupled pre-amplifier can achieve 95% detection accuracy while the power consumption is down to $\sim 120nW$. Circuit design and simulation details are given in Chapter 6

5. Designed memristor-based first-order ΔΣADC. The pre-amplifier operating as a conventional static amplifier can also be utilised in the ΔΣADC, which rejects up to $50mV$ DC offset by calibrating memristors. The designed transistor circuit can achieve 8.51 ENOB under the transient noise in the range of $[1Hz, 100MHz]$ (in Chapter 8).

1.5 List of Publications

Journal Articles

A. Serb, I. Kobyzev, J Wang and T. Prodromakis (2020) "A semi-holographic hyperdimensional representation system for hardware-friendly cognitive computing" *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378 (2164), 1-18. (doi:10.1098/rsta.2019.0162).

J. Wang, A. Serb, C. Papavassiliou, S. Maheshwari, and T. Prodromakis, (2021) "Analysing and measuring the performance of memristive integrating amplifiers" *International Journal of Circuit Theory and Applications*, 49(11), 3507-3525

S. Maheshwari, S. Stathopoulos, J. Wang, A. Serb, Y. Pan, A. Mifsud, L.B. Leene, J. Shen, C. Papavassiliou, T.G. Constandinou, and T. Prodromakis, (2021) "Design Flow for Hybrid CMOS/Memristor Systems—Part I: Modeling and Verification Steps" *IEEE Transactions on Circuits and Systems I: Regular Papers*, 68(12), 4862-4875

S. Maheshwari, S. Stathopoulos, J. Wang, A. Serb, Y. Pan, A. Mifsud, L.B. Leene, J. Shen, C. Papavassiliou, T.G. Constandinou, and T. Prodromakis, (2021) "Design Flow for Hybrid CMOS/Memristor Systems—Part II: Circuit Schematics and Layout" *IEEE Transactions on Circuits and Systems I: Regular Papers*, 68(12), 4876-4888.

Conference Proceedings

J. Wang, A. Serb, C. Papavassiliou, and T. Prodromakis (2021) "Accounting for Memristor IV Non-linearity in Low Power Memristive Amplifiers", *2021 IEEE International Symposium on Circuits and Systems* (*ISCAS*) (pp. 1-5). IEEE.

J. Wang, A. Serb, S. Wang, T. Prodromakis, (2022) "Hybrid CMOS/Memristor Front-End for Multiunit Activity Processing" accepted by *The IEEE International Symposium*

on Circuits and Systems (*ISCAS*).

J. Wang, A. Serb, S. Wang, T. Prodromakis, (2022) "Offset Rejection in a DC-Coupled Hybrid CMOS/Memristor Neural Front-End" accepted by *The IEEE International Symposium on Circuits and Systems* (*ISCAS*).

1.6 Thesis Organisation

The thesis is organised as follows. Chapter 2 presents literature on neural information processing. The introduction of applied memristive devices in terms of the operational behaviours and methodology of integration into Cadence presents in Chapter 3. Design concepts, the basic operation and transistor-level analysis of the pre-amplifier have been introduced in Chapter 4. The methodology for obtaining the full performance metric of the AC-coupled memristor-based pre-amplifier is given in Chapter 5, followed by the evaluation and design consideration. The pre-amplifier is also utilised to reject DC offset up to $50mV$ in the form of the open-loop circuit (in Chapter 6) and $\Delta\Sigma$ADC (in Chapter 8). The layout design of the single channel pre-amplifier and the testing result of the chip are included in Chapter 7. Finally, Chapter 9 contains the conclusion and the recommended future work.

Chapter 2

Neural Information Processing

The neural recording techniques allow scientists to investigate brain activity in depth and offer the potential to treat neuron disorders. The neural front-end should be designed after a thorough review of the signals of interest and the techniques employed by the various modules of the monitoring system. This chapter provides an overview of (i) characteristics of neural signals, (ii) electrode techniques related to signal acquisition, (iii) the state-of-the-art front-end designs (in the last five years) which feature a variety of offset rejection and signal processing techniques, and (iv) the primary reference design for the thesis.

2.1 Neural Signals

Neurons in the nervous system communicate with other nerve cells by passing electrical and chemical signals. A neuron consists of dendrites, cell body (soma), axon and axon terminals (Figure 2.1). Dendrites receive signals from other neurons in a neuron cell, while axon terminals send out messages. After dendrites are excited (by chemical signals), the cell body transmits information to the axon as an action potential [42]. Action potential transmission generates transmembrane potential, which the intracellular and extracellular neural recording can capture. By placing electrodes in the region of interest, intracellular and extracellular action potentials can be recorded [43]. Consequently, the action potential can be recorded as the potential difference between the electrode's tip and ground electrode [44].

The intracellular neural recording is a highly invasive technique that the transmembrane potential is measured by penetrating the membrane with the electrodes [45]. This technique is capable of capturing highly accurate information about the electrophysiology (e.g. a few ion channels) and obtaining a high amplitude of the intracellular action potential that is in the range of $70mV_{pp}$ [46]. However, due to the invasion of electrodes, this technique will cause irreversible damage to the neurons, so it can only be used for fundamental

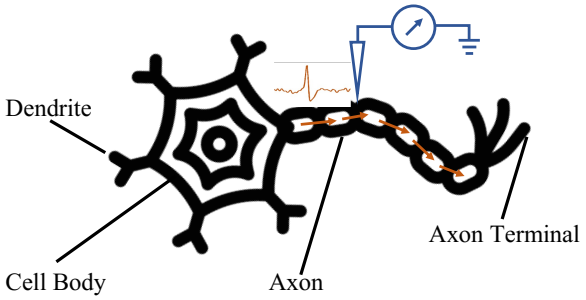


FIGURE 2.1: Dendrites, cell bodies, axons, and axon terminals make up a neuron. Moreover, the transmembrane potential of the neuron can be captured by the electrodes.

discoveries instead of medical interventions. As for the application, it is primarily applied for in-vitro measurement, while it is not appropriate for long-term or chronic recording [47].

In contrast, electrodes can be placed in the cortical tissue to record the extracellular action potential. This technique allows chronic monitoring of large groups of neural activities. However, compared to the extracellular recording, the recorded intracellular action potential has a weaker amplitude, ranging from $50\mu V_{pp}$ to $500\mu V_{pp}$ depending on the distance between the electrodes and the active neuron [48].

This thesis focuses on extracellular neuron recordings that are more appropriate for chronic monitoring. The recorded signals from the electrodes comprise extracellular action potential and Local Field Potential (LFP). The local field potential is generated by a group of neurons around the tip of the electrodes (50 - 350 μm) [49]. In addition, when the neuron is excited asynchronously, the action potential may be cancelled and the net field potential will be eliminated. Therefore, the LFP can be utilised to reflect the coordination of the population of neurons and the neural network dynamics. However, the contribution of individual spiking neurons to LFP has yet to be adequately described [45]. The LFP has a frequency of between 1mHz and 200Hz with an amplitude of $500\mu V_{pp}$ - $5mV_{pp}$.

As for the action potential, it contains both single-unit activity and multi-unit activity. The single-unit extracellular AP owns a frequency range from 100Hz to 10kHz. As for the multi-unit spiking, Choi et al. [50] obtained multi-unit activity by utilising a bandpass filter from 300Hz to 3000Hz. On the other hand, Stark et al. [51] acquired the multi-unit spiking from the root mean square of the signal in the frequency range of 300Hz to 6000Hz. In addition, Stolerman [52] only provided the vague frequency range of multi-unit spiking from 400Hz to a few kilohertz. Since the application of this design is to predict movement and epileptic seizure by counting the action potential that includes both single-

and multiple-unit spiking, we target the signal of interest in the frequency from 100Hz to 10kHz. The attributes of different neural signals are summarised in Table 2.1.

TABLE 2.1: Summary of the characteristics of different neural signal modalities.

Modality	Amplitude	Bandwidth	Potential for chronic recording
Intracellular action potential	10-70mV_{pp}	100-10kHz	low
Extracellular action potential	50-500μV_{pp}	100-10kHz	high
Local field potential	0.5-5mV_{pp}	1m-200Hz	high

2.2 Neural Electrode Techniques

Neural electrodes have significantly advanced the field of neuroscience by extracting and detecting neural signals. In addition, the extracellular action potential and local field potential are the main focus of this thesis. Spatial integration, stability and functionality are the three development axes for electrodes [53]. Consequently, we split the evaluation of innovative electrode approaches into three categories. Instead of describing each electrode in detail, we briefly describe each direction with a single example.

Spatial integration: interpreting the coordination of the brain demands a considerable increase in the number of monitored neurons. Moreover, this requires large arrays with multiplexed recording sites and minimum numbers of footprints [53]. The development of complementary metal–oxide–semiconductor (CMOS) technology features the probes with more compact multiplexing circuits (in shank) and I/O connections (in base) [54; 55]. In the thesis, we take the 'Neuropixels' as an example. The neuropixels 2.0 probe [56] contains four shanks and a base presented in Figure 2.2 with the dimensions. On the shank, 1280 electrodes are assembled, with each electrode containing an analogue

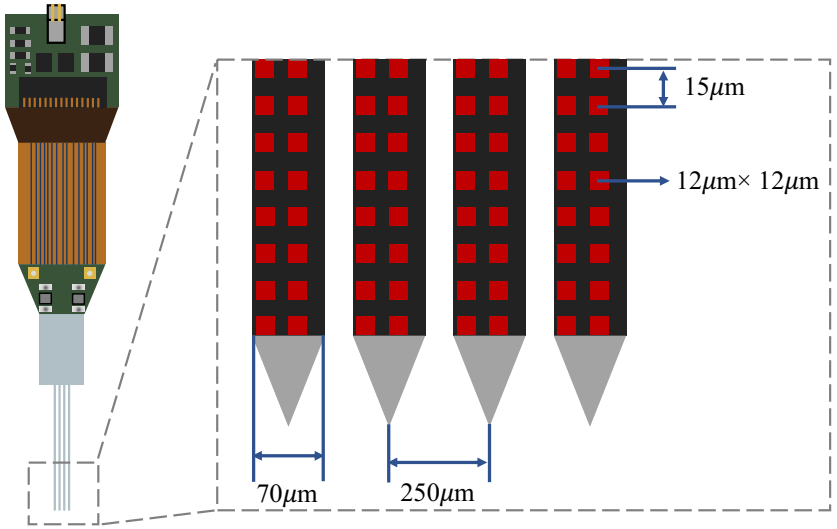


FIGURE 2.2: Architecture and dimensions of Neuropixels 2.0.

switch and 1-bit memory. The analogue switches determine which set of 384 electrodes is recorded concurrently. As for the non-implantable component, the base is fabricated by $130nm$ CMOS technology that comprises the electrical components responsible for multiplexing, amplification, digitisation, and power management units. Pre-processing the neural signals on the probe base prevents signal degradation during the data transmission from the probe to external devices. Neuropixels probe has been utilised to track the visual response in rats for up to two months.

Long-term stability: Access to time-dependent brain functions at the level of single neuron and the development of cognitive processes demands an implantable probe with long-term stability [57]. The interfaces should be featured with minimum perturbation to the brain tissues to achieve stable chronic monitoring. Therefore, scientists have concentrated on developing probes with significantly reduced mechanical stiffness than silicon or micro-wire probes [58; 59; 60]. To realise long-term recording, researchers (i) addressed the distance between the probe and brain tissues and (ii) designed the probes that 'appear' and 'behave' similarly to the tissues [53]. Mesh electronic probes were introduced to have similar maximum characteristic dimensions to individual neuronal somatic cells, with the same bending stiffness values as brain tissue (Figure 2.3). It has been applied for eight-month tracking of the age-related functional change of mouse neurons [60].

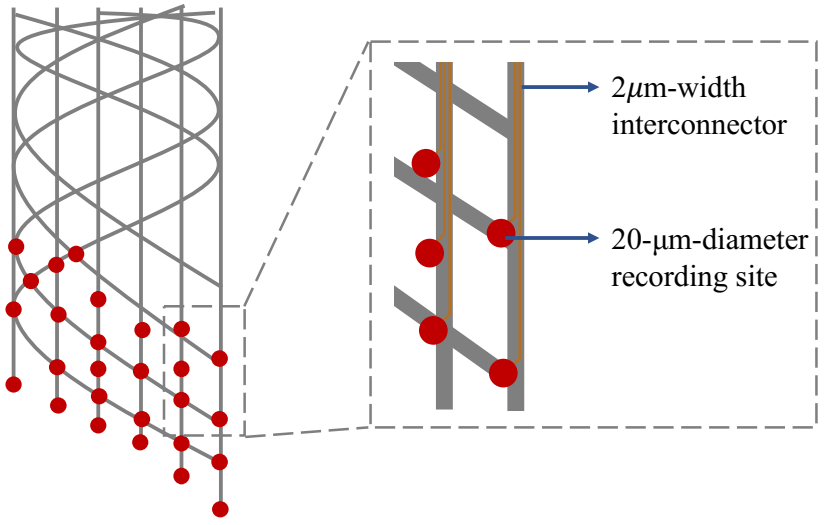


FIGURE 2.3: Architecture and dimensions of mesh electrodes.

Multiple functionalities: compared to the above electrodes, the multi-functional can achieve both recording the neural signals and stimulating the brain region. Multi-functional electrodes are mainly utilised to treat neurological disorders like Parkinson's disease and depression by real-time controlling and correcting the disorders [61]. For example, Shin et al. [62] proposed a multi-shank Micro-electromechanical systems (MEMS) electrode that consists of (i) recording sites, (ii) microfluidic channels for chemical delivery, and (iii) optical waveguides for optical stimulation in the brain. This functional integration is essential for accurately modulating neural circuits in-vivo. The architecture of the discussed electrode presents in Figure 2.4. The comparisons of the key characteristics of the

electrodes in three categories are presented in Table 2.2.

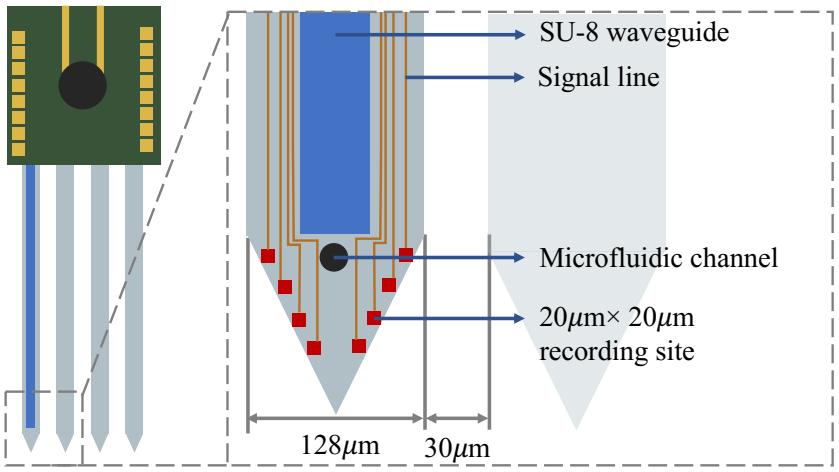


FIGURE 2.4: The architecture and dimensions of multi-shank electrodes, where SU-8 is a type of negative photoresist utilised as the core layer for waveguide.

TABLE 2.2: Comparisons of the novel electrodes.

Technology	**Multiplexity**	**Stability**
Neuropixels [56]	384 recording channels	data collection from the same neuron for more than two months
Mesh electrodes [60]	16-128 recording channels	tracking of the same neuron over eight months
multi-shank neural electrodes [62]	32 recording electrodes, 2-inlets microfludic mixer and SU-8 waveguides	causes litter immune responses over two weeks

After studying different categories of electrodes, we introduce the electrical model of the electrode-tissue interface that translates the neural activities into electrical signals [63]. The structure of the electrical model is in Figure 2.5 and the example data is shown in Table 2.3.

TABLE 2.3: Parameters of the electrical model from the measurement of different materials.

Parameter	***Pt* Electrode**	**I_rO_x Electrode**
V_{DL}	$15.77mV$	$3.12mV$
C_{DL}	$200pF$	$1.2nF$
R_{DL}	$10M\Omega$	$100M\Omega$
R_E	$5k\Omega$	$5k\Omega$

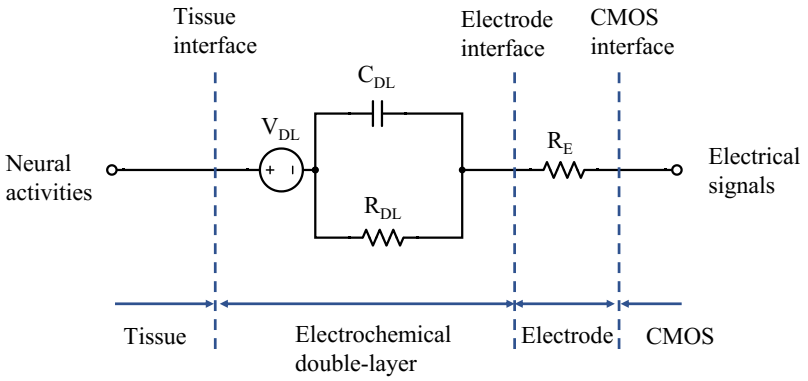


FIGURE 2.5: Electrical model of the electrode between the brain tissue and CMOS device, where V_{DL}, C_{DL} and R_{DL} are the double-layer potential, double-layer capacitance and resistance, respectively, and R_E is the resistance from the electrode to the CMOS device.

2.3 Neural Front-End

2.3.1 Overview of Neural Front-End

Neural signals from biology are transduced by a physical interface consisting of probes and electrodes. Typically, the amplitude of these raw neural signals ranges from 50 to $500\mu V$, with 5 to $10\mu V$ added background noise, and the frequency ranges from $0.1Hz$ to $10kHz$ [13]. The neural signal recording is typically realised through the approaches shown in Figure 1.1. Raw signals are passed to a front-end, which amplifies the signal to the millivolt range. The amplified waveform is then filtered band-pass filter, which eliminates interference (e.g. power line interference) while allowing action potentials to pass. After that, one approach is to pass the signal into back-end processing in the form of an analogue signal [64]. Alternatively, and indeed more typically, it can also be digitised by an ADC and subsequently post-processed in the digital domain [65]. The above belongs to front-end processing, which is presented in Figure 1.1 (black block diagram).

The signal is then transmitted to the back-end, which refines the signal and generally attempts to extract the computationally important information. As a rule, the back-end process consists of either spike detection or spike sorting. Spike detection involves determining whether a neural action potential (spike) occurs at any given time. The more complicated spike sorting process is then to also group detected spikes into clusters depending on the similarity of their shapes, i.e. to identify groups of spikes that are likely to originate from different neurons [66]. After processing in the back-end module, the neural signal will usually be transmitted off-chip for further processing and/or storage.

In digital processing systems, data integrity and power consumption increase as the ADC resolution is improved. Nevertheless, this approach has proved more prevalent and efficient than the alternative of processing the waveform in the analogue domain. The reason is primarily the ease of working with clean, digital signals when extracting information in

the back-end. Once the relevant information has been extracted, the digitally compressed neural signal waveform is significantly lower than the original analogue waveform.

Before reviewing the state-of-art neural front-end designs, we study the scheme of DC-offset rejection (AC and DC solutions) and the neural signal processing (spike detection and spike sorting) in the following.

2.3.2 Offset Rejection: AC-Coupled vs DC-Coupled Solution

One of the challenges of designing a neural front-end is to reject the DC offset from the electrodes. DC offset exists on the front-end's input due to the electrochemical effects at the electrode-tissue interface [67]. The electrical model is presented in Figure 2.5. The differential DC input is $1-10mV$ typically and can reach up to $50mV$. Such DC offset may saturate the differential amplifier. Various techniques have been proposed to overcome the input DC offset, including AC-coupled and DC-coupled solutions [68; 69; 70]. In the following, we study (i) the AC-coupled solution - closed-loop capacitive feedback network, (ii) chopper stabilisation, (iii) DC-servo loop and (iv) the DC-coupled solution by inducing Delta-Sigma ADC in the front-end.

Closed-loop capacitive-feedback network (CFN): This architecture in Figure 2.6 is firstly introduced by Harrison [71] and widely utilised in neural amplifier subsequently [72; 73]. In this topology, the resistor R_f is realised in MOS or bipolar element. From the frequency aspect, the large input capacitor C_i is utilised at the input to block the DC offset from tissue. The parallel design of capacitor C_f and resistor R_f in the feedback circuit operates as a high-pass pole: $\frac{-1}{R_f C_f}$. The OTA determines the low-pass pole, which falls in the range of $1Hz$. The huge capacitor of C_i is required to attain such a low cutoff frequency. The mid-band gain of this amplifier is $-\frac{C_i}{C_f}$. It leads to a large area, low mid-band gain and low input impedance of the amplifier.

In summary, the input capacitors (in the order of picofarad) occupy a sizeable on-chip channel area (in hundreds of μm^2) and prevent scaling up. Besides, the highly resistive triode-biased MOS transistor induces nonlinear behaviour in the presence of a large output voltage swing and DC operating point drift due to transistor leakage. This non-linearity causes distortion, process, voltage and temperature (PVT) variations which render the high-pass pole frequency time-variant.

Chopper-stabilised amplifier. The chopper is the continuous-time modulation technique in which signal and offset can be modulated at different frequencies [74; 75; 76]. In this review, we go through the principle of the chopper (in Figure 2.7) before inducing the chopper-stabilised ac-coupled amplifier (in Figure 2.8). The input voltage V_{in} is transformed to a square wave voltage at the frequency f_{ch} by passing through a chopper (see V_{in}). After initial signal modulation, the noise and offset are added to the modulated signal (see V_1). The modulated signal is then amplified with an input offset of its own.

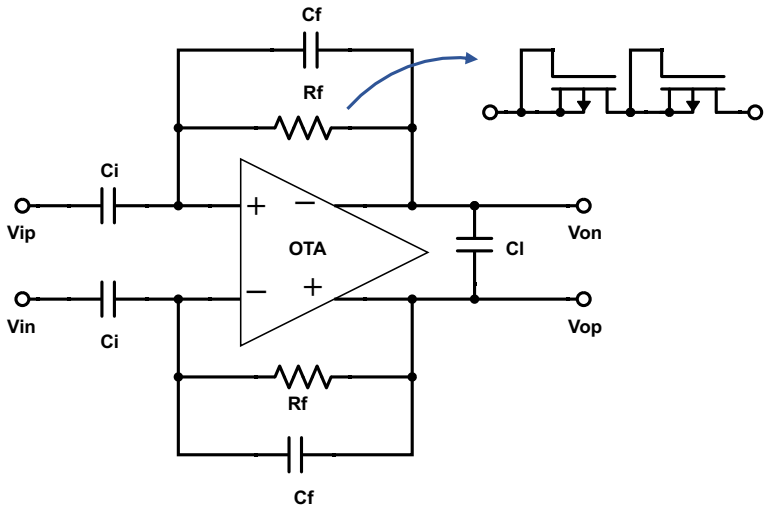


FIGURE 2.6: AC-coupled neural amplifier with capacitive-feedback network (CFN) topology. The resistor R_f is a MOS-bipolar pseudo-resistor.

After amplification and the second chopper, the modulated signal is demodulated back to DC. In contrast, the low-frequency noise and offset are modulated to the harmonics of the rotor frequency, resulting in a chopper ripple at the amplifier output (see V_2). A LPF is then applied to remove offset and 1/f noise from the modulated low-frequency signal (see V_{out}).

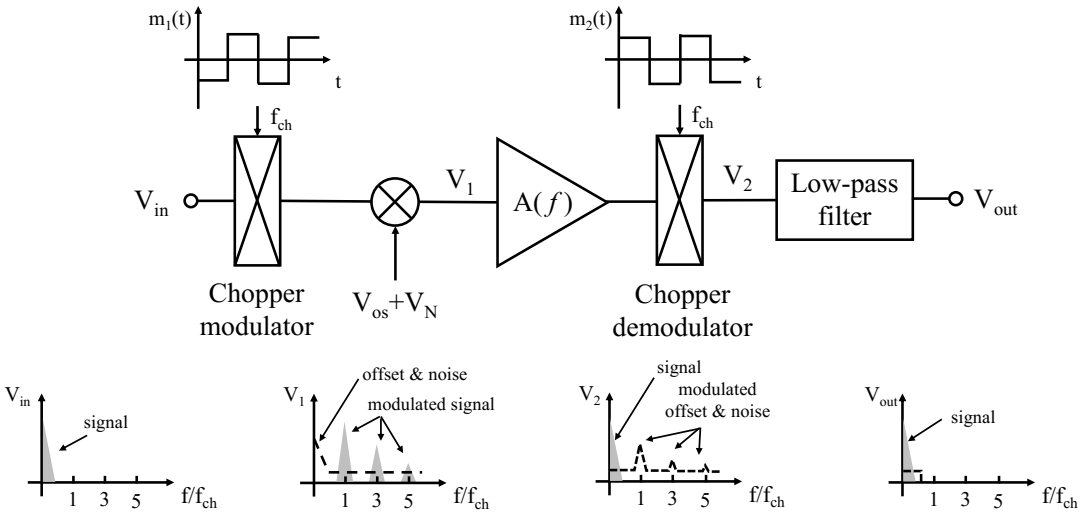


FIGURE 2.7: The general block diagram of the chopper shows the principles and signal transmission in the frequency domain.

On the other hand, the chopper often comes with the ac-coupled amplifier in the neural front-end that is referred to as a chopper-stabilised amplifier (in Figure 2.8). Compared to a pure chopper that employs a low-pass filter, the chopper-stabilised amplifier feeds the demodulated output back to the input. The amplifier then eliminates the low-frequency errors. Compared to the closed-loop capacitive-feedback network, the chopper-stabilised amplifier is capable of achieving considerable noise and offset reduction. Nonetheless,

this architecture still utilises extra large input capacitors and requires extra techniques to boost the input impedance.

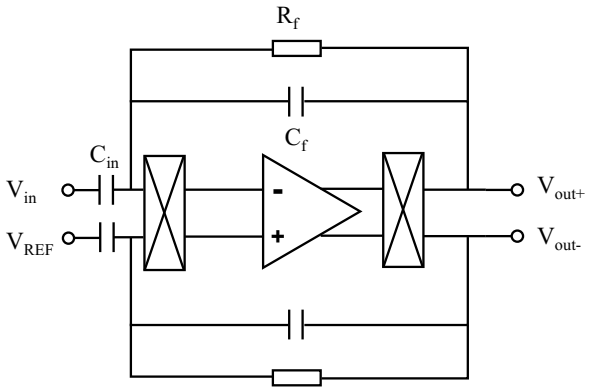


FIGURE 2.8: The conventional chopper-stabilised neural amplifier.

DC servo loop (DSL): The general schematic of the DC servo loop in Figure 2.9 illustrates the principle of DC offset cancellation. This system is capable of measuring the output voltage V_o and comparing it to the intended DC value V_{REF}. The voltage difference is then sent back to the input through an integrator, yielding a DC output voltage equal to the desired DC level (V_{REF}). The transfer function is

$$V_o(s) = V_i(s)\frac{s}{1+s/Gm} + V_{REF}(s)\frac{1}{1+s/Gm} \tag{2.1}$$

This architecture is able to eliminate DC components at the input by incorporating a high-pass filter for V_i and an integrator with a substantial time constant [77]. And it functions as a lowpass filter for the V_{REF}, which establishes the DC output level. The conventional DC-Coupled technique is to induce a DC servo feedback loop in order to eliminate the DC offset [70]. Pham et al. [70] lowers the large input capacitance by inducing additional capacitors, resistors, OTAs and chopper stabilisation modules into the DC servo loop. However, the opamp in the feedback circuit significantly increases the power consumption, and its various open-loop gain results in a variable high-pass pole. Even though it is effective at removing DC offset, it can be only applicable to systems with a restricted number of channels. This is because such a system requires substantial passive components to generate a low-frequency high-pass pole.

Delta-Sigma modulator (DSM): The delta-sigma modulator has become the mainstream front-end for achieving low noise level [78]. The block diagram of the one-bit delta-sigma modulator is in Figure 2.10. The analogue signal at the input will be sampled by multiple times, which is known as oversampling. Since the sampling rate is multiple times faster than the digital output, the individual sample is accumulated over time and then averaged with other input samples. The feedback loop in the modulator operates different functions to desired signal (signal transfer function - STF) and the noise (noise

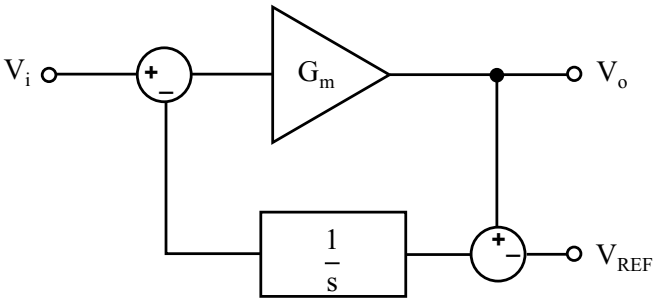


FIGURE 2.9: General schematic of DC servo loop.

transfer function - NTF):

$$V_o = V_i \frac{1}{s+1} + N \frac{s}{s+1} \tag{2.2}$$

It can be obtained that the integrator operates as a low-pass filter for input signals (V_i) and works as a high-pass filter for the quantisation noise (N) by summing the error voltage. The quantisation noise is spread to a higher frequency range, and it can be removed by a digital filter afterwards, which achieves low noise at the final output. The purpose of the feedback loop is to maintain the average output of the integrator near the reference level of the comparator, which averages the DC voltage rather than filtering it out. Thus, the delta-sigma modulator typically induces a chopper stabilisation module to cancel out the DC offset variation [79].

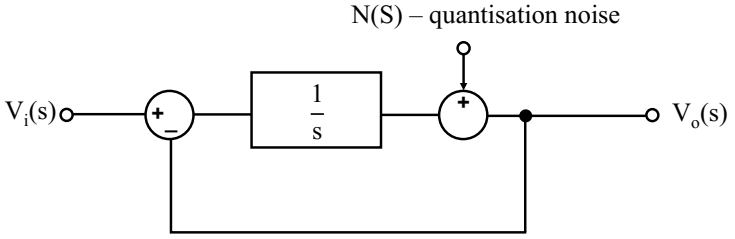


FIGURE 2.10: One-bit delta-sigma modulator in s-domain.

2.3.3 Neural Signal Processing: Spike Detection vs Spike Sorting

As the amount of brain data captured using microelectrodes increases, the transmission of the recorded information poses a growing challenge for microelectronics. The bandwidth of communication for implantable devices is fundamentally governed by the energy budget and total amount of heat dissipation permitted to prevent tissue injury. These constraints restrict the wireless transmission of such a large quantity of data. Consequently, it is crucial to identify feasible routes for performing a portion of signal processing on-chip while adhering to a minimal power budget. In this section, we review two processing schemes, including spike detection only and spike sorting.

Spike Sorting: The raw data collected from neurons can be effectively filtered into local field potentials (which reflect the dynamics of neural tissues around the electrodes and are generated by the input currents of the dendrites of the surrounding neurons) and 'spike trains' [80]. The effect of filtering is to make spikes in background noise visible. The filtered data is then separated into 'spikes' and 'background noise' (referred to as spike detection) using methods such as establishing user-defined thresholds. The spike shape is affected by a number of factors, such as the distance and orientation of the recording electrode from the neuron [81]. However, to classify the shapes, several features, such as spike amplitude and spike width, are selected, and based on these features, the spikes are classified into 'clusters' [82]. Each cluster is then linked to a single neuron, while clusters that cannot be recognised are linked to 'multi-unit' activity (typically characterised by a lower amplitude) [66]. In summary, the flowchart for spike sorting consists of an alignment of detected spikes, feature extraction, clustering, and classification, which is the identification of clusters with original neurons.

Spike Detection: Spike detection is an essential technique that distinguishes spikes (real brain activity/events) from background noise. Setting a threshold is the most typical method for performing this operation. The threshold setting can be fixed or adaptable. Fixed threshold utilises a constant circumstance throughout the duration of the recording. The adaptive threshold strategy, on the other hand, monitors background changes and modifies the threshold value accordingly. In any case, the threshold value must be selected with care because, if the threshold value is too low, noise fluctuations may produce 'false positive' events, and if the threshold value is too high, spikes with low amplitudes will be missed [83]. In the early days [84] of spike detection, the thresholds were manually set by the user. Whenever the voltage signal exceeded the threshold, a pulse would be created to represent an increase. If a spike waveform was required, specific user-specified points would be collected after each threshold.

Selection: The decision between spike sorting and spike detection can be influenced by the following factors:

- Application: not all applications require precise spike sorting to reflect neuronal activities. For instance, Kloosterman et al. [85] state that the signal 'read' by cuff electrodes derives from the multi-unit activity and that identification of individual spikes is not necessary. A measure of overall activity over time frames longer than individual spikes suffices.

- Design constraints and trade-offs: It is of great interest to the neuroscience community to implement spike sorting algorithms on a low-power device in order to enable wireless transmission of data and to provide ecologically and safely secure settings for animal research. Wendler et al. [86] highlight the trade-offs that must be addressed when creating a neural interface for the future generation. With the consideration of the power budget of the system, it will be necessary to weigh the

application and scheme trade-offs carefully.

2.3.4 State-Of-the-Art Neural Front-Ends

The comparison among the state-of-the-art neuron front ends (from 2018 to 2022) and the specification of our work is presented in Table 2.4. The majority of offset rejection has shifted from AC-coupled to DC-coupled solutions in order to increase the recording channel within a limited power and area budget. In the following sections, we reviewed the AC-coupled front-end from [69] and a DC-coupled solution from [68] as examples.

AC-coupled neural SoC: it includes the functions of recording and detecting spikes [69]. The analogue pixel comprises AC-coupled capacitive-feedback amplifiers, including low-noise amplifiers (LNA) and variable gain amplifiers (VGA) and 10-bit SAR ADCs for neural recording. Besides, the local digital processor conducts spike detection on-chip. The architecture is presented in Figure 2.11 and the detailed performance is listed in Table 2.4. The input AC-coupling capacitors are utilised to prevent the DC offset and provide the proper DC level to the input transistor of LNA in order to bias it in the correct operation region. In the feedback circuit, the pseudo-resistor determines the high-pass filter cutoff frequency and the DC input bias of the pixel. The tunable pseudo-resistor allows the cutoff frequency of the HPF for LFP (5 Hz) and AP to be adjusted (300 Hz), which features the system with higher flexibility. However, compared with other works with delta-sigma modulation [68; 87], this offset rejection method still occupies a larger channel area (shown in Table 2.4).

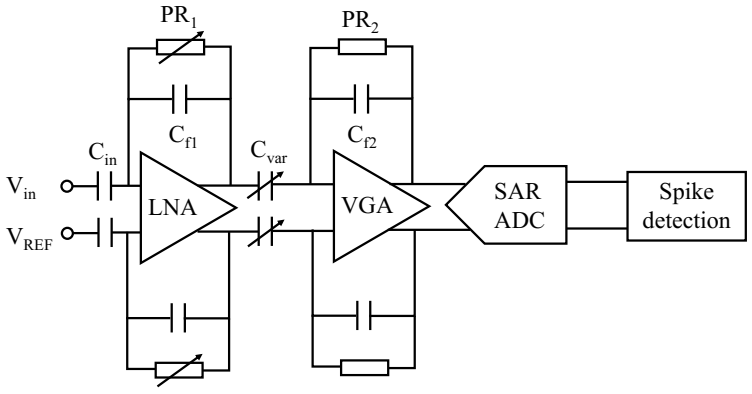


FIGURE 2.11: AC-coupled capacitive-feedback front-end.

Delta-sigma ADC: This front-end is fabricated under the electrode within an area of $70 \times 70\mu m^2$ is proposed in [68]. The presented first-order incremental $\Delta\Sigma$ ADC utilises Gm-C integrator and current feedback. The architecture and its performance are shown in Figure 2.12 and Table 2.4 respectively. The weak signal without pre-amplification does not require strict linearity, and thus, it allows us to utilise the Gm-C technique at the input, which can achieve higher area efficiency. As for the feedback path, the modulator chooses the current domain, which can avoid adding extra input transistor pair into the

system. In order to reduce the area, this front-end enables three scale modes of ADC that can cover the full-scale of ±11.25mV, ±22.5mV and ±45mV. The simple 1-bit ΔΣ ADC with Gm-C input utilises the three different current sources to accommodate a wider range of DC input voltage and occupies less area. In order to compensate for the DC offset, it utilises a dynamic reference voltage V_{REF} that is controlled by the digital module and can cover the offset of up to $\pm 200mV$. It inspires us to design the ΔΣ ADC in our neural front-end to achieve low noise, low area and low power consumption neural front-end in Chapter 8.

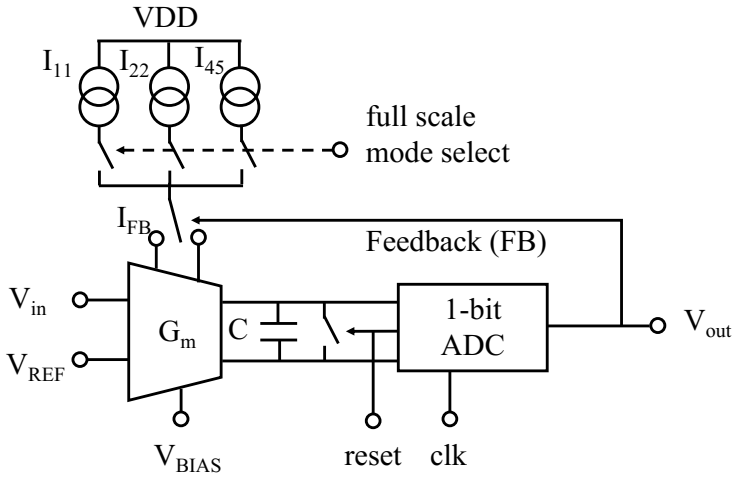


FIGURE 2.12: The architecture of one-bit delta-sigma ADC where utilises three scales of current sources to conduct calibration.

2.4 Main Reference Circuit: Neural Pre-amplifier

Our group has proposed a neural front-end for spike detection which implements adaptive thresholding by using memristive devices [30]. This design is referred to as the initial front-end of this thesis. This section covers the design concept and basic operations of the initial front-end. The author's work on furthermore performance analysis is in Chapter 4, and the following designs are in Chapter 5, 6 and 8.

2.4.1 Design Concept

It is important for implantable devices to increase the output data transfer rate and decrease power consumption. Thus, we keep the necessary functions in implantable devices and exclude other functions off-chip. It has been proved that it is sufficient to detect neural spikes and transfer to off-chip [45] since most of the information representing the extracellular activities is in APs. Therefore, we can include the main processing function of spike detection and digitisation in the system. The duty cycle of APs is around 2% to 20%, which implies large potential power reduction when the system only processes and

TABLE 2.4: Comparison among state-of-art neural front-end and the specification of this work.

Parameter	Dorigo_2018 [68]	Nikas_2019 [78]	Lee_2020 [88]	Wang_2021 [79]	Yoon_2021 [69]	Wendler_2022 [87]	Target
Technology	180nm CMOS	180nm CMOS	110nm CMOS	55nm CMOS	65nm CMOS	180nm CMOS	180nm CMOS + Memristor
Functionality	recording only				recording and detection	recording only	recording and detection
Coupling type	DC	DC	DC	DC	AC	DC	DC
Architecture	CT incremental $\Delta\Sigma$	IA+CTDSM	IA+CTDSM	$\Delta - \Delta\Sigma$	neural amplifier + SAR ADC	CT incremental $\Delta\Sigma$	CTDSM
Area per channel (mm^2)	0.0049*	0.088	0.078	0.0077	0.0062*	0.0046*	<0.005
Power (μW)	46.29	23	6.5	61.2	2.72 (recording module)	8.57	$\leq$ 0.75
DC rejection (mV)	/	12.5	$\pm$50	$\pm$70	/	120	$\geq \pm$50
Bandwidth (Hz)	10k	5k	10k	10k	10k	10k	10k
FSR (mV)	$\pm$45	26	300	148	/	$\pm$7.6	50
IRN (μV_{rms})	20.19	7.3	9.5	5.53$\pm$0.36	8.98	4.46	<10

*the area per channel of recording module only

transmits neural spikes and outputs redundant signal as zero/null [89]. Therefore, the objective of the processing system is to detect neural spikes and output them as digital signals.

However, the processing system is fed by the extracellular raw signal collected directly from the neural interface, which mixes both local field potential and action potential. The clean and stable AP contributes to higher accuracy of threshold detection of neural spikes. Since the LFP fluctuates on a large scale and may submerge the desired AP in voltage amplitude, we need to separate the AP from the raw signal from the aspect of frequency as the first step.

With action potentials fed into the system, it conducts detection after amplification generally so that the signal occupies a large voltage range and is less sensitive to noise. The aim is to output sparse neural spikes; thus, the amplification of discarded data points still occupies considerable energy. Another challenge is that APs are in the minute amplitude range that requires an expensive amplifier operating in low distortion and high sensitivity. With the consideration that action potential is in low frequency and sustains in milli-second, we can integrate the micron-volt signals to milli-volt slowly instead of purely amplifying. As for spike detection, a simple solution is to conduct threshold detection that biopotentials amplitude crosses the specified threshold voltage will be detected as a spike, which requires a reference voltage input. Therefore, we proposed a fully differential amplifier with appropriately sized load capacitors for enabling signal integration, which is an input of a train of APs and a reference voltage at each amplifier input terminal, respectively. Then, the output of the amplifier connects to ADC to realise digitisation, which output can be transmitted off-chip for further processing and/or storage.

Besides, there are other tricky design concepts that access less power consumption and high flexibility. Since input signals of interest are sparse, the system is not necessary to operate on the whole process which can be controlled by fixed external signals or asynchronous signals. In addition, the spikes provoked by the stimuli last for $1ms$ typically [90]. Thus, this architecture is controlled by clocking signals, switching between 'operate' and 'off' modes, which is a solution to energy reduction. With regard to offset, we induced memristive devices (whose resistive state can be programmed) in two current branches to tune the offset instead of designing an expensive offset compensation circuit, making the circuit more flexible.

2.4.2 Operational Principles

The schematic diagram is presented in Figure 2.13, which is a fully differential amplifier wrapped by the dynamic latch comparator, while the biasing circuit unit can be shared with multiple channels. The whole system mainly conducts the threshold detection for APs and outputs digital signals where ones represent spikes and zero as non-spikes. The

following part will access to four operation phases in detail and explain the main components of the circuit.

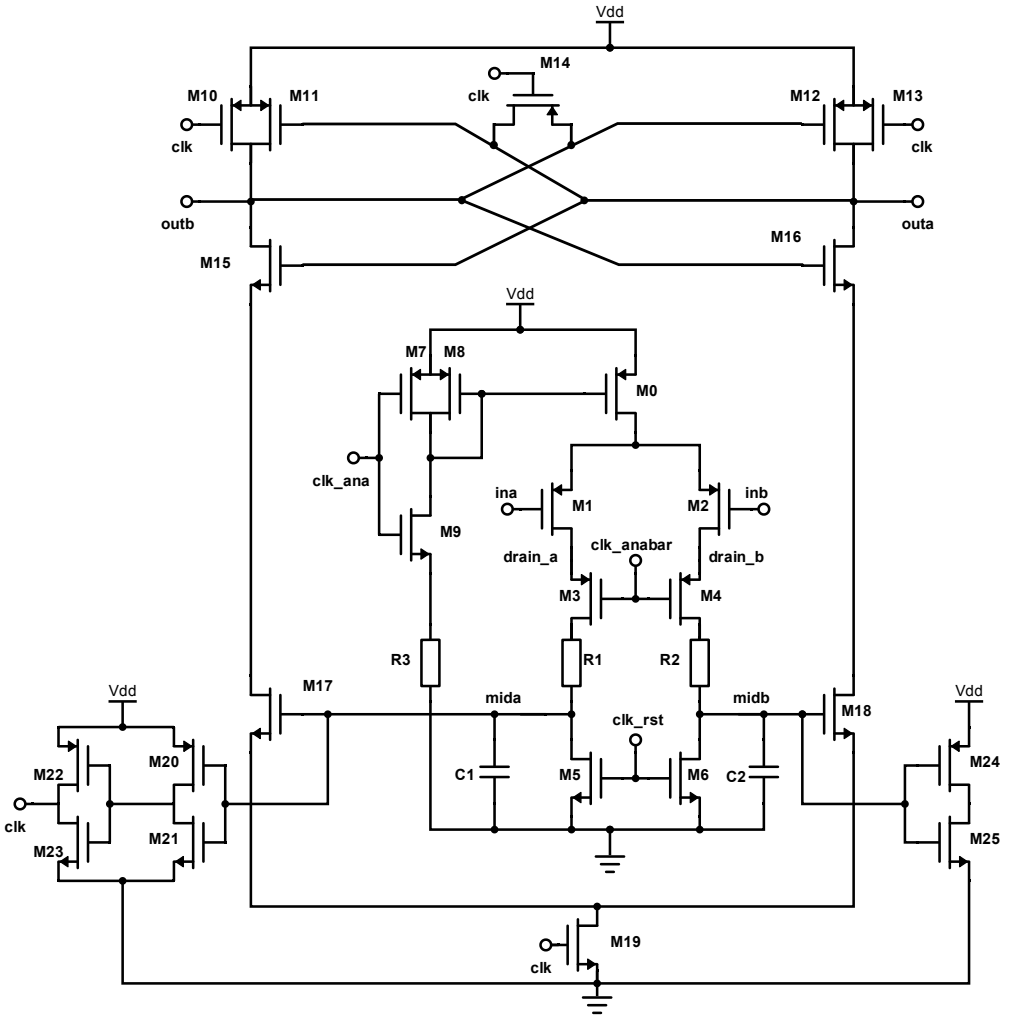


FIGURE 2.13: The architecture of pre-amplifier. It consists of three parts: (i) the core integrating amplifier, (ii) the dynamic latch comparator and (iii) the current bias control unit. Control signals, including *clk*, *clk_ana*, *clk_anabar*, *clk_rst* are all assumed to be generated by voltage sources which are strictly periodic.

A completed detection cycle consists of four phases, which have been shown in Figure 2.14: (i) reset, (ii)integration, (iii) digitisation and (iv) off phase. The detailed operation of each phase will be explained as follows:

In the reset phase (i) the core amplifier is on (*clk_ana*, *clk_rst*: high, *clk*, *clk_anabar*: low) and the load capacitors are discharged ($V_{mida/b} = 0$), so that voltage/current in core amplifier is initialised and cleared before integration commences in the next phase.

In the integrating phase (ii) (*clk_ana*: high, *clk_anabar*, *clk_rst*, *clk*: low) the reset transistors (M5&M6) are switched off and the currents flowing through the branches of the core amplifier drain into the load capacitors. From a 'large signal' perspective, V_{mida} and

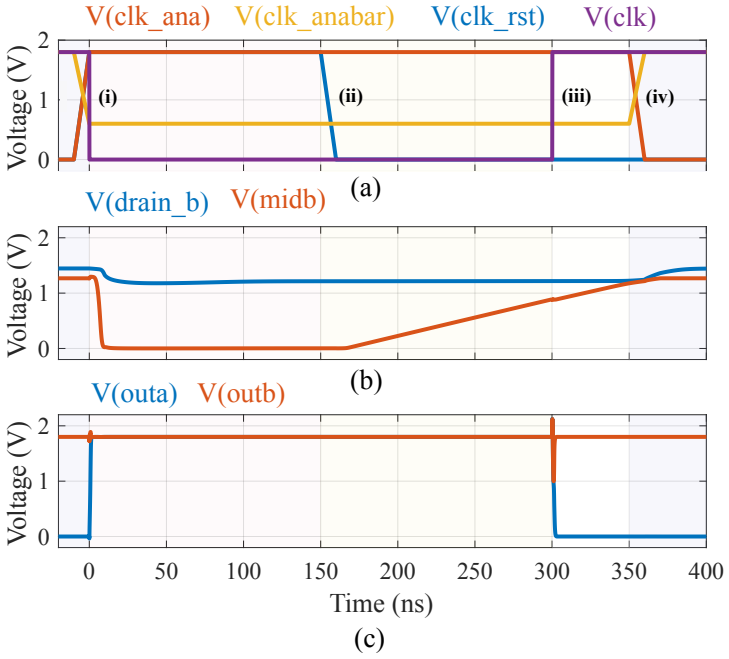


FIGURE 2.14: Transient simulation of one detection cycle of the neural spike. One detection with four phases completes in $350ns$ as an example, which has been labelled and highlighted with different backgrounds. It contains (i) reset, (ii)integration, (iii) digitisation and (iv) off phases. (a) presents the control of clocks; (b) compares the amplifier output ($midb$) with the drain of the input transistor ($drain\_b$); (c) presents the digital output from DLC. Input signals are set as: $ina = 1V + 50\mu V$ and $inb = 1V$.

V_{midb} continuously increase during integration. In terms of 'small signal', $\Delta V_{mida-midb}$ increases with time and normal operation is maintained so long as the cascode transistors M3&M4 remain in the saturation region. The voltage difference between nodes $mida$ and $midb$ is impacted by the charging speed/current and integration time. Memristors R1&R2 work as trimming devices and tune the offset of the core with very high sensitivity ($1\mu V/k\Omega$ shown in [30]). At the end of this phase, $V_{mida/b}$ should be high enough to successfully trigger the DLC and $\Delta V_{mida-midb}$ should be as large as possible for maximising gain.

In the digitisation phase (iii) (clk_ana, clk: high, clk_anabar, clk_rst: low) clk goes high, triggering the DLC to perform the conversion of $V_{mida/b}$ into the final digital outputs. By convention, we take the output from the branch where output '1' represents a spike while '0' represents the absence of a spike. Shortly after the decision is committed by the DLC, the core amplifier is turned off as the system re-enters the off phase.

Finally, in the off phase (iv) (clk: high, clk_ana, clk_anabar, clk_rst: low), the tail current is cut off by setting clk_ana to zero. The pre-amplifier is turned off and stops recording neural signals. clk_anabar is also deactivated (goes to high), thus preventing the accumulated charge across the large gate capacitances of M1&M3 from draining away.

After access to the basic operation of one detection cycle, the test with the differential input range of $\pm 500\mu V$ is presented in Figure 2.15. Input A (ina) is slowly swept between

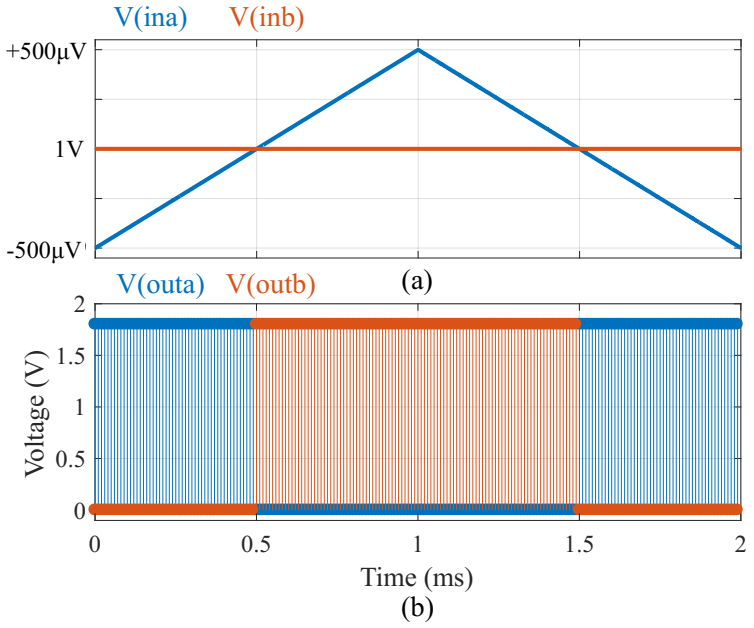


FIGURE 2.15: Pre-amplifier basic functionality test: Input A (ina) is slowly swept between $[1V - 100\mu V,\ 1V + 100\mu V]$ over $2ms$ while the pre-amplifier is carrying out a conversion every $10\mu s$ to detect the relationship between inputs A and B. Input B (inb) remains stable at $1\,V$ throughout. In this test, the amplifier was balanced ($R1 = R2$). When $V_{ina} < V_{inb}$, the left branch current is larger than the right branch current, inducing $V_{mida} - V_{midb} > 0$. The DLC captures this relation and generates binary signals: $V_{outa} = 1$ and $V_{outb} = 0$, which appears in the bottom panel as a predominantly orange output trace. Conversely, when $V_{ina} > V_{inb}$, $V_{outa} = 0$ and $V_{outb} = 1$, which appears as a combined orange/blue output trace. Note: this type of simulation can also be used to test the offset tuning range and tuning sensitivity on the resistive state of memristive devices. When $R1 > R2$, V_{ina} must be lower than V_{inb} to ensure a balanced output, creating an offset, this is read in the output trace as an encroachment of the blue region into the orange (and vice versa for $R1 < R2$).

$[1V - 500\mu V,\ 1V + 500\mu V]$ over $2ms$ while the pre-amplifier is carrying out a conversion every $10\mu s$ to detect the relationship between inputs A and B. Input B (inb) remains stable at $1\,V$ throughout. In this test, the amplifier was balanced ($R1 = R2$). When $V_{ina} < V_{inb}$, the left branch current is larger than the right branch current, inducing $V_{mida} - V_{midb} > 0$. The DLC captures this relation and generates binary signals: $V_{outa} = 1$ and $V_{outb} = 0$, which appears in the bottom trace as a predominantly orange output trace. Conversely, when $V_{ina} > V_{inb}$, $V_{outa} = 0$ and $V_{outb} = 1$, which appears as a combined orange/blue output trace. This type of simulation can also be used to test the offset tuning range and tuning sensitivity on the resistive state of memristive devices. When $R1 > R2$, V_{ina} must be lower than V_{inb} to ensure a balanced output, creating an offset. This is read in the output trace as an encroachment of the blue region into the orange (and vice versa for $R1 < R2$).

2.5 Summary

The development of the electrode techniques enables researchers to acquire a larger amount of neural data. Therefore, the primary responsibility of the front-end designer is to process the weak signals from the electrodes while adhering to low-noise requirements and power and area constraints. It necessitates the selection of appropriate DC offset rejection and signal processing techniques. In this instance, we use a memristive device in the CMOS technique to compensate for the DC offset while simultaneously integrating spike detection on-chip to realise a multi-functional front-end.

Chapter 3

Integrating Memristive Technology into CMOS Design

Memristor is an emerging two-terminal device whose resistance state (RS) can be changed by the applied bias voltage. It is a programmable and non-volatile device at the nanometer scale that is compatible with CMOS [31; 32]. Considering that the memristance is tuneable continuously [27], it can be regarded as an analogue device that processes the continuous signal. With these characteristics, it can be utilised as a trimming device with the feature of programmable resistance [30]. In this project, CMOS integrated circuits with memristor implants operate more flexibly because the RS can be altered by applying a bias voltage, operating in a 'write once-read many' regime.

Previously, we only had the Python model for this device in the range of [$20k\Omega$, $120k\Omega$], which is used on our instrument measurement interface. The first step in incorporating the memristive model into circuit design is to make the memristive Verilog-A model accessible to EDA tools. It can be utilised to represent the behaviour of memristive devices through physical equations, laying the foundations to enable the inclusion of these devices into integrated circuits. This chapter focuses on the specific Verilog-A memristor model (in the range of $20k\Omega$ to $120k\Omega$) that uses quadratic fitting as proposed in [36]. In this project, I converted the mathematical model into a Verilog-A model and made it available to Cadence, whose methodology refers to [37]. In this chapter, we cover the procedure and significant parameters of the Verilog-A model. Additionally, the integration of the memristor model into the Cadence Virtuoso design environment is documented, along with the methodology for integrating memristors with CMOS designs.

3.1 Verilog-A Memristor Model

The Verilog-A model provided in [37] represents the behaviour of memristive devices through physical equations. There are two mathematical models for our memristive device: exponential [37] and quadratic [36]. But we only had the exponential Verilog-A model, which can reflect the physical device with higher accuracy. However, the RS is limited to a narrow range of $[10k\Omega, 17k\Omega]$. The author applied the quadratic model to CMOS design to access the wider RS range of $20k\Omega$ to $120k\Omega$.

In this chapter, we convert the quadratic mathematical model in [36] into the Verilog-A model (in Listing 1), whose methodology refers to [37]. The explanation and implementation of the Verilog-A model in Cadence Virtuoso are given as follows.

- As a two-terminal device, p, n is defined as the 'inout' ports of the memristor, where the applied bias voltage (vin) can be used for calculation and the current through the device will be output (lines 1-3, 29, 45). The RS can be accessed through the nonlinear $I-V$ characteristic in lines 42-44.
- The fitting parameters, switching parameter (eta) and initial RS (Rinit) are defined from lines 4 to 17. The fitting parameters are extracted from the experimental results of an in-house fabricated device with a specific stimulus, while the details can be found in [37]. The switching direction parameter (eta) is 1 in this model, indicating the direction that the positive voltage induces higher RS. User-defined initial values for RS should be within the proper range determined in line 30.

Listing 1: the Verilog-A memristor model representing the in-house fabricated Pt/TiOx/Pt device using quadratic fitting

```
1.  module analytical (p, n);
2.      inout p, n;
3.  electrical p, n;
4.      parameter real Ap = 0.12340;
5.      parameter real tp = 2.74111;
6.      parameter real An = -0.33000;
7.      parameter real tn = 2.59685;
8.      parameter real rp0 = -40928.13784;
9.      parameter real rp1 = 55117.97865;
10.     parameter real rn0 = 41366.35820;
11.     parameter real rn1 = 7789.66771;
12.     parameter real Rinit = 40000;
13.     parameter real eta = 1;
14.     parameter real ap=0.225;
15.     parameter real bp=4.12;
16.     parameter real an=0.2801;
17.     parameter real bn=4.10;
18. real Rmp, Rmn, svp, svn, vin, RS, IVp, IVn, IV;
19. real first_iteration, R0_last, dt, it;
20. analog function integer stp;
21.     real arg; input arg;
22.     stp = (arg >= 0 ? 1 : 0 );
23. endfunction
24. analog begin
25.     if (first_iteration==0) begin
26.     it=0;     R0_last=Rinit;
27.     end
28.     dt=$abstime-it;
29.     vin=V(p,n);
30.     Rmp=rp0+rp1*vin;          Rmn=rn0+rn1*vin;
31.     if (vin>0) begin
32.     svp=Ap*(-1+exp(abs(vin)/tp));
33.     RS=(R0_last+svp*Rmp*(Rmp-
        R0_last)*dt)/(1+svp*(Rmp-R0_last)*dt);
34.     end
35.     else begin
36.     svn=An*(-1+exp(abs(vin)/tn));
37.     RS=(R0_last+svn*Rmn*(Rmn-
        R0_last)*dt)/(1+svn*(Rmn-R0_last)*dt);
38.     end
39.     if (RS>=Rmp && vin>0)          RS=R0_last;
40.     if (RS<=Rmn && vin<0)          RS=R0_last;
41.     if (abs(vin)<=0.5)             RS=R0_last;
42.     IVp=ap*(1/RS)*sinh(bp*vin);
43.     IVn=an*(1/RS)*sinh(bn*vin);
44.     IV=IVp*stp(vin)+IVn*stp(-vin);
45.     I(p, n)<+ IV;
46.     R0_last=RS;
47.     first_iteration=1;
48.     it=$abstime;
49. end
50. endmodule
```

- Lines 20-23 define a 'step function', which is then employed in line 44 to determine

the output current branch. For instance, when *vin* is positive, the result of 'stp(vin)' is one while 'stp(-vin)' is zero; thus, the current branch 'IVp' will be output.

- Analog block (lines 24-49) indicates two features of the device operation: iteration and transient. Lines 25 to 27 present the setup of the first iteration. In this case, the start time is set to zero and the latest RS is taken from the defined parameter Rinit. The time step can be obtained from lines 28 and 48 by subtracting the absolute time from the reference time. After deriving RS (lines 31-41), the current through the device will be updated (lines 42-44) and passed to ports (line 45). Then, the latest RS and absolute time will be updated for the next iteration (lines 46-48).

- The boundaries of RS can be derived from line 30. Within the RS boundaries, RS can be calculated under the condition of constant bias voltage in lines 31-38. However, the switching fails to realise if the latest RS exceeds the boundaries or the bias voltage drops below $0.5V$ (lines 39-41).

In conclusion, Verilog-A models can only be used to simulate transients and can be used to 1) conduct static current-voltage measurements within specific boundaries; 2) gather transient switching characterisation by applying bias voltages to initial RS. In order to process the above utilisation, we need to set up two stages: 1) read RS by applying triangular pulse with $0.5V$ amplitude, which prevents the device from switching; 2) write, or we said, change the RS, by applying pulses with defined duration/width, polarity, amplitude, and numbers of pulses. For extending the application, fitting parameters can be changed to represent other memristive devices. At this stage, the model does not incorporate AC analysis/small-signal modelling, noise performance, parasitics and device variation.

3.2 Integration of Memristor Model into Cadence

3.2.1 Import Verilog-A Model into Cadence

In order to utilise the device in circuit design as well as simulation, cooperating with electronic devices, we develop a symbol that links to the Verilog-A model so that it can be incorporated into circuits as an element. The instruction is shown as follows:

- Build the library and refer to chosen technologies. In order to cooperate memristor with CMOS, users are supposed to be familiar with the behaviour of the memristor quantitatively, such as the range of high/low resistive state, the allowed range of applied voltage/current, static $I - V$ characteristics and etc., which allows them to choose a more suitable technology, as well as define circuit specification with more realisable. In this case, we integrate the Verilog-A model in Cadence and choose

0.18um technology with four fabrication metal masks for demonstration, where the proposed memristor was fabricated with Metal4 as ports.

- In library Manager, we create a cell view in a specific library with Verilog-A type, named 'memristor' in library 'DesignMethodology' (Figure 3.1(a)). Then, we paste the provided Verilog-A code into the text editor and save the file.
- From the toolbar in the text editor, we can access create 'cellview from cellview' to create a symbol from Verilog-A memristor model (Figure 3.1(c).
- We assigned the position of ports on the left and right sides which can be automatically detected from the Verilog-A code, followed by the symbol design (Figure 3.1(b)).
- The memristor model can be applied to the schematic, operating with transistors, resistors, capacitors and etc. Besides, the model can be utilised flexibly, where fitting parameters (shown in Figure 3.1(d)) can be modified to represent other memristor models. In this case, we only have access to construct a single proposed memristor with voltage sources to investigate the switching characterisation.

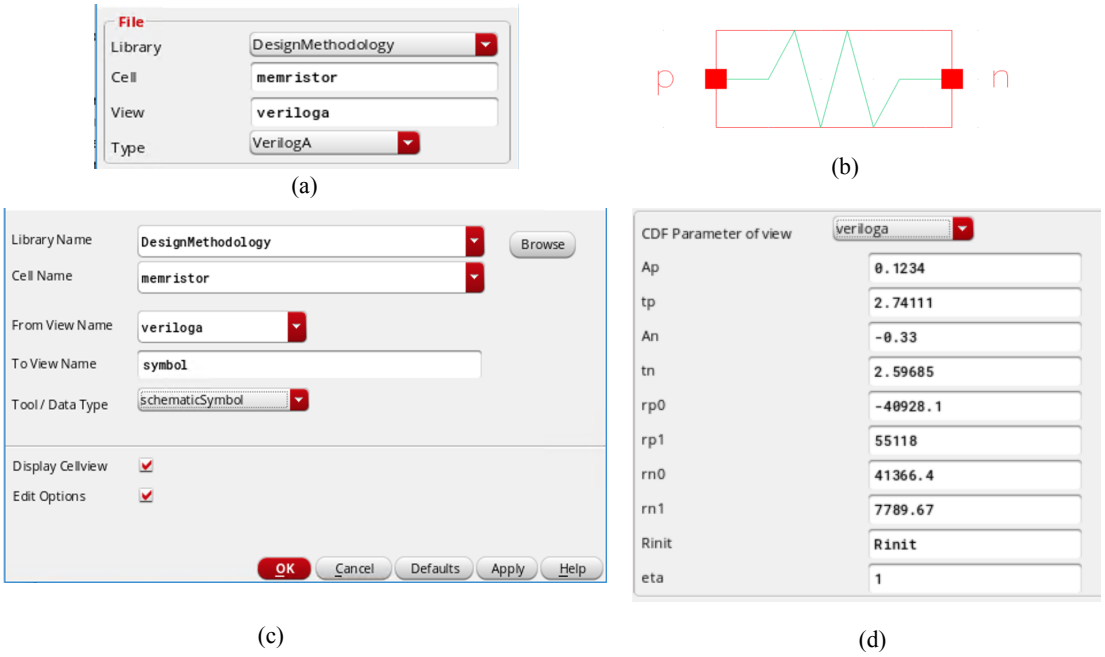
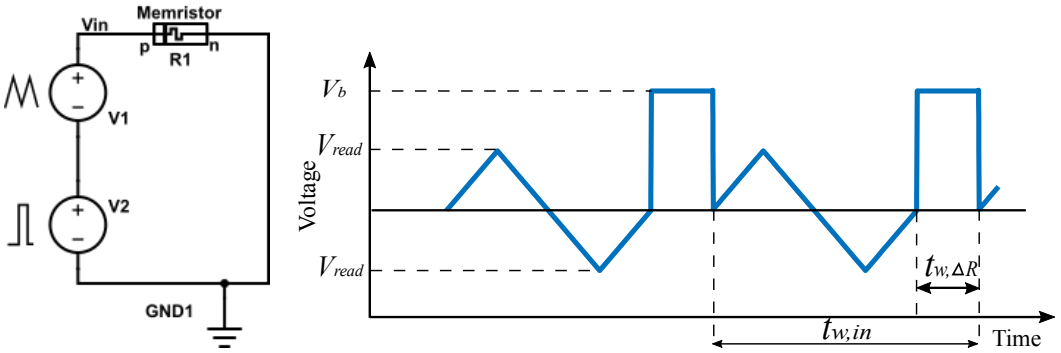


FIGURE 3.1: The operation sequence of importing memristor into Cadence. (a) Process of building a new cell view in Verilog-A type for memristor. (b) Symbol of the memristor. (c) Process of creating a symbol for memristor from Verilog-A cell view. (d) The parameters of the memristor can be modified in the object properties window.

3.2.2 Simulation Setup for Memristor-based Circuit

Considering that the Verilog-A memristor model calculates RS against time and keeps tracking the change of RS, the model is supposed to run in transient simulation to process

both write and read operations. In this case, we took a single memristor with voltage sources as a simulation example that was provoked by a chain of pulses followed by a triangular wave to conduct the write and read of the memristor, respectively. The operation steps will be shown as follows:



(a) (b)

FIGURE 3.2: (a) Testbench of applying pulse and triangular wave to memristor to conduct the write and read operations in the transient simulation. (b) Pulse chain that provokes the memristor. The input signal sequence can be divided into two parts: triangular wave and pulse. For all the simulations of our device, the read voltage is defined as $V_{read} = 0.5V$ with $1ms$ duration. The pulses can be determined with specific duration/width ($t_{w,\Delta R}$), amplitude (V_b) and the number of pulses to provoke the memristor. With the combination of two stages, the RS of the memristor model can be tracked for each stimulus.

- In this demonstration, we built the schematic with memristor, piece-wise linear (PWL) and pulse voltage sources (Figure 3.2(a)). In the schematic, we have defined the direction that voltage sources are connected to 'positive' (p) port of the memristor. The positive bias voltage ($V_b > 0$) from 'p' provokes the memristor to higher RS, while the RS will decrease when positive voltage applies at 'n' port. In testbench (Figure 3.2(a)), the proposed model is in OFF transitions under positive voltage, while negative voltage leads to ON transitions. Two voltage sources generate triangular waves and pulse alternately to process that the pulse changes RS of memristor followed by a triangular wave which keeps tracking RS. The detail of the input signal has been shown in Figure 3.2(b). In this measurement for the proposed model, the identical triangular wave is defined as $0.5V$ peak amplitude (V_{read}) within $1ms$ duration. The pulse mainly has three variables: width ($t_{w,\Delta R}$), amplitude (V_b) and the number of pulses.

- After setting up the testbench and running the transient simulation, we are supposed to record the change of RS. In this case, we need to record every current across the memristor at $V_{read} = 0.5V$. With both read voltage and current, the RS can be obtained by dividing $V_{read} = 0.5V$ by read current. For detail operation, users can 1) process the division in Cadence calculator and plot RS; 2) send the RS to the table and save it as a '.csv' file; 3) process the data in Matlab in order to extract and plot the RS at $V_{read} = 0.5V$.

The simulation example of the Verilog-A model in Cadence is given in Figure 3.3. This type of simulation allows us to elaborate on the effect of the resistance state from the applied pulse width, amplitude and polarity.

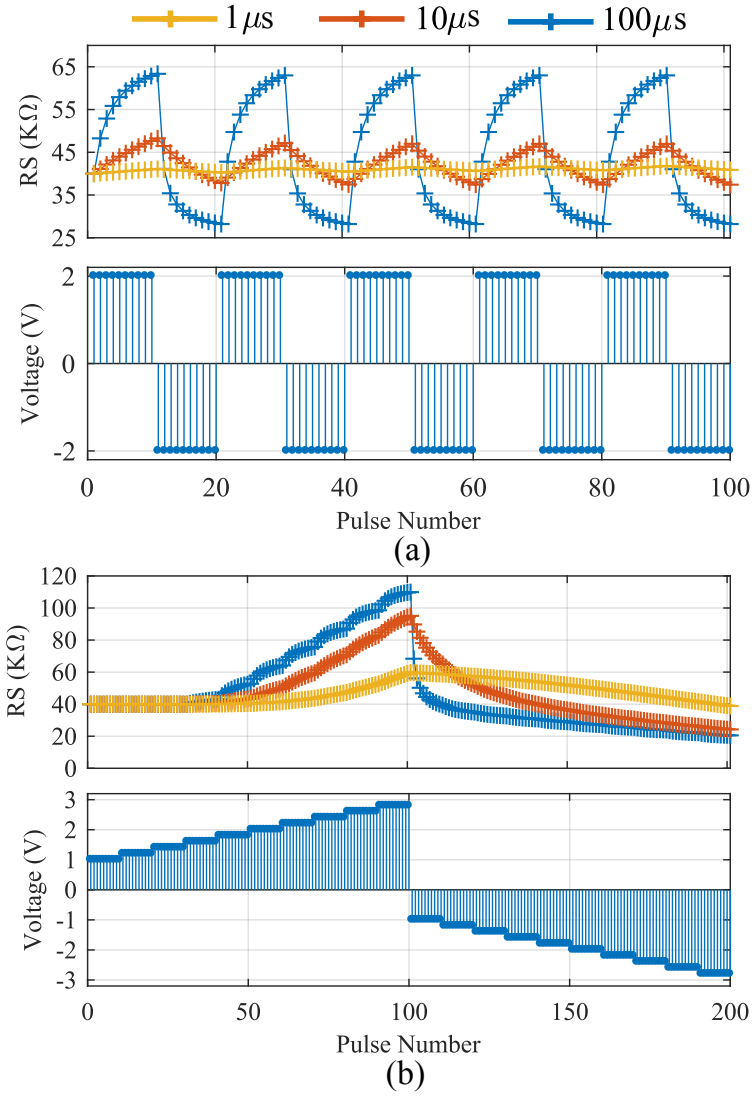


FIGURE 3.3: Simulation results for the quadratic memristive Verilog-A model in Cadence. (a) shows the responses of the model to $1, 10, 100\mu s$ pulse widths with constant amplitudes ($\pm 2V$). (b) presents the model response to the voltage ramp from 1V to 3V.

3.3 Summary

This chapter shows the behaviours of memristor from explaining the Verilog-A model in terms of non-linear $I - V$ relationship and RS switching due to bias voltage. Besides, the methodology of utilising the memristive device into Cadence for circuit design and simulation is given. It can be regarded as the foundation for the below chapters that apply the memristor to the amplifier, including defining and measuring the performance of the proposed amplifier.

Chapter 4

A Memristor-based Pre-amplifier Topology

Based on the design requirements and application provided in Chapter 2, this chapter analyses the pre-amplifier at the transistor level and provides theoretical calculation of the core memristive OTA. Since the proposed OTA operates in a 'start-stop' scheme, its performance metrics are slightly different from the conventional OTA. And the differences and explanations are given below. In this thesis, we modified the initial front-end (Figure 2.13) and the proposed schematic is shown in Figure 4.1. In the initial front-end [30], the clocking signal (clk) is generated by the output of core OTA (mida). On the other hand, the clk in the proposed OTA is set to be generated by the external clocking module, which guarantees the DLC capture the sufficient voltage difference from the core OTA. The next chapter presents a transient simulation for signal detection with more details.

4.1 Transistor Level Analysis

After introducing basic functions of the pre-amplifier in Chapter 2, we analyse the behaviour of significant transistors of memristive OTA for afterwards design. The dynamic OTA operates in four phases: i) initial, ii) integration, iii) digitisation and iv) off.

The input differential pair (M4&M5) provides the main transconductance for the core amplifier which converts the effect of gate voltage into current in CMOS device. To achieve maximum gain, the g_m needs to achieve the highest value (for a given bias current), which indicates the input transistors need to bias to the subthreshold region. In addition, a maximum g_m makes a contribution to reducing thermal noise component of the input-referred noise and the use of large input transistors achieves a lower flicker noise component.

The cascode transistors (M6&M7) in the core amplifier mainly contribute to increasing output resistance and being shield devices to isolate the input pair from memristive devices

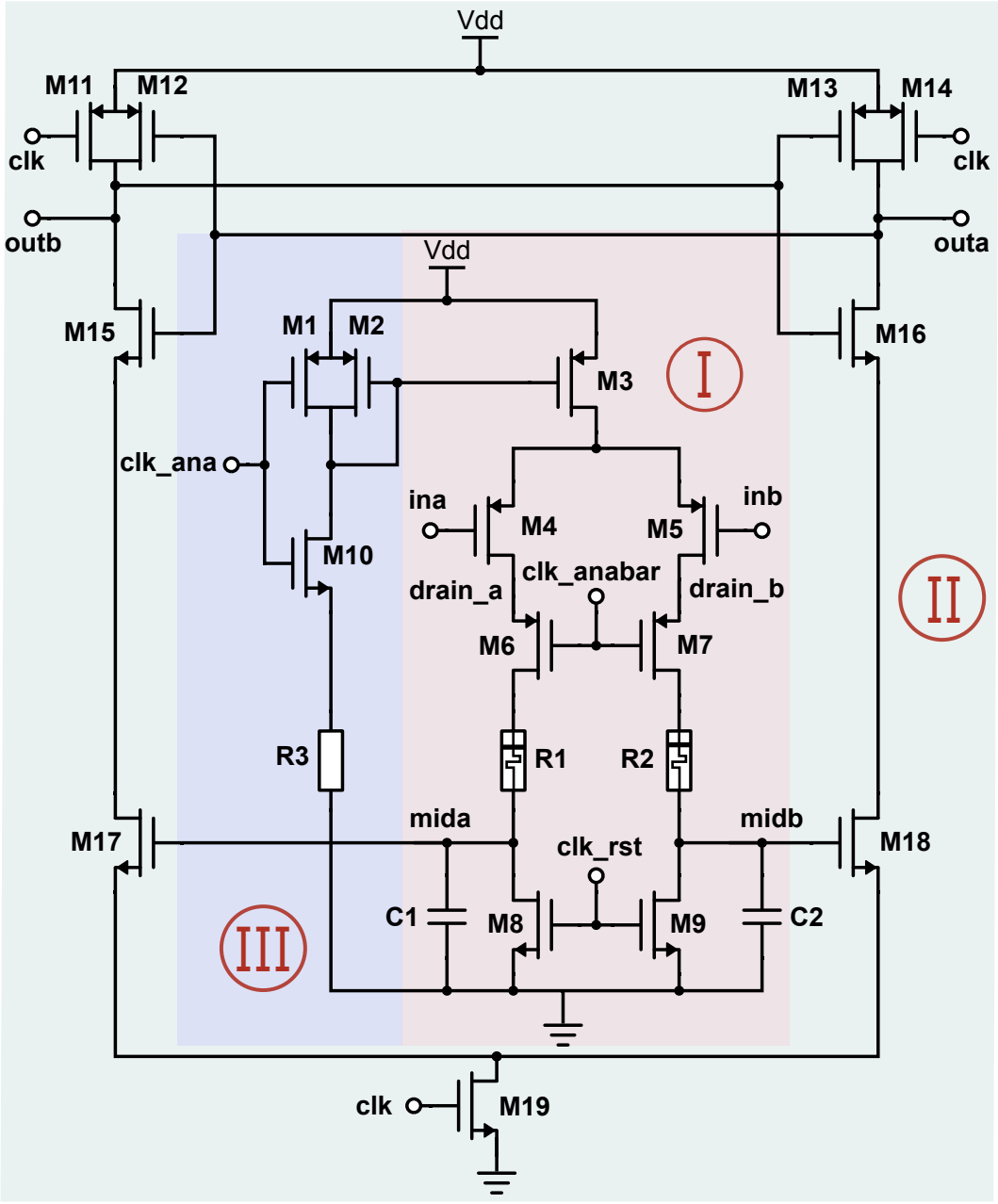


FIGURE 4.1: The schematic of the proposed pre-amplifier. Compared to the initial design in Figure 2.13, this circuit uses external *clk* to control that the DLC can be fed by sufficient voltage difference and increase detection accuracy.

and dynamic voltage changing due to integration. Thus, drains of input pair ($drain\_a/b$) can be maintained at a constant voltage level, which avoids fluctuating input pair.

The memristor, along the current branch, plays as a trimming device in the core amplifier. Generally, the circuit requires an adjustment circuit to cancel or compensate the inherent offset voltage to prevent drift after fabrication. But in this case, we do not intend to cancel the inherent offset; on the other hand, we take advantage of the inherent offset as part of reference/threshold voltage during threshold detection. The differential output $\Delta V_{mid} = V_{midb} - V_{mida}$ is impacted by the inclined current with the same integration time.

We applied memristors (R1&R2) that can be programmed to trim the effective resistance which inclines the balance of two current branches. Thus, it requires that the tuneable range should cover the range of $\Delta V_{spike,max} \pm |V_{os,max}|$, where $V_{spike,max}$ is the maximum potential of neural spike and $V_{os,max}$ is the maximum offset voltage of the core amplifier.

As for the bottom transistors (M8&M9), they operate as switches rather than active loads. During the 'initial phase,' the bottom transistors are activated to discharge the load capacitors (C1&C2) and initialise the status of the input and cascaded pairs. And they are turned off during the 'integration phase' to guarantee that charges flow into the capacitors to realise integration and amplification. As a result, the design of the bottom transistors should prioritise minimising parasitics and leakage current.

In this section, we highlight some significant components and preliminary design considerations. The detailed theoretical analysis and calculation of specified indicators are given below.

4.2 Theoretical Analysis of Performance Metric

Considering that the core amplifier is a dynamic amplifier where the integrated output signals ($mida/b$) keep changing the DC operation point, the performance of the amplifier needs to be re-defined based on the transient analysis since the traditional DC/AC analysis based on the static operation point is unsuitable for this case. In this section, we assess the foundation of the performance metric and tailor it to the integrating amplifier. Most of the indicators mainly depend on the 'integrating phase' (iii in Figure 2.14) when the amplification is processed. Thus, the analysis will be focused on the core amplifier in the integrating phase.

Furthermore, the bias current (I_{tail}) distributes into differential branches with a slight incline (Δi) depending on weak input signals. This allows us to conduct both 'large' and 'small' signal analysis by using transient simulations to obtain the relevant data. Thus, we define the V_{mida}, V_{midb} and I_{tail} as 'large' signals in this case, while the voltage difference $\Delta V_{mid} = V_{midb} - V_{mida}$ and inclined current Δi belong to 'small' signal. The following access to each performance indicator one by one with the insight of definition foundation and the derivation of tailored definition. This chapter mainly focuses on the performance analysis of core OTA, whose parameters are listed in Table 4.1.

4.2.1 Overview of Performance Metric

The indicators are divided into different types in Table 4.1. Moreover, we briefly describe significant indicators and provide a detailed explanation for the specific ones in the following sections.

TABLE 4.1: Specification of the memristive integrating OTA.

Type	Parameter	Target
DC	Bias current Input common mode range V_{ICR}	$\geq$ 50mV
	Input differential voltage range V_{IDR}	$\geq$ 5mV
	Offset compensation by memristors V_{COMP}	$\geq V_{OS}$+500μV
AC	Open loop voltage gain A_{OL}	$\geq$ 26dB
	Bandwidth BW	10kHz
	Power supply rejection ratio PSRR	$\geq$ 100dB
	Common mode rejection ratio CMRR	$\geq$ 100dB
Transient	Sampling rate F_s	$\geq$ 20kHz
Noise	Input-referred noise e_n	$\leq 10\mu V_{rms}$
Minimum ratings	Power dissipation	$\leq 750nW$

DC There are two types of input ranges: i) common mode range and ii) differential voltage range. The input common mode range V_{ICR} is defined as the voltage that keeps the input transistors in the subthreshold region, which analysis is in section 4.2.2. The input differential voltage range V_{IDR} is relative to the linearity where the output is linear to the input voltage. Due to the electrode can induce up to 50mV DC offset to the amplifier, to guarantee the input transistors are in subthreshold region, V_{ICR} is required to be larger than 50mV. As for the differential input, it contains both LFP and AP that we need to set $V_{ICR} \geq V_{LFP,max} + V_{spike,max} \approx 5mV$.

The other highlighted parameters are relevant to offset. The inherent offset voltage is dependent on the sizing and layout design. And we determine that the compensated voltage by memristor V_{COMP} needs to cover both the spikes and inherent offset of the circuit ($V_{spike,max} \pm |V_{os,max}|$), where the amplitude of neural spike is $V_{spike,max} = 500\mu V$.

AC Due to the amplification being realised by integrating the capacitor, the open loop voltage gain A_{OL}/output voltage Δmid should be relative to the accumulated charges in the capacitors. Furthermore, the detailed calculation is in section 4.2.3. We required the OTA to amplify the minimum spike $50\mu V$ to millivolt-level, which requires 20V/V (26dB) amplification. As for the frequency response, the bandwidth is set in the range that covers the neural signals which are up to 10kHz. Due to the front-end being required to reject the DC offset, the OTA should operate as the bandpass filter by applying it into the open/closed loop network as an AC-coupled front-end. In addition, the OTA can be applied to the DC-coupled solution described in Chapter 6.

Transient The maximum frequency of the neural spike is 10kHz. According to the Nyquist theorem, the sampling frequency must be at least twice the signal of interest; we set the sampling frequency is required to be higher than 20kHz.

Noise Noise from the input amplifier is typically lower than that from electrodes [91]. Moreover, the noise introduced by the electrodes is typically $10\mu V_{rms}$ [13]; thus, we target

the input-referred noise e_n to be lower than $10\mu V_{rms}$.

Power consumption As estimated in Chapter 1, the maximum power consumption of the implanted integrated circuit is restricted to 15mW. Moreover, we estimated that the transmission module occupies 50% of the power budget; the 10,000-channel front-end can only consume 750nW per channel.

4.2.2 Input Range

As the analysis above, the input transistors M4&M5 are biased in the subthreshold region to achieve higher voltage gain and lower input-referred noise. It implies two operating conditions: (1) A minimum drain-source voltage $|V_{ds,min}| = 3 \cdot V_T$, where V_T is the thermal voltage and good rule of thumb for ensuring in subthreshold region [92]. (2) the gate-source voltage $|V_{gs,4}|(< |V_{th,4}|)$ that allows the transistor to pass $\approx I_{tail/2}$ in subthreshold region. These two conditions can be unfolded with the aid of Figure 4.2 as follows.

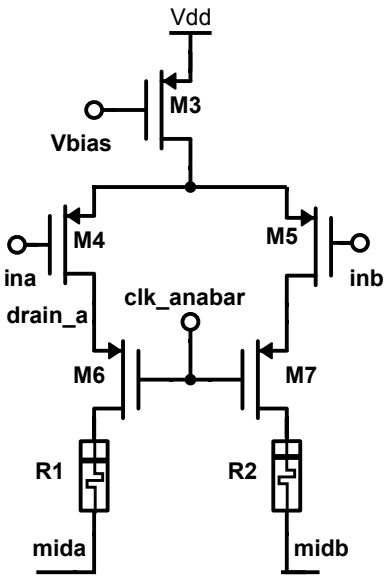


FIGURE 4.2: Part of schematic of core amplifier aiding the input range analysis.

Looking into Figure 4.2 from VDD to the input, the common mode input voltage V_{CM} reaches the top bounded:

$$V_{DD} - |V_{ds,sat,3}| - |V_{gs,4}| \geq V_{CM} \tag{4.1}$$

where $V_{ds,sat,x}$ is the drain-source saturation voltage of transistor x. Exceeding the boundary causes M3 to triode and simultaneously encroach on $V_{gs,4}$, progressively shutting the amplifier down.

The bottom boundary hinges on maintaining the input differential pair in the subthreshold region ($|V_{ds,4}| \geq |V_{ds,min,4}| = 3 \cdot V_T$):

$$|V_{ds,4}| \approx (V_{CM} + |V_{gs,4}|) - (V_{clk\_anabar} + |V_{gs,6}|) \geq 3 \cdot V_T \tag{4.2}$$

where $V_{gs,6}$ is the gate-source voltage allowing the cascode transistor to pass $\approx I_{tail/2}$. This is also treated as approximately constant in this analysis. Under normal operation, the second term is recognised as V_{drain_a}. This unfolds to:

$$V_{CM} \geq V_{clk\_anabar} + |V_{gs,6}| - |V_{gs,4}| + 3 \cdot V_T \tag{4.3}$$

Here, the cascode transistor M6 enforces a specific and relatively fixed value of V_{drain_a} under the control of V_{clk_anabar} (similarly for M7 and V_{drain_b}). Combining Equations 4.1 and 4.3, we can find the approximate value of V_{clk_anabar} above which the input differential pair runs out of the common mode range:

$$V_{anabar,low} = V_{DD} - |V_{ds,sat,3}| - 3 \cdot V_T - |V_{gs,6}| \tag{4.4}$$

We can see the trade-off between common mode and integration voltage ranges (directly connected to gain), depending on the *clk_anabar*. If V_{clk_anabar} increases, push the bottom boundary of V_{CM} to narrow the common input range; meanwhile, it allows large headroom for the signal integration at nodes *miba/b* which boosts the voltage gain.

4.2.3 Open Loop Voltage Gain

For a fully differential amplifier, the voltage gain is defined as the ratio of the output amplitude difference over the input amplitude difference: $\Delta V_{out}/\Delta V_{in}$. As for the proposed amplifier, the voltage gain will be translated into $\Delta V_{mid}/\Delta V_{in}$, where ΔV_{mid} is captured at the end of integrating phase and ΔV_{in} is the input difference between APs and reference voltage in practice. The APs are in low frequency that each spike features of the order of $100s$ micron seconds, while a completed detection cycle lasts within hundreds of nanosecond levels, which implies the input voltage can be regarded constant for analysis.

First, we go through the signal conversion and flow within the core amplifier during the integrating phase step by step. In this structure, the input pair (M4&M5) contributes the main transconductance to convert the voltage into current through the device. The cascode pair (M6&M7) mainly contributes to output resistance and shields the voltage fluctuation from nodes *mida* and *midb*, maintaining the stability of input transistors. Then, the branch current charges the load capacitors that induce both output voltage V_{mid} and the voltage difference ΔV_{mid} increases gradually, where converts the current back to voltage and realises amplification. The process of deriving the formula will be given as follows.

The input signal and reference voltage are input into the amplifier through M4&M5 where the voltage difference ΔV_{in} is converted into current difference Δi. The conversion degree depends on the transconductance g_m of the input pair, which represents the signal amplification ability. The current difference will be:

$$\Delta i = \Delta V_{in} \cdot g_m \tag{4.5}$$

As the branch current drains into load capacitors within the same period, the charges in the capacitor will be accumulated and the charge difference ΔQ will be increased during this process. The difference of charge on the load capacitors:

$$\Delta Q = \Delta i \cdot \tau \tag{4.6}$$

where τ is the integration phase duration. Finally, this gets transformed into the voltage difference we observe at ΔV_{mid} through the load capacitance C:

$$\Delta V_{mid} = \Delta Q / C \tag{4.7}$$

Combining the above yields the gain (G):

$$G = \frac{\Delta V_{mid}}{\Delta V_{in}} = \frac{g_m \cdot \tau}{C} \tag{4.8}$$

From the view of the 'large' signal, currents filling each load capacitor are approximately constant and equal, and the output V_{mid} is mainly dependent on the half-tail current ($I_{tail/2}$) integration within τ:

$$V_{mid} = \frac{Q}{C} \approx \frac{I_{tail/2} \cdot \tau}{C} \tag{4.9}$$

where Q is the total charge accumulated on each node ($mida/b$) as a result of the tail current. Since the load capacitor has discharged clearly in the reset phase, we induce Q to represent the accumulated charge during integration. Similarly, V_{mid} represents the 'large' voltage signal at node $mida/b$ during integration. We can easily get that the integration time τ is mainly dependent on the 'large' signal integration, and τ can be expressed as:

$$\tau = V_{mid} \cdot C / I_{tail/2} \tag{4.10}$$

Then, we can substitute Equation 4.10 into Equation 4.8 and obtain gain (G) as:

$$G = \frac{g_m \cdot V_{mid}}{I_{tail/2}} \tag{4.11}$$

We can divide the formula into two parts: the transconductor efficiency factor of the input transistor pair ($g_m/I_{tail/2}$) and the 'large' signal output voltage (V_{mid}), both of which depend on the differential gain of integrating amplifier. The maximum $g_m/I_{tail/2}$ can be obtained when the input transistors are biased in the subthreshold region. As for the output voltage V_{mid}, it needs to be integrated large sufficient to boost the gain, while V_{mid} should be lower than the voltage that forces the cascode pair (M6&M7) into triode region.

4.2.4 Noise Performance

The input-referred noise derivation in continuous mode is given in this section. The standard MOSFET input-referred noise model containing both thermal and flicker noise is given by the following expression for spectral density [93]:

$$\overline{V_{in\_std}^2}(f) = 4kT\gamma\frac{1}{g_m} + \frac{K}{C_{ox}WLf} \tag{4.12}$$

where k is Boltzmann's constant, T is the absolute temperature, $\gamma = \frac{2}{3}$ for long-channel transistors and higher for shorter channel devices, K a typically empirically determined factor scaling $1/f$ noise, C_{ox} the gate capacitance, W, L the transistor sizes and f denotes (linear) frequency.

The analysis of each component in the core amplifier on noise contribution during the integrating phase will be conducted. The criterion is whether the transistor contributes to AC signal increment from input to output. M3 is beyond the differential branch and mainly contributes to common-mode noise. The input pair (M4&M5) induces the primary increment on the AC signal; thus, it is an effective noise source at *mid* nodes. The cascode pair (M6&M7) operates as a 'common-gate' that contribute zero increments on AC signal, which can be omitted in noise analysis. Finally, the reset transistors (M8&M9) are off during the integrating phase. Similar to continuous mode amplifiers such as the Harrison [94]: the input differential pair provides substantial gain through its gm, mitigating the input-referred contributions from downstream elements (primarily the cascode transistors and the memristive devices). Therefore, in a fully differential amplifier, the input-referred noise will be doubled and given by:

$$\overline{V_{in}^2}(f) = 2 \cdot (4kT\gamma\frac{1}{g_m} + \frac{K}{C_{ox}WLf}) \tag{4.13}$$

where all the component parameters in terms of g_m, $K/C_{ox}WL$ represent the input transistors M4&M5.

4.2.5 Tuneable Range and Sensitivity

The memristive devices applied in the current branches regulate the charging speed to load capacitors by modulating the effective output resistance of the core amplifier as seen by the capacitive load. The calculation of branch resistance will be divided into two parts: (1) find out the impedance of a drain-degenerated of M6, looking into the source of M6, which schematic has been shown in Figure 4.3; (2) apply the drain degeneration analysis again on the source of M4 (Figure 4.1) to calculate the branch impedance.

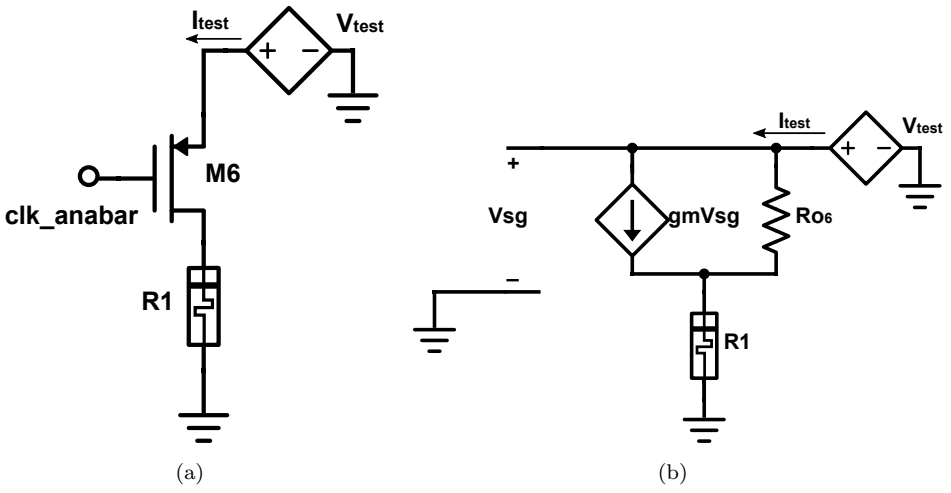


FIGURE 4.3: The circuit diagram of M6 and R1 with a test voltage source presented as (a) a large signal and (b) a small signal. R_o represents the drain-source resistance.

To obtain the impedance, a test voltage source is applied on the source of M6 in Figure 4.3, with the assumption of I_{test}. The effective impedance (Z_{s6}) can be obtained by: V_{test}/I_{test}. Looking into the current path from the voltage source to R0, R1 and GND, we can obtain:

$$V_{test} = I_{test} \cdot R_1 + (I_{test} - g_{m6}V_{sg,6}) \cdot R_{o6} \tag{4.14}$$

To obtain the item V_{test}/I_{test}, Equation 4.14 can be translated into:

$$Z_{s6} = \frac{V_{test}}{I_{test}} = \frac{R_1 + R_{o6}}{1 + g_{m6}R_{o6}} \approx \frac{R_1 + R_{o6}}{g_{m6}R_{o6}} = \frac{1}{g_{m6}}\left(1 + \frac{R_1}{R_{o6}}\right) \tag{4.15}$$

where the approximation is induced due to in saturation operation: $g_{m6}R_o \gg 1$. Z_{s6} is the impedance looking into the source of M6, g_{m6} is the differential transconductance of M6, and R_{o6} the output resistance of M6.

Extending this principle to calculate the impedance of M4, as drain-degenerated by the M6-R1 cascade we obtain:

$$Z_{s4} \approx \frac{1}{g_{m4}} \left(1 + \frac{Z_{s6}}{R_{o4}}\right) \tag{4.16}$$

which eventually unfolds to:

$$Z_{s4} \approx \frac{1}{g_{m4}} + \frac{1}{g_{m6} g_{m4} R_{o4}} + \frac{R_1}{g_{m6} R_{o6} g_{m4} R_{o4}} \tag{4.17}$$

A similar expression also applies for the right current branch.

Setting $A = \frac{1}{g_{m4}} + \frac{1}{g_{m6} g_{m4} R_{o4}}$ and $B = \frac{1}{g_{m6} R_{06} g_{m4} R_{o4}}$ we can express the impedances seen by M3 looking into each current branch as:

$$Z_l \approx A + BR_1 \tag{4.18}$$

$$Z_r \approx A + BR_2 \tag{4.19}$$

where $Z_l = Z_{s4}$ is the left current branch impedance and Z_r is the right branch impedance.

Next, examining the distribution of tail current across the branches we obtain an expression for the left branch current i_l as follows:

$$i_l \approx i_T \frac{A + BR_2}{2A + B(R_1 + R_2)} \tag{4.20}$$

where $i_T = i_3$ is the tail current. Given that $B \ll 1$ (as it is the product of two maximum FET amplifier gains), i_l can be further approximated as follows:

$$i_l \approx \frac{i_T}{2} \left(1 - \frac{B}{2A}(R_1 - R_2)\right) \tag{4.21}$$

Similarly for the right branch current i_r:

$$i_r \approx \frac{i_T}{2} \left(1 + \frac{B}{2A}(R_1 - R_2)\right) \tag{4.22}$$

This yields a total current imbalance of:

$$i_l - i_r \approx \Delta i = -i_T \cdot \frac{B}{2A}(R_1 - R_2) \tag{4.23}$$

which if divided by the common transconductance of the input differential pair transistors yields the required voltage offset to rebalance the branches as a function of the difference

in memristor resistive states:

$$V_{os} \approx V_{ina} - V_{inb} = \frac{\Delta i}{g_{m4,5}} \tag{4.24}$$

which when fully unfolded yields:

$$V_{os} \approx -\frac{(R_1 - R_2) i_T}{2 R_{o,cas} g_{m,in} (1 + g_{m,cas} R_{o,in})} \tag{4.25}$$

where we have renamed our variables to explicitly stress the common values of output impedances and differential transconductances of the input differential pair and cascode transistors ($R_{o,cas}$ = output impedance of cascode transistor, $g_{m,in}$ = transconductance of the input differential pair).

Overall, Equation 4.25 shows that in small-signal conditions, the offset voltage of the core amplifier is proportional to the difference in memristor resistive states divided by the maximum transistor gains of the input diff pair and cascode transistors. This division explains the extreme fineness of tuning achievable.

The tuning range can, in principle, be extended under the rule of Equation 4.25 for as long as the underlying assumptions hold. We note two important limiting conditions: 1) If the current imbalance becomes large, the assumption of equal g_ms on both current branches collapses. Strictly when this occurs depends on the tightness of the specifications. 2) If the voltage dropped across the larger of the pair $R_{1,2}$ becomes comparable to the capacitor voltage range through which the amplifier can integrate while maintaining transistor saturation (regular operation), eventually, the amplifier will run out of integration voltage headroom.

4.3 Summary

This chapter accesses to transistor-level analysis for the proposed pre-amplifier and performance metric theoretically with equations. Since the proposed memristive OTA operates dynamically, some indicators differ from the conventional static OTA. We highlighted and explained simulations to verify the differences and these in the next chapter.

Chapter 5

Methodology of Measuring Memristor-based AC-Coupled Pre-amplifier

After defining and analysing the circuit's performance, the suitably defined performance parameters from the previous chapter are assessed for an example design in simulation. Since the capacitor integration and the control signal design are time-varying, the amplifier evaluation is based on transient simulation and periodic steady-state (PSS) analysis rather than direct DC/AC analysis. The overview of the simulation setup is listed in Table 5.1. For each performance, we develop a simulation setup and analysis. Additionally, power consumption is discussed separately. Pre-amplifier was applied to an open-loop network (Figure 5.1) to reject the DC offset from electrodes. For these simulations, we used a commercially available $0.18\mu m$ CMOS technology with VDD $= 1.8V$. The simulation of this chapter is run based on the schematic in Figure 5.1, where the detailed OTA and DLC designs are in Figure 4.1 and the sizes of transistors are in Table 7.1. A description of the methodology and simulation setup will be followed by an evaluation and design consideration based on the architecture.

5.1 Transient Simulation Setup

As mentioned above, the direct DC/AC simulation cannot be applied to this dynamic amplifier. DC indicators, transient indicators and power consumption can be obtained by utilising transient simulation in this design. Thus, the following provides a timing diagram of the whole detection cycle at first to present the operation and function of the circuit before analysing detailed performance. The timing diagram of one detection cycle is presented in Figure 5.2, and the transient simulations are based on this operation scheme.

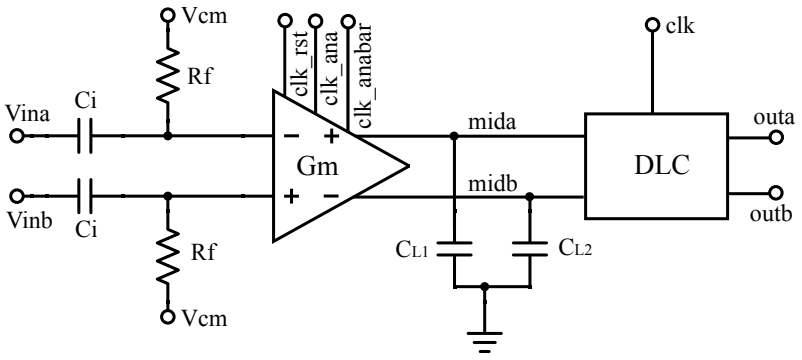


FIGURE 5.1: Block diagram of the proposed neural recording front-end. The open-loop network OTA filters the raw data to obtain the MUA. The minute signals charge the load capacitors ($C_{L1,2}$) to realise the integration and boost the voltage difference of *mida* & *midb*, which triggers the DLC to process digitisation.

TABLE 5.1: Specification and overview of simulation methodology for the memristor-based OTA.

Type	**Parameter**	**Target**	**Simulation**
DC	V_{ICR}	$\geq$ 100mV	Transient
	V_{IDR}	$\geq$ 5mV	
	V_{COMP}	$\geq V_{OS}$+500μV	
AC	A_{OL}	$\geq$ 26dB	PSS
	BW	100Hz-10kHz	
	PSRR	$\geq$ 100dB	
	CMRR	$\geq$ 100dB	
Transient	F_s	$\geq$ 20kHz	Transient
Noise	e_n	$\leq 10\mu V_{rms}$	Transient/PSS
Minimum ratings	Power dissipation	$\leq 750nW$	Transient

5.1.1 Input Range

In order to experimentally demonstrate the input range of the amplifier, we performed a series of experiments querying different potential range limitation factors in practice. First, we checked the system's behaviour at different stages as a function of common mode voltage by running a series of integration cycles whilst sweeping V_{CM} from 0V to VDD in steps of $0.1V$. At each run, the differential input was $50\mu V$ and the outputs were registered after integrating for $70ns$. Results were registered at: i) V_{midb}, ii) ΔV_{mid} and iii) the overall system output after the DLC. Results are shown in Figure 5.3. Note: To check for possible input signal history dependence during these tests, three integration cycles with $V_{CM} = 1.8V$ preceded each test integration cycle. We have sample-tested a few runs with initial V_{CM} between 0.1V and 1.8V and confirm that the history-dependence effect is negligible.

From the results in Figure 5.3 we can draw three key conclusions: 1) The DLC successfully triggers for V_{CM} between zero and $1.3V$. This means that V_{midb} is sufficiently high for the

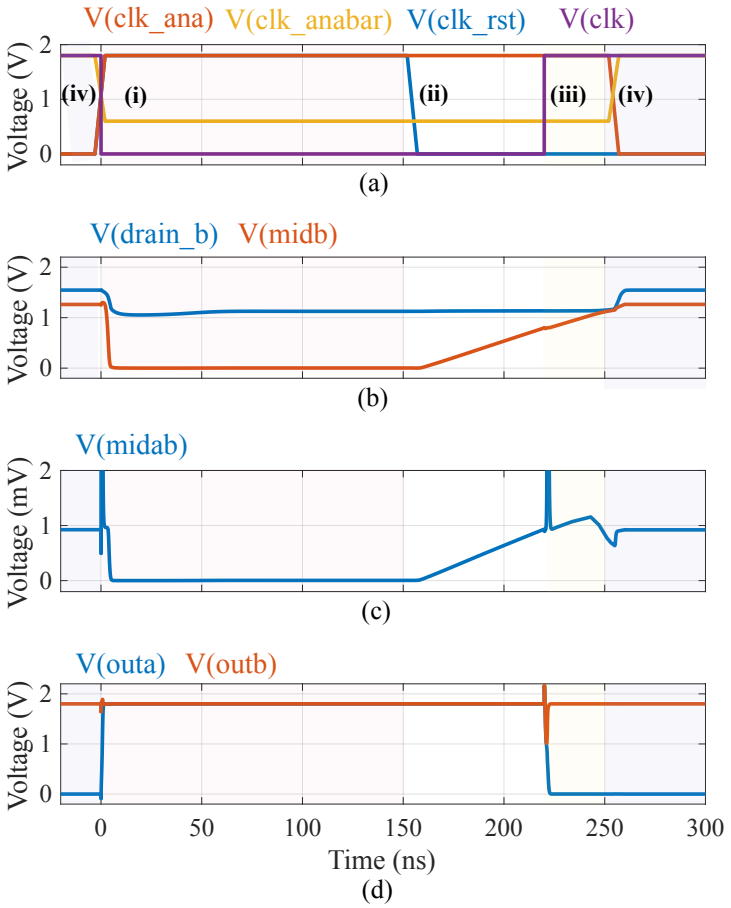


FIGURE 5.2: Transient simulation of one detection cycle of the neural spike. One detection with four phases completes in $250ns$, under the condition of $V_{ina} = 1V + 50\mu V$ and $V_{inb} = 1V$. It contains (i) reset, (ii) integration, (iii) detection and (iv) off phases. Full detection cycle is presented, including (a) control signals, (b) output of the amplifier ($midb$) and drain voltage of input transistor ($drain\_b$), (c) differential output of the OTA (ΔV_{mid}) and (d) digital outputs.

DLC to settle to an output within $30ns$ of its triggering (which occurs when clk goes high). 2) In this case, the DLC provides the correct answer so long as it triggers, but this might change towards the edges of the range once we consider noise. 3) The actual analogue gain of the amplifier remains close to the maximum ($\approx 26dB$) within a narrower region: $\approx [0.7, 1.2]V$. We recommend that the maximum gain area be taken as the effective V_{CM} range to maximise the chances of correctly capturing small differential inputs under noisy conditions. Nevertheless, this shows that by de-rating the specification of the amplifier to higher ΔV_{mid} we can extend its effective input range.

In order to visualise the effects leading to loss of gain outside the region $V_{CM} \in [0.7, 1.2]V$ we ran some unrestricted integration tests as shown in Figure 5.3 for different values of V_{CM}. The results are shown in Figure 5.4 where we observe that for V_{CM} between $0.8V$ and $1.2V$ the integration traces follow each other very closely, with traces at $0.7V$ and $1.3V$ beginning to show more substantial deviations. We note how excessively low V_{CM}s

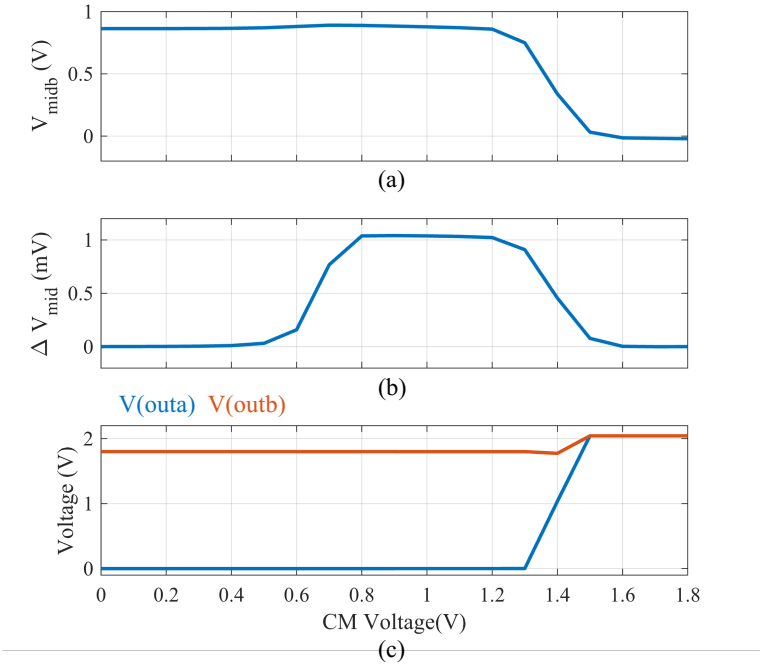


FIGURE 5.3: Input range results of pre-amplifier. In this simulation, the common mode voltage was swept from zero to $1.8V$ with $0.1V$ differential input. For $V_{CM} \in [0 - 1.3]V$, we notice that (a) V_{midb} reaches sufficiently high voltage to prompt a stable output from the (c) DLC for our chosen differential input the output is always correct. However, the core's analogue gain in (b) is maximised in the narrower range $[0.8, 1.2]V$.

shorten the peak without shifting (a result of desaturating the input differential pair but not changing the integration range) whilst excessively high V_{CM}s shift the peak without changing its magnitude.

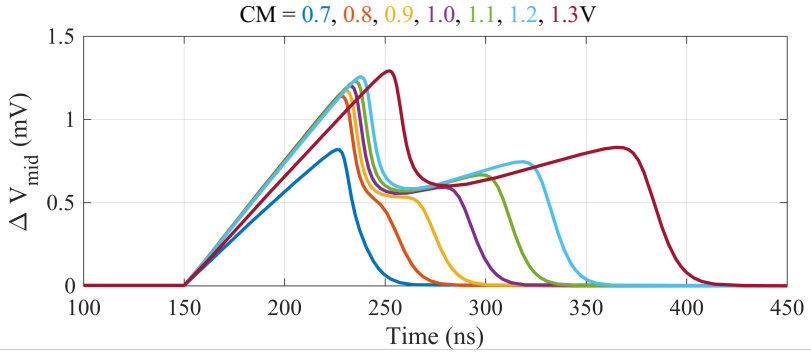


FIGURE 5.4: Intermediate differential output ΔV_{mid} evolution as a function of V_{CM}. Differential input voltage is $50\mu V$ and the integration phase is not time-constrained (see Figure 5.3). Voltage traces for different V_{CM}s follow each other closely except in the edge cases $V_{CM} \in \{0.7V, 1.2\}$.

5.1.2 Voltage Gain and Linearity

For amplifier gain analysis, we have run multiple, single data-point amplification transients sweeping a range of input differential voltages centred around zero. These simulations are under nominal conditions for this study: no added noise, mismatch or process variation was included.

There are two main experiments: First, we set an integration phase run where $\Delta V_{in} \neq 0$ and the *clk* signal does not interrupt the integration process but instead lets it run its course until both $V_{mida/b}$ saturate. Thus, the crucial features of the resulting waveform (e.g. position of peaks) are revealed. A key question we seek to answer here is whether there is an optimum time to stop the amplification from obtaining maximum gain reliably and, if so, when that occurs. The second experiment uses a fixed clock to explore the gain linearity for a fixed integration period: we run multiple simulations with ΔV_{in} swept from $-500\mu V$ to $500\mu V$ with integration period $\tau = 70ns$. The critical question is whether the amplifier has a usable linear range centred around the $0V$ differential input and, if so, how wide it is.

The first experiment is illustrated in Figure 5.5(a). We observe that for all test inputs ΔV_{in} from $-500\mu V$ to $500\mu V$ in the step of $100\mu V$. ΔV_{mid} increases linearly to a global peak at $237ns$ into the integration phase and then gradually decreases to zero. At this point, both $V_{mida/b}$ have saturated and any potential difference they had is erased. The peak occurs because as we keep integrating, the voltage at $mida/b$ nodes eventually increases to the point where the cascode transistors enter the triode mode. This causes the rate of voltage accumulation on whichever V_{mid} node is highest to slow first, allowing the other node to catch up (and leading to the post-peak drop in ΔV_{mid}). At this point, we are past maximum gain and continuing the integration eventually equalises the V_{mid}s.

Next, we note that the peak gain time is nearly perfectly aligned for all input samples; the maximum peak time difference is only $1ps$. The high quality of alignment arises because the time at which the V_{mid} voltages start trioding the cascode transistors is determined primarily by the tail current and not the differential currents. The slight discrepancy is explained by the fact that the peak gain time is technically determined by the time at which the first of $V_{mida/b}$ reaches the point where it triodes its cascode transistor. This has two critical engineering implications: 1) It allows us to set a universally optimal DLC triggering time. 2) It states that the optimal trigger time is bounded by the trioding time obtained for V_{mida} = min and V_{midb} = max (or vice versa), in which case we have the fastest trioding corner.

The results from the second experiment are shown in Figure 5.5(b). The differential output voltages ΔV_{mid} for $\tau = 70ns$ are plotted versus input differential voltage ΔV_{in}. We notice excellent gain linearity arising again from the minimal effect that the differential voltages have on the behaviour of the voltages at $V_{mida,b}$. For this experiment, the differential

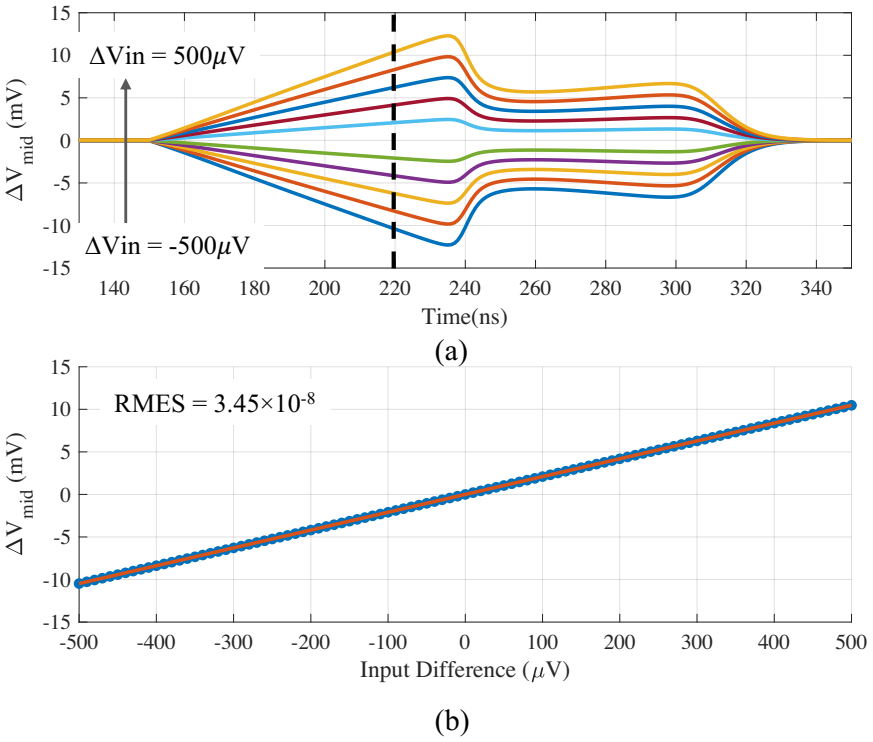


FIGURE 5.5: Simulation results of the differential gain analysis. In this simulation, *inb* was set at $1V$ while *ina* was swept from $1V - 500\mu V$ to $1V + 500\mu V$ with in steps of $10\mu V$. (a) ΔV_{mid} throughout an intentionally excessively long integration phase. As $V_{mida,b}$ increases, the cascode transistors eventually triode causing the gain to peak and decrease. Peak gain times occur at $t = 237ns$ and are aligned within $1ps$ difference. An indicative integration time leaving a substantial margin for error can be set to, e.g. $220ns$ (dashed line in (a)). (b) Output voltage difference ΔV_{mid} at integration time $\tau = 220ns$ vs input differential voltage. A linear curve excellently fits the result. The gain is constant at approx. $G = 20$.

input voltage was swept based on a fixed input $V_{midb} = 1V$ and a swept input $V_{inb} \in [1V - 500\mu V, 1V + 500\mu V]$ in the step of $10\mu V$. Results were linearly fitted yielding a gain of $G = 20V/V$ $(26dB)$ with excellent linearity throughout the range ($RMSE = 3.45 \times 10^{-8}$).

5.1.3 Input Offset Voltage

In order to access the input offset voltage, we utilise the transient simulation by multiple transient simulations such as those seen in Figure 2.15. The measurement of offset range is conducted in transient simulation by feeding a slow-increasing signal and a constant reference voltage at the inputs and recording the input voltage that changes the outputs of DLC. By tracking at what difference ΔV_{in} the outputs flip value, we can obtain an estimate for the offset. The quality of the estimate is calculated as follows: if at cycle n we had $V_{outa} = 0$ and at cycle $n + 1$ we obtained $V_{outa} = 1.8$, it means that somewhere between $\Delta V_{in}|n$ and $\Delta V_{in}|(n + 1)$ we crossed the amplifier's offset voltage. The tracking will be applied in ascending and descending phases, after which offset voltage will be

averaged. Assuming that the amplifier always decides at approximately the same relative time in each cycle, the interval is fixed and proportional to the total swept range over the number of sampling cycles.

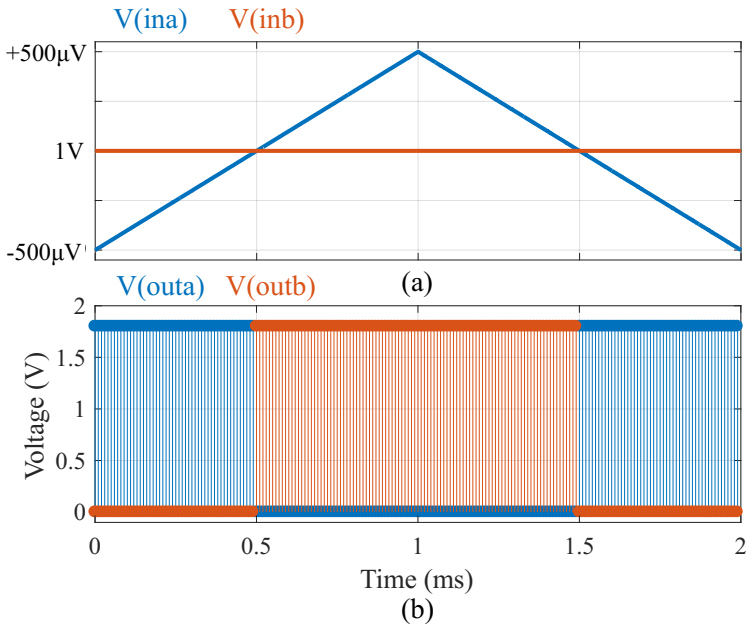


FIGURE 5.6: Pre-amplifier basic functionality test: Input A (ina) is slowly swept between $[1V - 500\mu V, 1V + 500\mu V]$ over $2ms$ while the pre-amplifier is carrying out a conversion every $1\mu s$ to detect the relationship between inputs A and B. Input B (inb) remains stable at $1V$ throughout. In this test, the amplifier was balanced ($R1 = R2$).

In our case, we run 2000 cycles ($1\mu s$/cycle for a total duration of $2ms$) and sweep the input across a range of $1000\mu V$ (ascending from $-500\mu V$ to $500\mu V$ and descending from $500\mu V$ to $-500\mu V$), yielding an offset estimate resolution of $1\mu V$. After setting up the transient simulation, we apply this to Monte Carlo simulation that includes both process and mismatch to obtain the input offset voltage. The result of the Monte Carlo simulation is presented in Figure 5.7 and the offset voltage is summarised in Table 5.2. The 99.7% (3σ) of the input offset voltage is within around $200\mu V$. Therefore, we can seek the tuneable range covering both neural spikes and circuit inherent offset which requires the input differential is greater than $700\mu V$.

TABLE 5.2: Results of a normal distribution from Monte Carlo simulation.

Parameters	Offset (μV)	Parameters	Offset (μV)
Mean	6.58	2σ	138.92
StdDev	-0.93	-2σ	-125.76
σ	72.75	3σ	205.09
$-\sigma$	59.60	-3σ	191.39

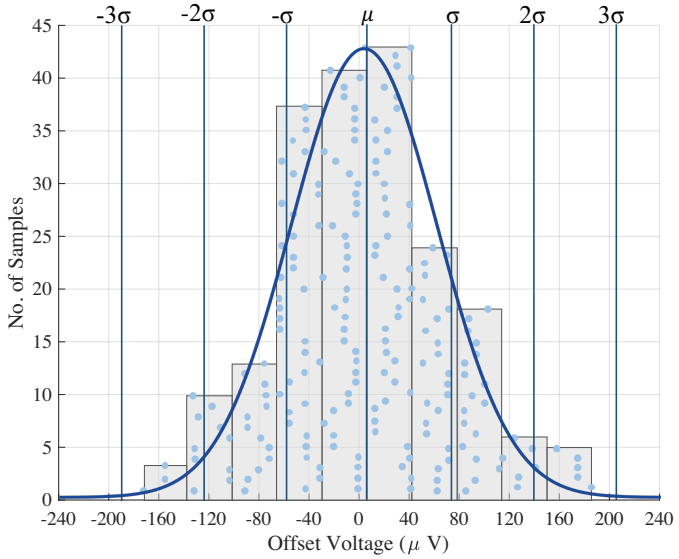


FIGURE 5.7: Histogram of Monte Carlo simulation with 200 samples of the input offset voltage.

5.1.4 Tunable Range and Sensitivity

To obtain the tuneable range and sensitivity of implanted memristive devices, multiple transient simulations such as those seen in Figure 5.6 can be repeated while sweeping both memristor device resistive states ($R1$ and $R2$).

Table 5.3 shows the offset voltage as a function of $R1, R2$. From there, we observe: 1) The overall trimming range for this particular design can cover the neural spike amplitude. 2) As expected, the maximum induced offset occurs at the maximum $R1, R2$ imbalance corners. 3) The offset sensitivity is close to $2\mu V/k\Omega$ for any combination of $R1, R2$. 4) The table is almost symmetric (as expected). The slight asymmetry indicates that the common mode voltage influences the offset voltage. This effect will be the subject of a dedicated study. Finally, the quoted offsets were checked and are the same both on the upward and the downward slopes, indicating no history dependence.

5.2 PSS Analysis

After transient simulation, we induce periodic steady-state analysis to simulate the frequency response of the proposed front-end. The front-end is operated under a discrete mode controlled by clock signals. The nodes of $mida/b$ keep increasing during the integration, which features the amplifier dynamic operation points. Moreover, the DLC is dynamic that does not have a static operation point. The conventional DC/AC analysis is not applicable to the proposed front-end. We can use the periodic steady-state analysis to complete the performance metric in this case. The PSS analysis linearises the periodic time-varying operating points and executes frequency conversion, which can be

TABLE 5.3: The offset voltage of pre-amplifier vs memristor device resistive state is quoted at $1\mu V$ resolution.

R1\R2	$1k\Omega$	$10k\Omega$	$20k\Omega$	$40k\Omega$	$60k\Omega$	$100k\Omega$	$140k\Omega$	$190k\Omega$	$240k\Omega$	$250k\Omega$
$1k\Omega$	0	50	95	160	213	319	406	512	646	701
$10k\Omega$	-53	0	48	117	172	266	353	460	519	633
$20k\Omega$	-101	-48	0	70	125	219	306	412	542	601
$40k\Omega$	-169	-117	-70	0	55	150	237	343	472	501
$60k\Omega$	-224	-172	-125	-55	0	95	182	288	419	458
$100k\Omega$	-319	-267	-219	-150	-95	0	87	192	323	362
$140k\Omega$	-406	-353	-306	-237	-182	-87	0	105	229	265
$190k\Omega$	-512	-460	-413	-343	-288	-192	-105	0	128	201
$240k\Omega$	-646	-592	-543	-472	-4109	-323	-230	-128	0	101
$250k\Omega$	-701	-633	-601	-512	-458	-362	-265	-201	-101	0

used for the periodic small-signal analysis [95]. It can assist us in conducting Fourier transform based on periodic operation points and provide more visual results in the frequency domain. The frequency data from PSS is utilised in periodic AC (PAC), periodic transfer function (PXF) and periodic noise (pnoise) that takes the noise folding effects into account. Thus, periodic analysis can be applied to this dynamic circuit to obtain the frequency-related performance such as voltage gain, bandwidth, common-mode rejection ratio (CMRR), power supply rejection ratio (PSRR) and input-referred noise.

The pre-amplifier is dynamic, and its significant state is at $220ns$ in Fig. 5.2 when *mida&midb* trigger the DLC and the outputs of DLC are in metastable point ($\Delta V_{out} = 5mV$), especially for measuring the performance of voltage gain and noise. Transient simulation needs to be conducted first to locate the time when the output metastable point appears. Then, the sample and hold module can be applied to collect the status from both *mida/b* and *outa/b* for PSS analysis. So, the simulator can present the performance directly, such as 1) mid-band voltage gain and bandwidth of OTA, 2) transfer functions for both CMRR and PSRR, and 3) noise performance at a significant state.

5.2.1 Open Loop Voltage Gain and Bandwidth

From the results from Figure 5.8(a), the $-3dB$ bandwidth of this system is $[8Hz, 16.3kHz]$ with $26dB$ mid-band gain that is eligible to cover the frequency of spikes, which is consistent to the transient simulation result. The open-loop network OTA only sets the lower cut-off frequency and maintains the initial pre-amplifiers mid-band gain and higher cut-off frequency. Meanwhile, it rejects the DC offset and allows the memristive devices to compensate only for the offset arising from PVT variations and mismatch. Compared to the previous work, the transition to a triggering scheme has reduced the gain in this implementation. However, this self-triggering regime is more robust on clock jitter and

period variations.

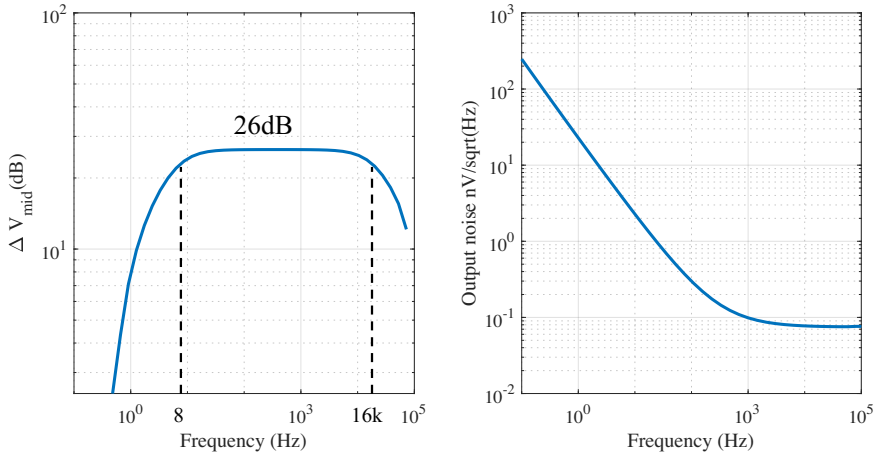


FIGURE 5.8: OTA gain and input noise as a function of frequency. (a) The open-loop network OTA presents bandpass filtering with a mid-band gain of 26 dB. (b) Output noise of the detection cycle from (periodic) noise simulation.

5.2.2 Noise Performance and Detection Accuracy

Before analysing pnoise and transient noise, we look into the kickback noise on two voltage nodes: 1) the input of the core amplifier (ina/b) and 2) the input of the DLC ($mida/b$). The kickback noise on the ina/b appears when turning on/off the core amplifier. When turning on the amplifier, bias current charges the parasitic capacitors of input transistors. The kickback noise can be discharged in $150ns$ reset phase (Fig. 5.2), and does not influence the detection phase. Moreover, the charge distribution induces kickback noise on the input nodes when we turn off the amplifier, which does not impact the operation of spike detection. In addition, the voltage changes in latch nodes are coupled to the input of DLC ($mida/b$) via the parasitic capacitors of transistors, which happens when resetting the DLC. Under the same regime, the kickback noise is settled in the reset phase. Thus, we proved that the $150ns$ reset phase eliminates the impact of kickback noise on the detection result.

pnoise analysis The pnoise simulation of the discrete-time system is different from the continuous-time system. In a discrete-time system, the noise from all phases will be integrated into the power spectral density (PSD), considering the noise folding effects. In contrast, the noise of a continuous-time system is only based on one static operation point within the frequency ranges. To measure the noise, we refer to [96] which illustrates the principle of measuring the comparator-based circuit where to capture the noise jitter (shown in Figure 5.9). As shown in Figure 5.9, 'periodic noise jitter analysis' captures the differential output level (internal latching nodes) of the DLC at the threshold point. In this case, we set the threshold point as $5mV$ where the comparator is still in unlatch stage.

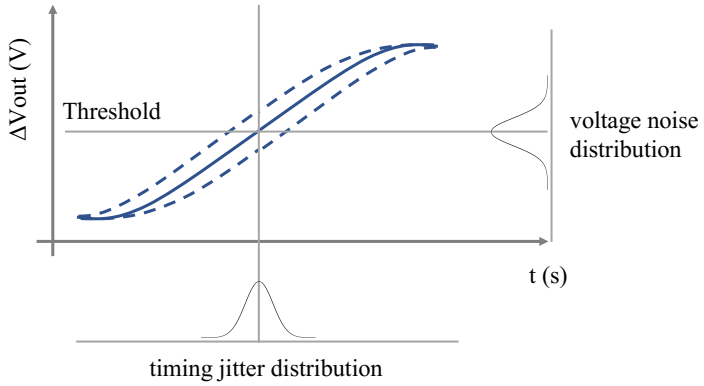


FIGURE 5.9: Principle of periodic noise jitter analysis at the threshold point.

This result can reflect the noise contribution from the OTA and the input stage of the DLC. Output noise is presented in Figure 5.8(b) and the integrated noise from $1Hz$ to $10kHz$ is $9.84 \times 10^{-4}V$. Since the noise contribution is captured at the $\Delta V_{out} = 5mV$, the input-referred noise is equal to the integrated output noise divided by voltage gain. And it yields $9.84\mu V_{rms}$ input-referred noise. The low f_c indicates that the flicker noise does not affect the system since it is out of the operating frequency range.

Detection Accuracy Transient noise simulation for testing the detection accuracy was also conducted to complement noise analysis [73]. The principle is that we fit the pre-amplifier detection accuracy to the Gaussian distribution and determine the input-referred noise as the distribution's standard derivation (σ). In this case, the transient noise should be included in this simulation. Under the condition of noise within the frequency range of [$1Hz$,$100MHz$], we ran 100 detection cycles for different ΔV_{in} from $-200\mu V$ to $200\mu V$ in the steps of $10\mu V$ respectively and obtained the accuracy in Figure 5.10. The accuracy is counted when the system outputs the correct answer as expected by the input signal. After obtaining the discrete data of the detection accuracy versus the differential voltage, we can fit these discrete data into Gaussian distribution to obtain the input-referred noise as the standard derivation (σ).

The simulation was taken under two circumstances: $R1 = R2 = 10k\Omega$ and $R1 = R2 = 250k\Omega$ to test the degree that increasing the RS induces noise and reduces the accuracy. The results show that the $250k\Omega$ RS only reduce the maximum of 2% accuracy under the low differential input voltage. It indicates that increased RS to $250k\Omega$ does not materially impact detection accuracy. In this case, the input-referred noise is $54.36\mu V_{rms}$ from the Gaussian fitting which is much larger than the pnoise result. It is because the transient result includes the latch stage of the DLC. Under the application scenario of detecting a spike, we can increase the sampling frequency to improve the detection accuracy. Considering that one detection cycle is $250ns$, the system can reach maximum $4MHz$ sampling rate.

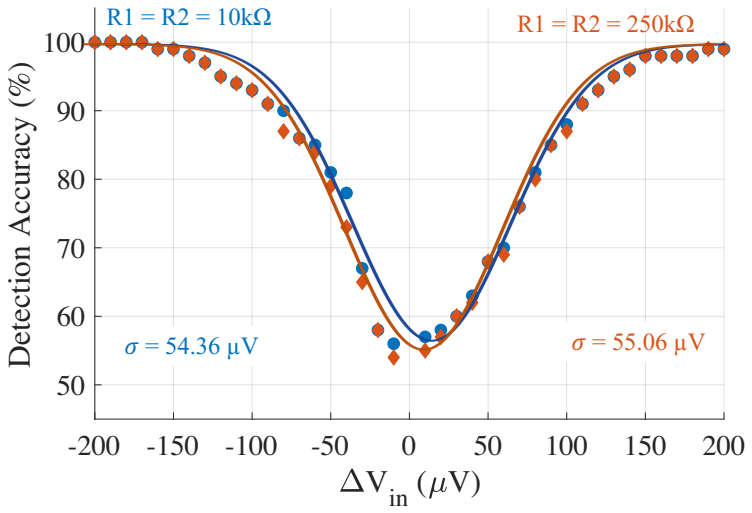


FIGURE 5.10: Detection accuracy of different ΔV_{in} from transient noise simulation under the conditions of $R1 = R2 = 10k\Omega$ (blue) and $R1 = R2 = 250k\Omega$ (orange). The Gaussian fittings and input-referred noise/standard derivation (σ) are given.

5.2.3 CMRR and PSRR

Both CMRR and PSRR are measured in PXF analysis to complement the performance metric. In PXF analysis, we sample the output nodes of DLC when $\Delta V_{out} = 5mV$ and select the response to the common mode voltage source and power supply voltage, respectively. Both CMRR and PSRR in this design are above $100dB$ within the bandwidth, presented in Figure 5.11. The performance is summarised in Table 5.4.

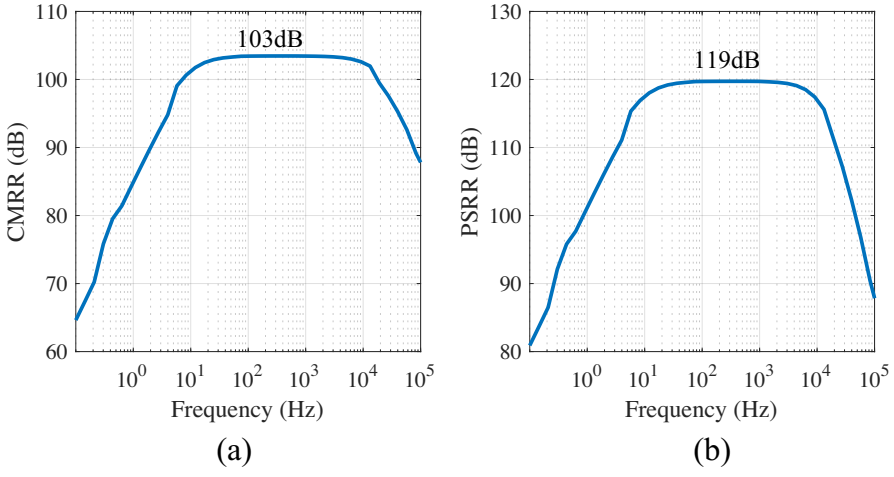


FIGURE 5.11: CMRR and PSRR of the open-loop network pre-amplifier.

TABLE 5.4: Performance metrics of the discrete-mode threshold detection system.

Performance	Value
CMOS Tech.	$0.18\mu m$
Power Supply	$1.8V$
Bias Current	$3\mu A$
Duration of one detection	(i)150 + (ii)70+(iii)30=250ns
Gain (midband)	$26dB$
Bandwidth	$8Hz - 16kHz$
CMRR	$103dB$
PSRR	$119dB$
InputNoise	$9.84\mu V_{rms}$ @ input stage $54\mu V$ with latching
Memristor tuning range	$250k\Omega$ covers $\pm 700\mu V$
Memristor tuning sensivity	$2\mu V/k\Omega$
Energy ($250ns$)	$1.44pJ$
Power ($250ns$)	$5.77\mu W$

5.3 Power Consumption

The power consumption has to be assessed for all operating phases of the pre-amplifier. The most power-hungry phase is the reset phase since it is the only one with a DC path between the power supplies. For this reason, the reset phase should be kept as short as possible. However, during the reset phase, the core amplifier reaches a steady state at all nodes so that the integrating phase can commence without any history dependence, i.e. influence from or 'memory of' its previous inputs. Next, the cost associated with the integration and digitisation phases can be split into two main components. First, the integration cost is equal to charging the core amplifier's capacitors from GND to their equilibrium level, where the integration self-terminates ($\approx 1.26V$ in our case. Noted how this integration cost currently spans both integration and digitisation phases because we do not stop the integration once we trigger the DLCs). Second, the comparison cost equals the energy needed to operate the DLC. Finally, during the 'off' phase, power dissipation is mainly down to leakages.

In this design, the tail current in the operation mode ($t_{op} = 250ns$) is $I_{op} = 3.2\mu A$ with a supply voltage of $V_{DD} = 1.8V$. Thus, the power consumption during operation is

$$P_{op} = I_{op} \times V_{DD} = 5.77\mu W \tag{5.1}$$

However, we utilised the 'start-stop' operation scheme, which has a long duration of off mode. If we operate the amplifier at sampling rates of $20k$Hz that is twice the neural spikes (one detection is in $T = 50\mu s$), the total energy consumption is

$$E = P_{op} \times t_{op} + I_{off} \times V_{DD} \times t_{off} = 1.444pJ \tag{5.2}$$

where the leakage current is $I_{off} = 17pA$ in the off mode ($t_{off} = T - t_{op} = 49.75\mu s$). Thus, the average power consumption is

$$P_{avg} = \frac{E}{T} = 28.88nW \tag{5.3}$$

In summary, this 'start-stop' operation scheme allows the system to achieve $\sim 30nW$ average power consumption per channel, whose power consumption is two orders of magnitude lower than the front-ends reviewed in Table 2.4.

5.4 Clock Skew and Jitter

This section discusses the clock skew and jitter effects on the front-end operation. It is a dynamic system with a 'start-stop' scheme, where the clocking signals control the four phases: *clk_ana*, *clk_anabar*, *clk_rst*, and *clk*. Due to the DLC being triggered by *clk* signal, we refer to this 'forced-trigger' operation, which is compared to the 'self-trigger' operation in section 6.7. The clocking signals are presented in Figure 5.12 with jitter effects presented in dash lines.

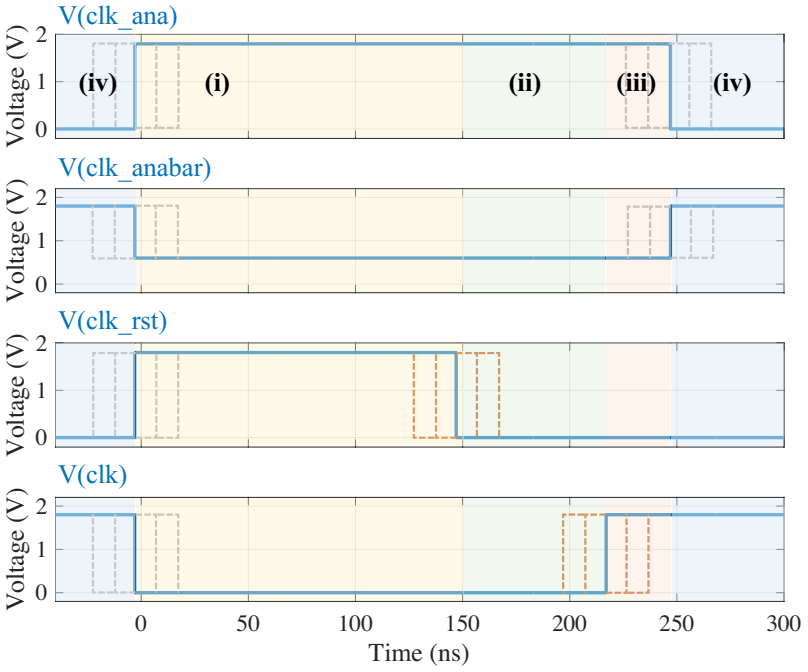


FIGURE 5.12: Clocking of the AC-coupled front-end with the jitter effect presented in dash lines. The timing diagram presents four phases: i) reset, ii) integration, iii) digitisation and iv) off phases.

Clock Skew means different arrival times of the clocking signal to the different channels. The propagation delay among channels can be neglected in this circuit for the following reasons. The operation mode only occupies $250ns$ in every detection cycle which lasts for $50\mu s$. The impact from clock skew that shifts the operation mode within the long detection cycle $50\mu s$ can be neglected in this case.

Clock Jitter means the clock signal varies at the ideal edge that the duty cycle of the clock signals is different to the ideal value, which is presented as dash lines in Figure 5.12. In this case, we divide the discussion into two parts: i) grey lines and ii) orange lines in Figure 5.12. As for grey lines, clock jitters affect the reset, digitisation and off phases, where we have set sufficient time margin that can tolerate $15ns$ jitters, which is explained below. All clocking signals control the reset phase while clk_rst determines the duration of this phase mainly. The reset phase is designed for discharging the load capacitors, which can be completed in $120ns$ and has a $30ns$ time margin. The worst acceptable case is that clk_rst shortens the reset phase to $120ns$, which indicates the front-end can tolerate $15ns$ time jitters with the consideration of the reset phase. As for the digitisation phase, it is controlled by clk_ana, clk_anabar and clk signals. In this case, $mida/b$ is fed into the DLC, and digitisation occurs in picoseconds when clk activates and activates the DLC. The worst acceptable scenario is to shorten the duration of this phase to picoseconds. Thus, we can obtain that the digitisation phase can tolerate $15ns$ jitter.

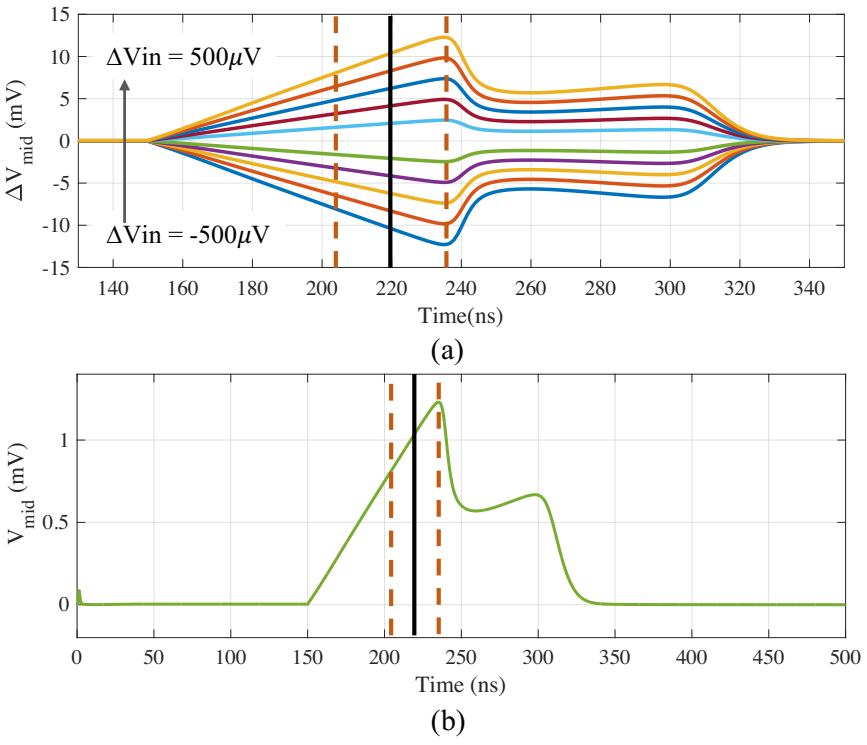


FIGURE 5.13: Simulation results of (a) differential output voltage of OTA (ΔV_{mid}) and (b) output voltage (V_{mid}). The clk signal occurs at 220ns ideally and samples the output of OTA, whilst the tolerant variation range is within orange dash lines.

The most vulnerable part is the integration phase, where the voltage gain depends on the time (shown in orange dash line in Figure 5.12 and Figure 5.13). It can be observed in Figure 5.13 that *clk* are supposed to trigger the DLC before $237ns$ in order to keep all transistors in OTA in the saturation region (details are given in section 5.1.2). The lower boundary of *clk* can be set at $180ns$ to guarantee that the output of OTA V_{mid} can be integrated to 0.5V to meet the threshold voltage DLC. With the consideration of symmetry, we determine the tolerant variation range is between $203ns$ and $237ns$ with the ideal triggering time at $220ns$. The integration phase lasts for 70ns ideally and its acceptable range is in $\pm 17ns$. Since the integration phase is controlled by *clk_rst* and *clk* with two clock edges, the tolerant jitter for each is $8.5ns$ in this case. In summary, the time jitters of *clk_ana* and *clk_anabar* are acceptable within $15ns$. The jitters of *clk_rst* and *clk* are supposed to be confined to $8.5ns$.

5.5 Simulation with Spike Train

In this section, we simulate the proposed front-end with a modelled 'spike train' with a maximum amplitude of $500\mu V$ and the frequency is $10kHz$. The spike train generation method is in section 5.5.1, followed by the simulation setup and analysis in section 5.5.2.

5.5.1 Spike Train Generation

The spike train generation schemas are presented in Figure 5.14, where the spike train outputs at the port 'AP'. In this circuit, we utilise the periodic pulse and switch to sample the sine wave to model the neural spikes train. The simulation results are present in Figure 5.15.

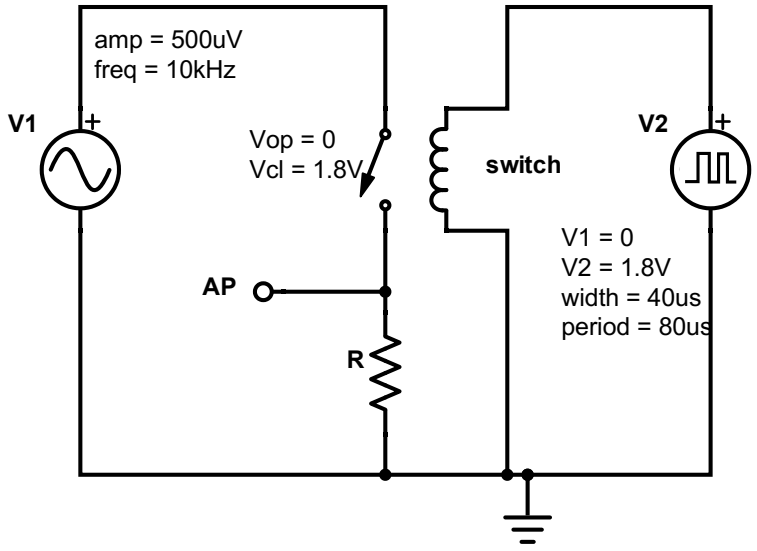


FIGURE 5.14: Schematic of spike train generation, where the output is at the port 'AP'.

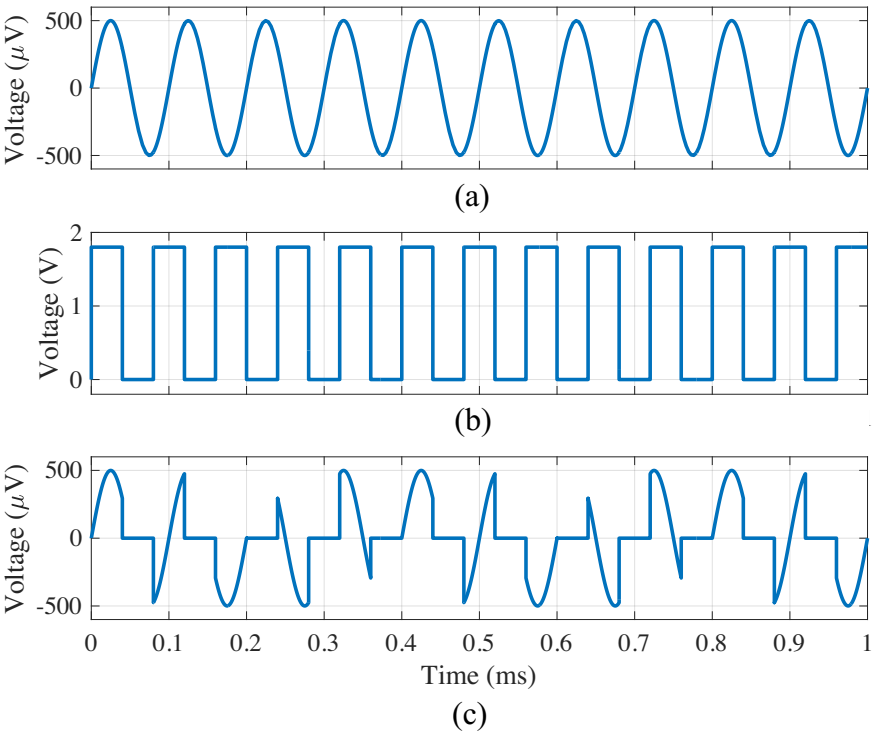


FIGURE 5.15: The timing diagram of spike train generation. (a) is sine wave with the amplitude of $500\mu V$ and frequency of $10kHz$. (b) is the periodic pulse, and (c) is the generated spike train.

5.5.2 Spike Train Simulation

After generating the spike train, we feed it into the electrode model (in Figure 2.5 and Table 2.3) to append $15.77mV$ DC offset. Then, the stimulus will be fed into the circuit under test to model the practical detection. The diagram of testbench is presented in Figure 5.16.

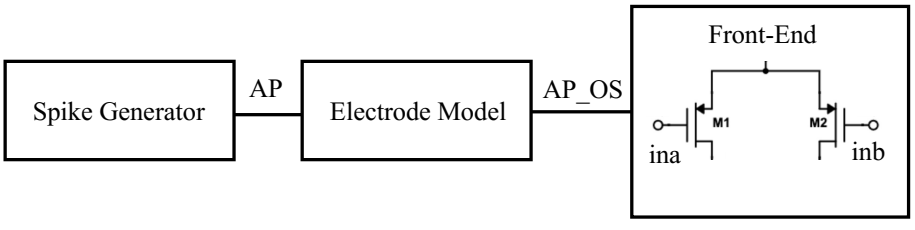


FIGURE 5.16: The block diagram presents the module and connection of the testbench.

The simulation result of the input signals is given in Figure 5.17. It can be seen in Figure 5.17(b)-(e) that the glitches occur when the OTA is turned on and off. The glitches represent the kickback noise that occurs when voltage variation in the circuit is coupled to the input, thereby perturbing the input signals. However, there exists a reset phase to discharge the load capacitors and initialise the OTA for detection. Consequently, the kickback noise can be disregarded in this design. Figure 5.17 (b), (c) and (e) describe

the two processes: i) the spike is appended with $15.77mV$ DC offset and ii) the DC offset is rejected by the highpass module at the input of OTA (in Figure 5.1).

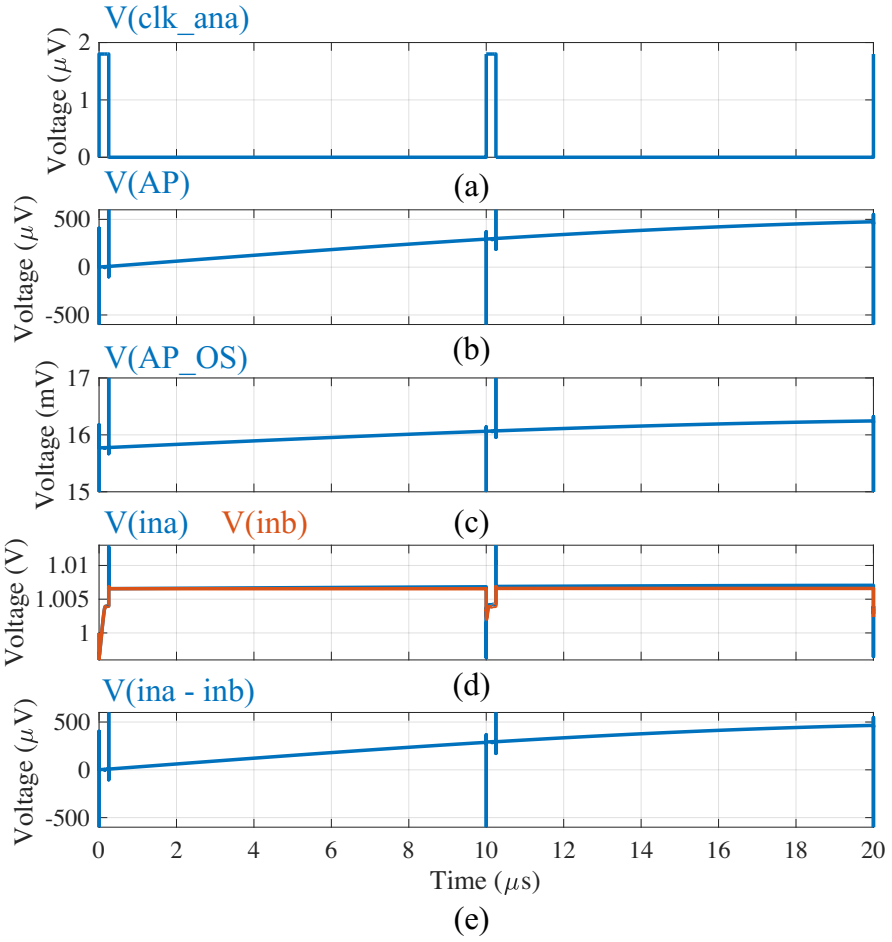


FIGURE 5.17: The timing diagram of the control signal and spikes. (a) *clk_ana* controls the on and off status of the OTA. (b) Spike train (AP) from spike generator. (c) shows the spike with 15.77mV DC offset appended. (d) presents the input voltages of OTA. (e) is the differential input voltage of OTA.

The simulation results of spike detection are presented in Figure 5.18 within $50\mu s$. In this simulation, the threshold voltage can be determined by setting the memristor RSs as $R1 = 1k\Omega$ and $R2 = 5k\Omega$ according to Table 5.3. In this case, we set the detection cycle as $10\mu s$ (sampling rate as 100kHz) to demonstrate that the system can detect the spike (above $50\mu V$) and output the digital signal in *outb*. (Noted that this simulation is noiseless.)

The practical detection is supposed to consider the noise; thus, we conduct the transient noise simulation below to determine the detection accuracy under the noise in the frequency range of $[1Hz$,$100MHz]$. Figure 5.19 compares the digital outputs of spike detection in the condition of (b) the noiseless simulation and (c) transient noise simulation, respectively. The blue lines in Figure 5.19(b)-(c) are the actual signals and we highlight the output by orange lines for more explicit demonstration. The detection is 100% ac-

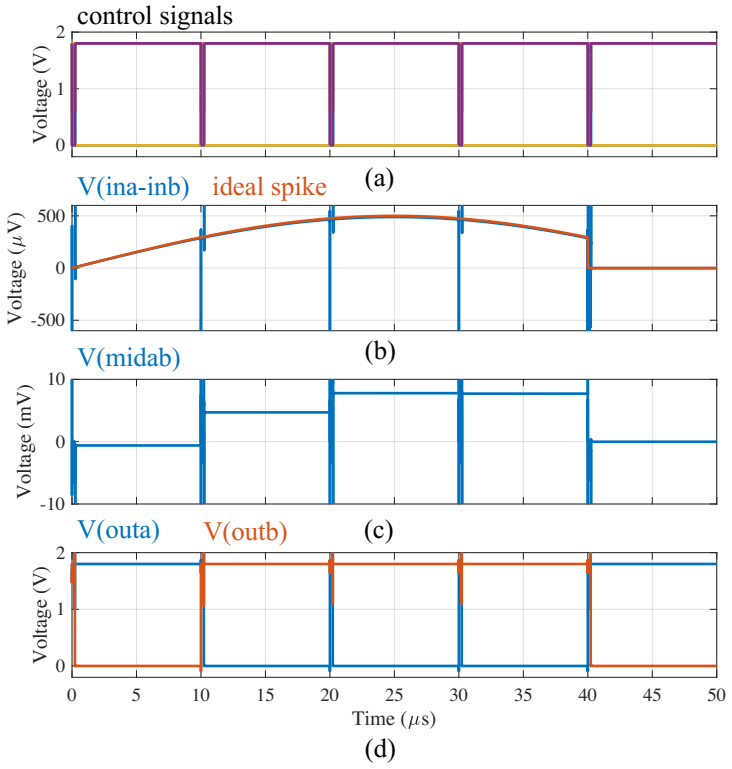


FIGURE 5.18: Simulation result of spike detection. (a) presents control signals which are the same as Figure 5.2. (b) presents the differential input voltage $ina - inb$ and the ideal spike. (c) shows the differential output voltage of OTA, and (d) presents the digital outputs.

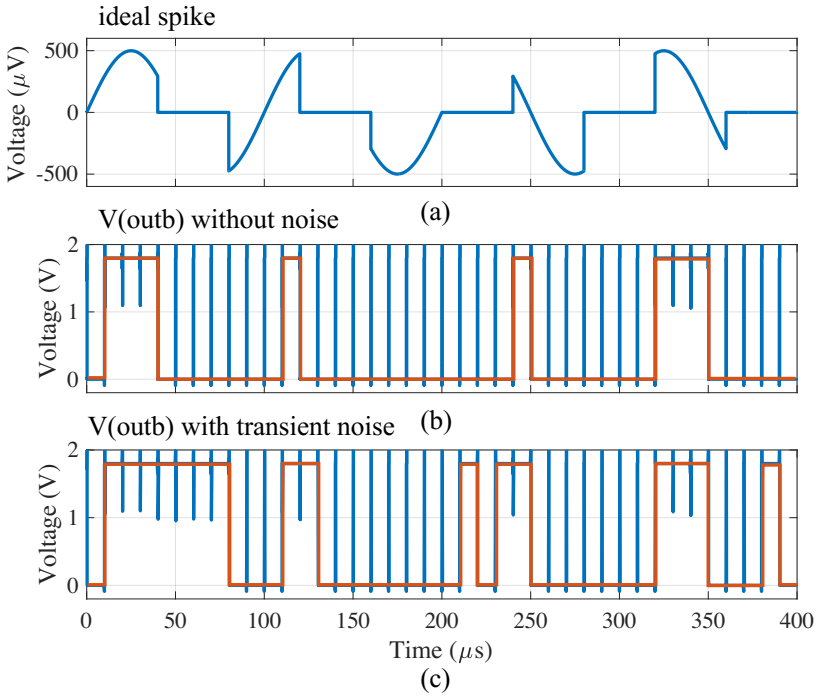


FIGURE 5.19: Simulation result of spike detection within $400\mu s$. (a) presents the ideal spike. (b) is digital output in noiseless simulation, and (c) is obtained from transient noise simulation.

curate in noiseless simulation, while it reaches 80% accuracy with the consideration of noise.

As discussed above, 'detection accuracy' refers to the proportion of correct results. The 'detection accuracy' is fixed when the circuit is designed. Furthermore, to determine whether a spike exists, we need to induce a voting module in the back-end. We can utilise the sampling frequency and 'detection accuracy' to determine whether there is a spike. For example, if the oversampling ratio is 11, accurate detection equal to or greater than six can be considered a successful capture of the spike. This can be accomplished by using permutation and combination. We refer to this as 'spike detection accuracy'. Thus, in Figure 5.19, the 'detection accuracy' is 80% and the 'spike detection accuracy' is 87.5% (correct spike detection: seven out of eight, presented in Figure 5.20(a)). In this case, the detection cycle is $10\mu s$, and the average power consumption of the front-end is $144nW$. Moreover, when we increase the sampling frequency to $200kHz$ (each detection cycle is $5\mu s$), the spike detection accuracy reaches 100%, and the average power consumption is only $288nW$.

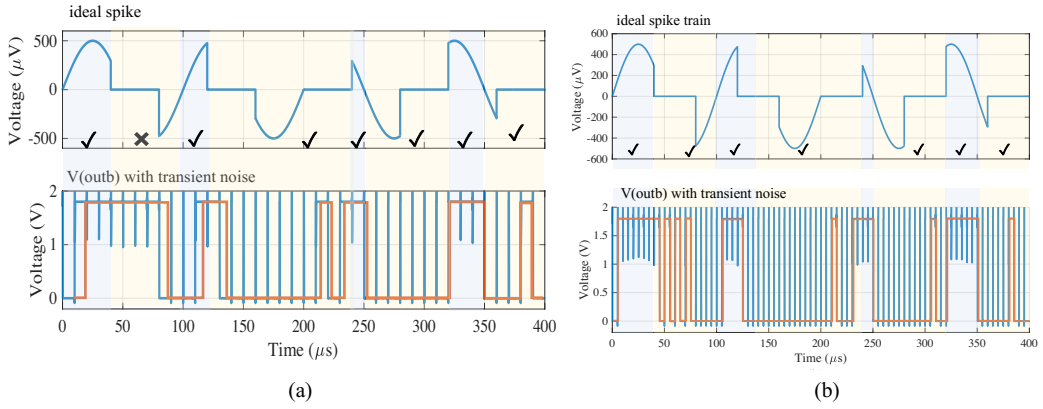


FIGURE 5.20: The timing diagram is to present the spike detection accuracy. The detection cycle is $10\mu s$ in (a) and $5\mu s$ in (b), respectively.

5.6 Discussion

From the analysis and simulation of the integrating amplifier, we highlight some key conclusions:

First, the performance improvement of the integrating amplifier over more traditional designs relies on the integration process, which enhances the gain and decreases the effective bandwidth (helping reduce noise in the process). To visualise this, let us consider an integrating amplifier using the same tail current as a standard OTA first stage. During integration, the power dissipation is the same, but the gain and bandwidth differ. In this sense, the design represents a trade-off between gain and bandwidth without changing power dissipation or using feedback.

Next, we note that there is a natural trade-off between tail current and integration time while keeping the overall energy dissipation approximately constant. This is the result of the fixed duration of the reset phase (just enough to clear any residual charge at the V_{mid} nodes) and the fact that energy consumption during the integration phase only depends on the size of the load caps and the voltage change across them during that phase. Thus, we can design a wide range of required sampling rates or bandwidths for the same energy budget.

The trade is not entirely free: Changing the tail current affects gain, bandwidth and noise performance by altering the g_ms of all transistors involved and the integration period. Furthermore, if using real memristor devices with non-linear IV curves, changing the tail current also changes the static resistance of the memristor devices. Together with changes in transistor g_ms, the tuneability range is also affected since it depends on the impedance balance between memristor and transistors. Thus, whilst the integrating amplifier offers a lot of design flexibility, the precise design trade-off space is also not trivial, much like it is for OpAmps. This is a critical subject meriting its dedicated study.

The last design decision to highlight concerns the size of the load capacitors C. The gain analysis shows that C does not affect the gain, but it does affect the integration period and, therefore, can be used to adjust the bandwidth if, for some reason, that cannot be achieved by tweaking the tail current. Effectively it is a design parameter that trades away energy for design flexibility.

In terms of operation, we note the importance of ensuring that the integrating amplifier is cleared properly in preparation for each integrating phase to avoid the output's history dependence. This means that all node voltages should be equalised across the left and right branches prior to the commencement of the sensitive integration phase. The current design achieves this by forcefully flushing the system during the reset phase. However, more energy-efficient approaches are under development as the rest phase represents a substantial fraction of the energy budget.

5.7 Summary

In this chapter, we apply the proposed OTA to the open-loop network and conduct the transient simulation and periodic steady-state analysis to complete the performance metric of the pre-amplifier. This AC-coupled pre-amplifier utilises resistors and large input capacitors to reject the DC offset. From this design, we derive the design trade-off and design consideration of this start-stop scheme. Moreover, the proposed AC-coupled front-end is proven to remove DC offset and achieve 100% spike detection accuracy (in demonstration), only consuming $288nW$ average power consumption per channel. Furthermore, this version pre-amplifier has been fabricated, and its layout design and the testing result are presented in Chapter 7.

Chapter 6

A Memristor-based DC-Coupled Pre-amplifier Topology

6.1 The Proposed Circuit

In this chapter, we propose a hybrid CMOS/memristor front-end that can compensate the input DC offset up to $50mV$. The block diagram of the front-end is presented in Figure 6.1(a). The electrode's reference voltage and neural signal are fed into the memristive OTA from nodes of *inb* and *ina*, respectively. The tissue-electrode model in Figure 6.1(b) is referred to Jochum et al. [97]. The simplified schematic is shown in Figure 6.1(c). The schematic of this OTA is similar to the AC-coupled circuit in Chapter 4 and 5, except for the position of memristors. The detailed schematic is in Figure 4.1 and the information on transistor sizes is in Table 6.1. The memristors are assembled below the input transistors to tune the branch currents in a wider range, compared to the AC-coupled solution in Chapter 5. Note that the methodology of simulating the circuit is the same with Chapter 5 provided.

TABLE 6.1: Sizes of devices in the proposed architecture, where the bias current of the core amplifier is $I_{tail} = 3\mu A$. $R3$ is replaced by a diode-connected NMOS. The detailed schematic is in Figure 4.1.

Devices	W/L ($\mu m/\mu m$)	Devices	W/L ($\mu m/\mu m$)
M1, M2	1.2/2	R3	1/1
M3	1.2/2	M11-M14	2/0.6
M4, M5	16/1	M15,M16	1/0.6
M6, M7	4/1	M17, M18	2/0.6
M8, M9	1/0.6	M19	1/0.6
M10	2/1	C1, C2	100 fF

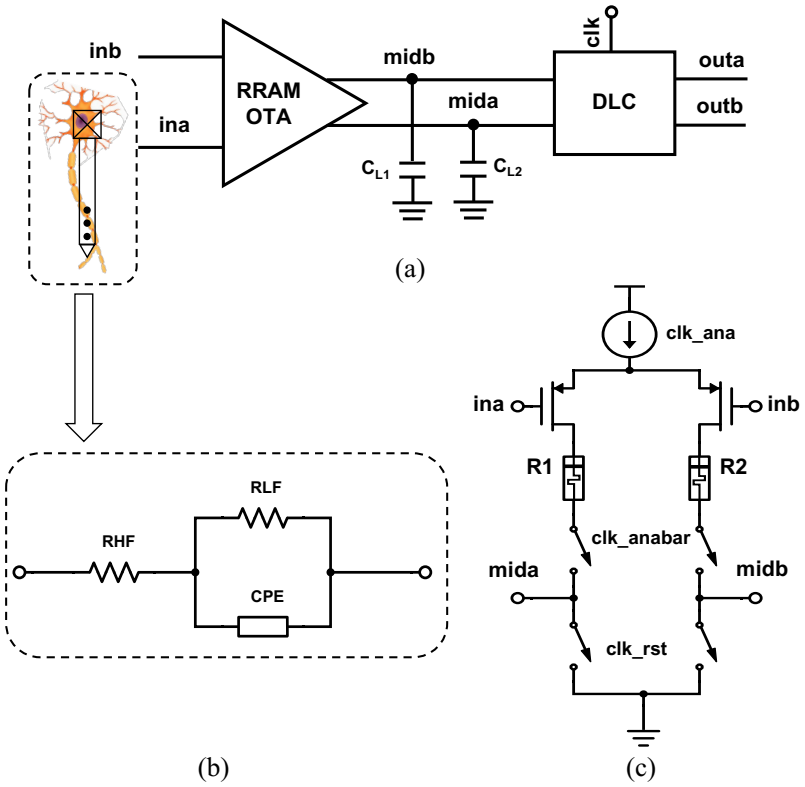


FIGURE 6.1: DC-coupled front-end. (a) Block diagram of the neural recording front-end. (b) The electrical model of the tissue-electrode interface. (c) Simplified schematic of the OTA assembled with memristive devices.

6.2 Transient Analysis

This chapter presents a new clocking scheme in Figure 6.2.The system is controlled by four clocking signals: *clk_ana*, *clk_anabar*, *clk_rst* and *clk* in Figure 6.2(a). The discrete-mode front-end is divided into three operation phases: (i) reset, (ii) detection and (iii) off phase. In the reset phase, the amplifier is switched on and the current branch from VDD to the ground is connected. The $150ns$ reset phase helps initialise the amplifier and DLC and discharge the load capacitors $C_{L1,2}$.

In the detection phase, the system conducts signal integration and digitisation in DLC. Once the $mida/b$ integrate above the threshold voltage of DLC ($0.5V$), the DLC is triggered and processes the digitisation based on the differential voltage $\Delta mid = V_{midb} - V_{mida}$ (in Figure 6.2(a)-(d)). We highlight that the 'detection phase' combines the 'integration phase' and 'digitisation phase' in Figure 2.14. In the 'integration phase' of the initial design, the DLC is off until ΔV_{mid} is integrated to the maximum voltage and the pre-amplifier enters the 'detection phase' (we named this scheme as 'forced-trigger'). On the other hand, the DLC is enabled and awaiting activation by the output of the core OTA (termed 'self-triggered'). This scheme prevents the jitters of the controlled clocking signals from fluctuating the input signals of the DLC and causing glitches.

Once the detection completes, the amplifier is turned off. The duration of APs is about a few milliseconds, and APs occur from 10 to 120 per second [45]. It indicates that the front-end can be in the off mode in the interval of spikes to save power consumption. Besides, the low spike frequency provides much headroom for oversampling.

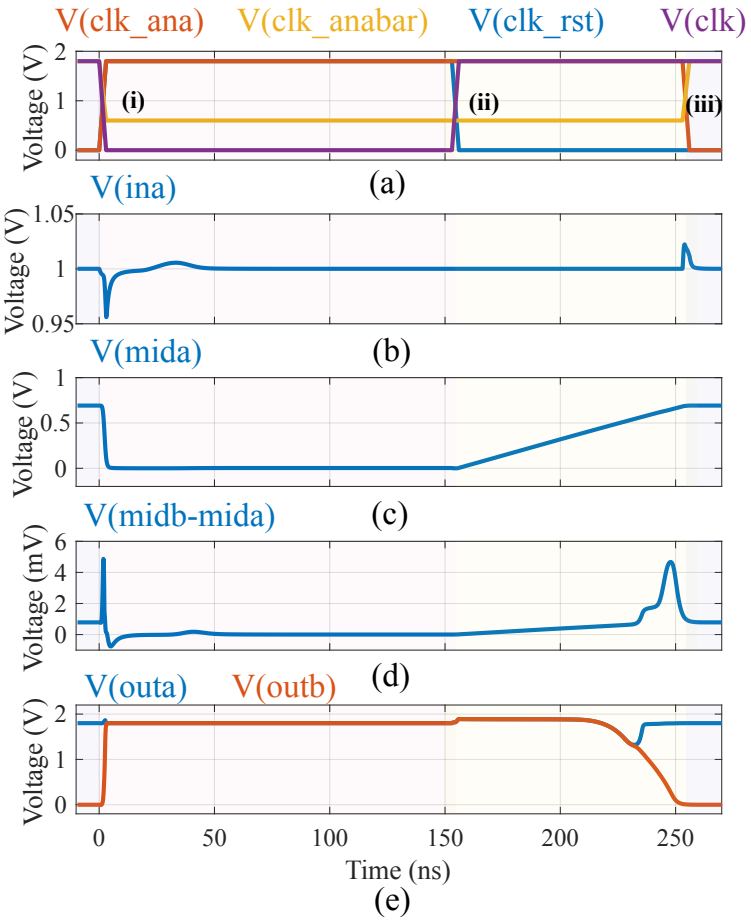


FIGURE 6.2: The timing diagram of one neural detection cycle ($250ns$). (a) clocking scheme. (b) input voltage of the front-end from the electrode. (c) the output voltage of the core amplifier. (d) the differential output voltage of the amplifier. (e) digital outputs from the DLC. (In this simulation, $ina = 1V + 50\mu V$ and $inb = 1V$.)

6.3 PSS Analysis

The primary performance metrics of the front-end are presented in Table 6.2, where most of the indicators are measured at the nodes of $mida/b$, presenting the performance of the memristive amplifier. Moreover, the energy and power consumption is of the whole front-end. The integration process can amplify the differential input signal from micro-volt to milli-volt, which makes it recognised by the DLC. The gain is $21.2dB$ within the bandwidth of $13.7kHz$. The CMRR of the memristive amplifier reaches $98.7dB$, while the PSRR is $63dB$ from the PXF analysis. Since we stated that the input-referred noise from discrete mode is not comparable to the conventional amplifier, we can only extract

that the front-end can operate in accurate DC with $394nV/\sqrt{Hz}$ white noise.

The energy and average power are measured for one detection cycle within $250ns$. As the neural spike is in low frequency and at a low occurrence rate, the average power consumption also depends on the sampling rate. Therefore, if the front-end operates under $20kHz$ sampling rate, the average power can be reduced to $29nW$. The discrete-mode operation with low power consumption provides lots of headroom to increase the sampling rate and detection accuracy.

TABLE 6.2: Performance metric of memristor amplifier (The energy and average power consumption are of one detection cycle with $250ns$, including both memristor amplifier and DLC).

Performance	Value
Power Supply	$1.8V$
Bias Current	$3\mu A$
Gain	$21.2dB$
Bandwidth	$13.7kHz$
CMRR	$98.7dB$
PSRR	$63.0dB$
Input Noise	$1/f$ corner: $21.9Hz$ thermal noise: $394nV/\sqrt{Hz}$
Energy ($250ns$)	$1.44pJ$
Power ($250ns$)	$5.77\mu W$

6.4 Tuneable Offset Range

In this simulation, we import the memristor model by Maheshwari et al. [35]. In order to measure the tuneable offset range, a testbench is designed with the resolution of $10\mu V$. For example, if the input voltage is from zero to $1mV$, we set the detection to appear every microsecond and run the simulation for 100 cycles. Then, the digital output $outa/b$ will switch at some point when the ΔV_{in} keeps increasing. Furthermore, the ΔV_{in} corresponding to the switch point is the offset input voltage.

The tuneable offset range is summarised in Table 6.4, where $R1 = 10k\Omega$ and measures the $R2$ in the step of $10k\Omega$ till the offset range reaches $50mV$. The applied memristor can reach $130k\Omega$ [27]. We can obtain from the table that: 1) the tuning sensitivity varies from different $R2$ that the larger $R2$ induces lower sensitivity, where the resolution can reach $19\mu V/k\Omega$; 2) 1% RS variation causes maximum 6.7% drift of the offset range. The accurate RSs that compensate the offset to $\{1, 10, 50\}mV$ are measured as well (shown in Table 6.3).

TABLE 6.3: The compensated DC offset versus the resistive state of memristors under R1 = 10 $k\Omega$.

R2 ($k\Omega$)	Offset Compensation (mV)
49	1
83.6	10
127.3	50

TABLE 6.4: Memristor tuneable offset range when $R1 = 10k\Omega$ and $R2$ is from $10k\Omega$ to $130k\Omega$. The third column presents the tuneable range with 1% RS variation from two memristors. The offset resolution is $10\mu V$.

R2 ($k\Omega$)	Nominal offset (V)	Tuneable offset with 1% RS variation (V)	Offset variation (%)
10	0	$-10\mu \sim 10\mu$	/
20	150μ	$140\mu \sim 150\mu$	6.7%
30	340μ	$330\mu \sim 340\mu$	2.9%
40	610μ	$600\mu \sim 630\mu$	3.3%
50	$1.09m$	$1.05m \sim 1.12m$	3.7%
60	$2.05m$	$1.97m \sim 2.14m$	4.4%
70	$4.14m$	$3.94m \sim 4.35m$	5.1%
80	$8.08m$	$7.68m \sim 8.48m$	5.0%
90	$14.09m$	$13.47m \sim 14.73m$	4.5%
100	$21.96m$	$21.10m \sim 22.84m$	4.0%
110	$31.41m$	$30.22m \sim 32.42m$	3.8%
120	$41.82m$	$40.51m \sim 43.14m$	3.2%
130	$53.17m$	$51.66m \sim 54.70m$	2.9%

6.5 Noise Analysis

Kickback noise appears on the inputs of both the memristive amplifier and the DLC. In Figure 6.1(b), the kickback noise happens when switching on/off the amplifier. When the amplifier is turned on, the bias current charges the parasitic capacitors of input transistors. However, it does not impact the detection result because the long reset phase allows the kickback to be settled before the detection phase. Furthermore, when we turn off the amplifier, the kickback noise appears due to the charge distribution, which happens in the off phase and does not influence the spike detection. From Figure 6.2(d), kickback noise exists on $mida/b$ due to the voltage changes in the latch nodes when resetting the DLC. Still, it was eliminated before the detection phase.

The detection accuracy is measured using a transient noise simulation with 250 runs. In order to estimate the noise effect, we apply resistors as memristors because there is no mature memristor noise model. In this case, we keep $R1$ as $10k\Omega$ and set $R2 \in \{49k, 83.6k, 127.3k\}\Omega$ to imitate the circuit status with $\{1m, 10m, 50m\}V$ offset compensation. Then, ΔV_{in} is swept from $10\mu V$ to $500\mu V$ in the step of $10\mu V$. The detection accuracy is in Figure 6.3.

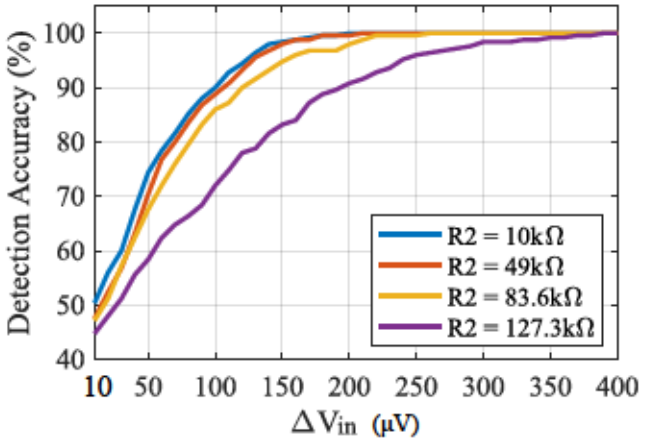


FIGURE 6.3: Detection accuracy of front-end system vs differential input voltages ΔV_{in} from $10\mu V$ to $400\mu V$. In this case, $R1 = 10k\Omega$, $R2 \in \{10, 49, 83.6, 127.3\}k\Omega$ in order to compensate the DC offset of $\{1m, 10m, 50m\}V$.

When two resistors are in $10k\Omega$ and the $\Delta V_{in} = 50\mu V$, the detection accuracy reaches 74.4%. And the accuracy can reach 100% when the differential input exceeds $200\mu V$. Even though RS is increased to compensate for $10mV$ offset, the accuracy reduces by a maximum of 9%. There exists more considerable accuracy down when the RS is capable of covering $50mV$ offset. The accuracy in this situation reaches 58.4% for $50\mu V$ differential input. However, the detection accuracy can be improved through oversampling, as described in the following section.

6.6 Oversampling

The above detection accuracy is obtained when the front-end only detects once for every neural spike. Increasing the sampling rate of the front-end and inducing majority voting on the back-end can increase the spike detection accuracy. For example, if the oversampling ratio is 11, accurate detection equal to or greater than six can be considered a successful capture of the spike. Assuming the oversampling ratio is m (an odd number), the detection accuracy is A and inaccurate possibility is $\bar{A}$, the accuracy is

$$C_m^{\frac{m+1}{2}}(A^{\frac{m+1}{2}} \cdot \bar{A}^{\frac{m-1}{2}}) + \cdots + C_m^{m-1}(A^{m-1} \cdot \bar{A}) + C_m^m(A^m) \tag{6.1}$$

which can be summarised as

$$Accuracy = \sum_{n=\frac{m+1}{2}}^{m} C_m^n(A^n \cdot \bar{A}^{m-n}) \tag{6.2}$$

The detection accuracy under the conditions of $R2 = 83.6k, 127.3k\Omega$ and $\Delta V_{in} = 50\mu V$ are 58.4% and 67.6% respectively. Moreover, the initial value can be utilised to estimate

the oversampling effect, which is presented in Figure 6.4. It can be estimated that when we aim to compensate typical $10mV$ DC offset, the spike detection accuracy can reach 95% when the oversampling ratio is 21. In this case, we simulate the input signal with a maximum frequency of $10kHz$. Conducting 21 times indicates one detection cycle is $\sim 47\mu s$ consumes $302.4nW$ average power of the front-end. Moreover, under the extreme situation of $R2 = 127.3k\Omega$, to guarantee the spike detection accuracy is 95%, the oversampling ratio needs to reach 95 and the system consumes $1.368\mu W$.

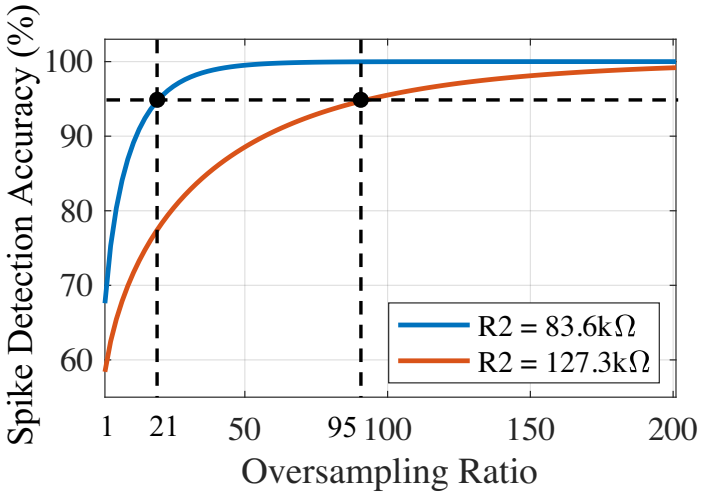


FIGURE 6.4: Spike detection accuracy under oversampling. The initial value takes from Fig. 6.3, where the accuracy of $R2 = 83.6k\Omega$ and $R2 = 127.2k\Omega$ are 67.6% and 58.4% respectively under $\Delta V_{in} = 50\mu V$.

6.7 Clock Skew and Jitter

In this section, we discuss the clock skew and jitter effects on the operation of the proposed DC-coupled front-end. This front-end operation is similar to the AC-coupled pre-amplifier operation (in section 5.4), except for the triggering of the DLC. In this design, DLC is activated after resetting the OTA and waiting for its output to trigger. Thus, we refer to this 'self-trigger' DLC. There are three phases: i) reset, ii) detection and iii) off phase, which are controlled by the clocking signals shown in Figure 6.5. The clock skew analysis has been conducted in section 5.4. And we focus on the jitter analysis in this section.

In the reset phase, the worst case that can be accepted is that the duration is shortened from $150ns$ to $120ns$ when the load capacitors are fully discharged. Thus, for each edge variation, the tolerant value is $(150 - 120)/2 = 15ns$. The detection phase contains both integration and digitisation. The detailed operation in Figure 6.2 presents that it takes $20ns$ to complete the digitisation, which indicates there is no time margin for clock jitters in this case. In order to design a front-end that is robust to clock jitters, we need to extend the duration of the detection phase. For instance, to tolerate the $15ns$ clock jitter,

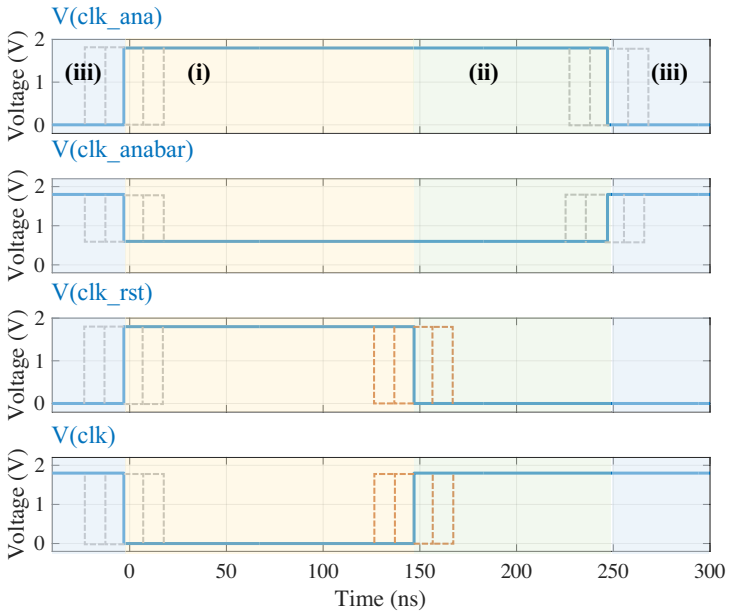


FIGURE 6.5: Clocking of the AC-coupled front-end with the jitter effect presented in dash line. The timing diagram presents three phases: i) reset, ii) detection and iii) off phases.

we need to extend the detection phase from $100ns$ to $130ns$. The operation time will be extended and increase power consumption. For example, if the front-end operates at the sampling frequency of $20kHz$, the power consumption increases from $29nW$ to $33nW$ per channel. It indicates a design trade-off between robustness to clock jitter and power consumption.

6.8 Spike Train Detection

In this section, we conduct the practical detection of spike train with the same simulation setup in section 5.5. The detailed simulation within $50\mu s$ is given in Figure 6.6, followed by the spike detection accuracy analysis.

The simulation results are present in Figure 6.6. (a) presents the ideal control signals which are the same as Figure 6.2. (b) presents the spike with an amplitude up to $500\mu V$. There are glitches at the edge of the operation mode and off mode represent the kickback noise. (c) presents the differential input signals of OTA and it can be seen that the voltage of *ina* is $\sim 15mV$ larger than the *inb* (reference). The OTA is fed with tens of DC offset and utilises the memristor to compensate. In this case, we set the $R1 = 10k\Omega$ and $R2 = 91.2k\Omega$ to compensate the DC offset ($15.77mV$) and detection threshold voltage ($50\mu V$).

This circuit operates with a 'self-trigger' DLC, which means the DLC is activated after the reset phase. After that, the outputs of OTA increase gradually through integration

and trigger the DLC once the voltage reaches the DLC threshold. The digitisation time, in this case, is longer than the 'forced-trigger' scheme due to the low input voltage of DLC. Thus, the long settling time can be seen in Figure 6.6(e) compared to the 'forced-trigger' scheme in Figure 5.18(d).

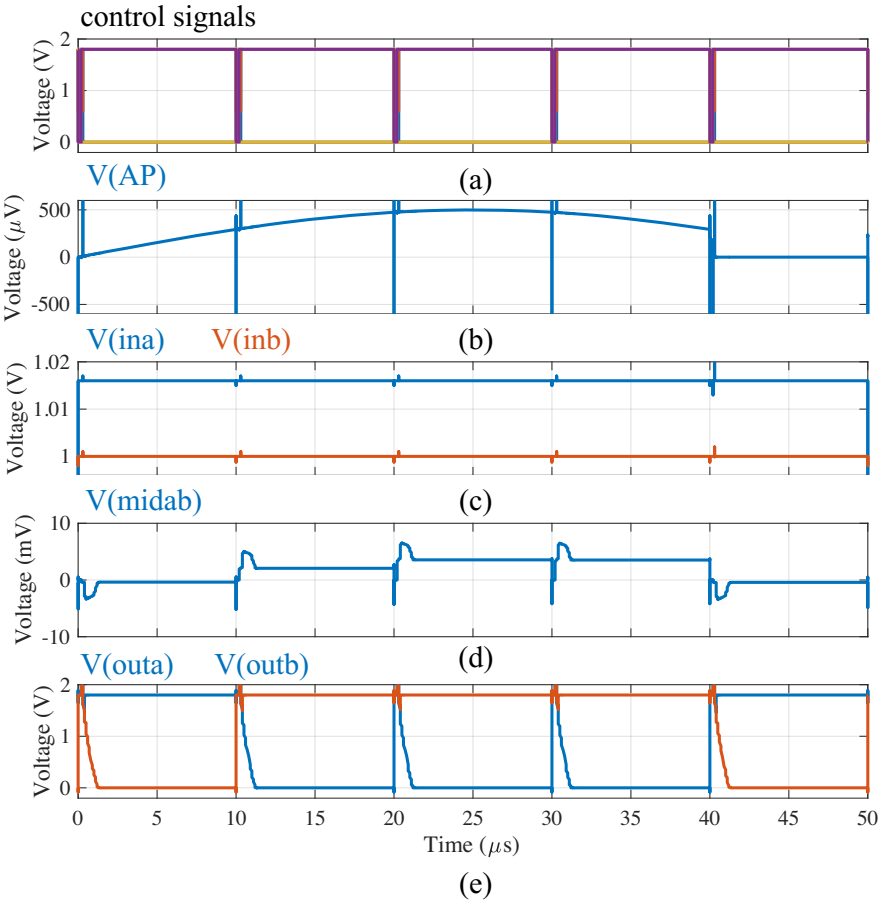


FIGURE 6.6: Simulation result of spike detection. (a) presents control signals which are the same as Figure 6.2. (b) is the spike train generated in Figure 5.16. (c) presents the input voltages of OTA. (d) shows the differential output voltage of OTA, and (e) presents the digital outputs.

The detection results with the sampling frequency of $100kHz$ and $200kHz$ are depicted in Figure 6.7. The simulation is performed in the presence of $[1Hz, 100MHz]$ transient noise. Figure 6.7 demonstrates that the spike detection accuracy can reach 87.5% even though the detection accuracy is 72.5%. It indicates that the user can detect the presence of a spike with an accuracy of 87.5%. In addition, the spike detection accuracy can reach 100% with a sampling frequency of $200kHz$. According to the clock jitter analysis in section 6.7, we can extend the operation mode from $250ns$ to $280ns$ to enhance the robustness of clock jitter at the cost of $33nW$ per channel. Furthermore, if this system is utilised at a sampling rate of $200kHz$, the average power consumption is $323nW$.

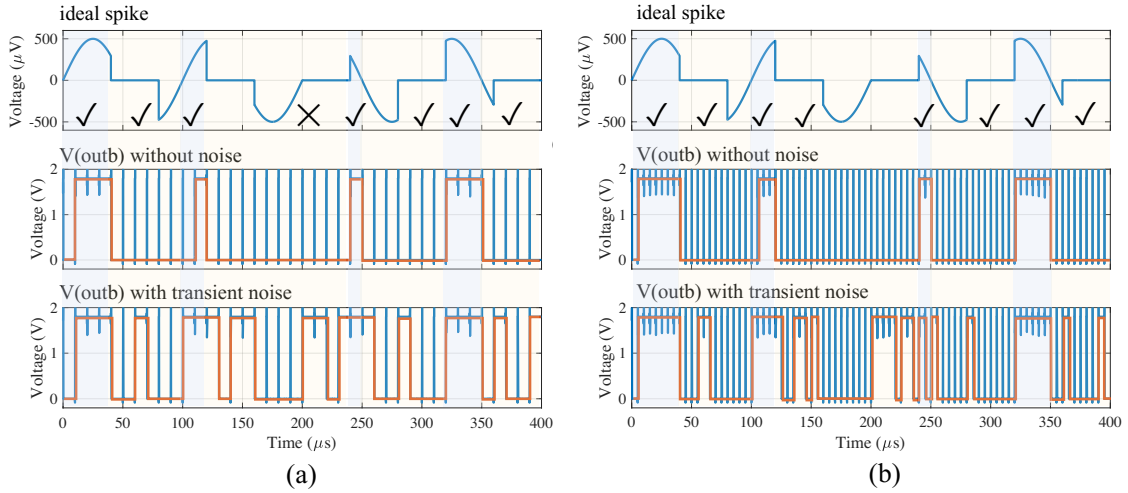


FIGURE 6.7: The timing diagram is to present the spike detection accuracy. The detection cycle is $10\mu s$ in (a) and $5\mu s$ in (b), respectively.

6.9 Summary

Memristive devices adjust the offset voltage to eliminate the need for large capacitors and resistors at the input of a DC-coupled pre-amplifier to reject DC offset. Even though the primary performance of a memristive amplifier is inferior to that of a conventional state-of-the-art neural amplifier, it is adequate for use as an integrator followed by a DLC. Simulations of transient noise have been used to test stability and robustness. The results indicate that the front-end can compensate for large DC offsets of up to $50mV$. From spike train simulation, it can be derived that the DC-coupled front-end can achieve 100% spike detection accuracy (in demonstration) at the cost of $323nW$ average power consumption.

Chapter 7

Layout Design and Chip Testing

This chapter focuses on the layout design of single channel pre-amplifier and chip test after tape-out. The details and techniques of layout design with the program unit of the memristive device are provided, followed by the post-layout simulation. In this work, we used a commercially available $180nm$ CMOS technology with $VDD = 1.8V$. Furthermore, the preparation for the chip test, including test strategy and test board design, is described. We conclude with a demonstration of the results of chip tests and the insights gained from this experience.

7.1 Layout Design

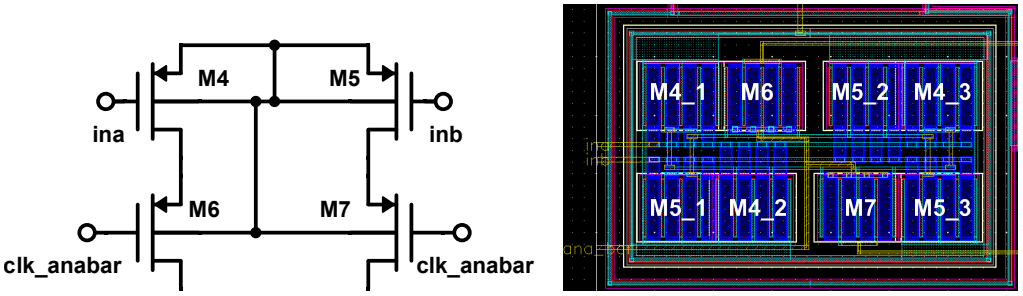
The pre-amplifier consists of both analogue and digital parts, whereas the analogue layout design requires high matching and low noise. Firstly, the $1.8V$ power supply of analogue and digital modules needs to be separated, and the programming unit is supplied by $5V$ voltage source. The most sensitive part of the circuit is the differential input pair and cascode pair; thus, the bulks are connected to the source of the input pair (in Figure 7.1(a)) with the protection of two guarding rings. There are double protections: i) the guarding ring surrounds the input pair and the cascode pair, and ii) the others isolate these two elements from other components to prevent the sensitive circuit from interfering with other modules (in Figure 7.1(b)).

In addition, we induced interdigitation and common centroid techniques in layout design to access high matching. The sizes of each component have been shown in Table 7.1, where the width of the differential input pair ($60\mu m$) is three times larger than the cascode pair ($20\mu m$). Thus, we divided the input pair into three parts and distributed them according to the common centroid in Figure 7.1(b).

The programming circuit in Figure 7.2(a) consists of a memristor and two transmission gates with the combination of NMOS and PMOS. The transmission gate is utilised to

TABLE 7.1: Sizes of devices in the proposed architecture, where the bias current of core amplifier is $I_{tail} = 3\mu A$. $R3$ is replaced by a diode connected NMOS.

Devices	W/L ($\mu m/\mu m$)	Devices	W/L ($\mu m/\mu m$)
M1, M2	3/3	R3	1/1
M3	3/3	M11-M14	2/0.6
M4, M5	60/1	M15,M16	1/0.6
M6, M7	20/1	M17, M18	2/0.6
M8, M9	1/0.6	M19	1/0.6
M10	2/1	C1, C2	100 fF



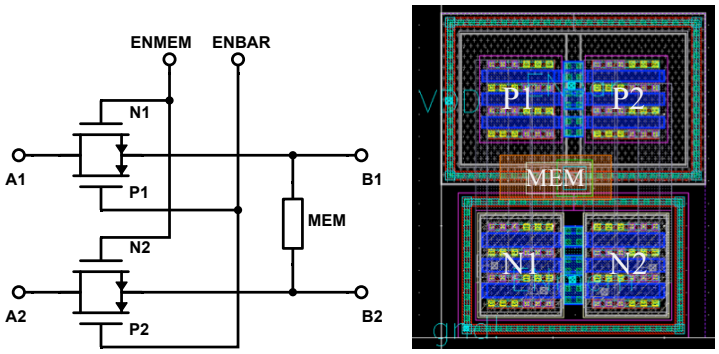
(a) (b)

FIGURE 7.1: Differential input and cascode pair in both (a) schematic and (b) layout view. The bulks of the above components are connected to the source of M4&M5.

isolate the pre-amplifier ($1.8V$ supply voltage) with the high-voltage control signal. In this case, the terminals A1&A2 connect to the pre-amplifier, while terminals B1&B2 directly connect to specified pads so that the designated signals can program the memristive device. Since the programming signal is up to $5V$, the pass-transistors are applied to shield the core circuit from damage by high voltage. In this user-isolation switch, when $ENMEM = 0$ and $ENBAR = 1$, the transmission gates are off and the memristor can be programmed by the signal from the pad, referring back to Chapter 3. When $ENMEM = 1$ and $ENBAR = 0$, the transmission gates are on so that the memristor is in the application status. The layout of the switch is shown in Figure 7.2(b).

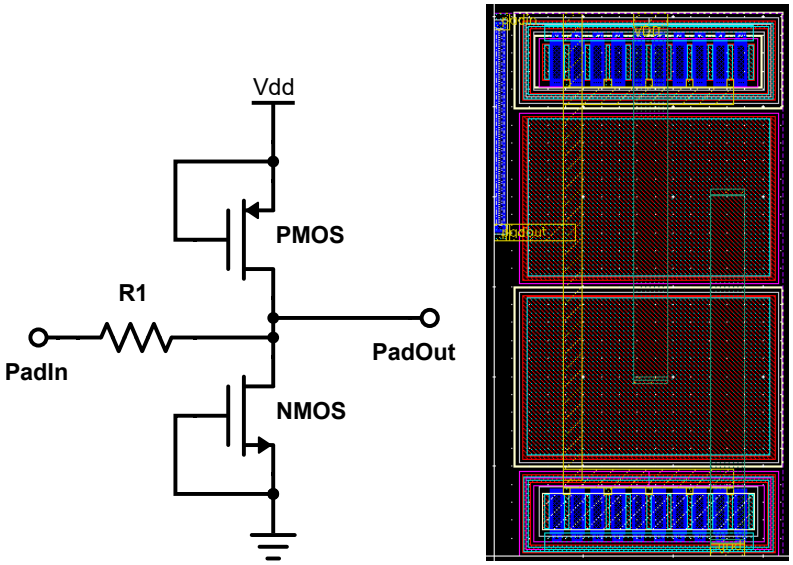
The ESD (electro-static discharge) cell is utilised for analogue and digital inputs/outputs, except for memristor programming input. For the input pin, the ESD cell works as a protection network that makes it operate under passive conditions. For the output pin, the protection element is supposed to minimise the performance and voltage degradation of the I/O circuit. The schematic of the ESD cell referring to the $180nm$ has been shown in Figure 7.3(a), and the layout following the design instruction is in Figure 7.3(b). An ESD cell occupies the area of $16.120\mu m \times 30.210\mu m$.

After accessing the essential components of the layout, we provide the layout of the core circuit in Figure 7.4. In this layout, we included 21 pins which are presented in Table 7.2, where we separated the analogue power ($1.8V$) with memristor program unit ($5V$) and the digital ($1.8V$) ones so that we can prevent the sensitive analogue module will be



(a) (b)

FIGURE 7.2: Schematic and layout of the programming circuit of the memristive device. (a) Terminals A and B can be the user terminal and programming terminal or in reverse. The status of memristive in the application or programming is controlled by two signals, $ENMEM$ and $ENBAR$. In layout (b), the memristive device is in the middle and connects to PMOS (upper device) and NMOS (lower device).



(a) (b)

FIGURE 7.3: The schematic (a) and layout (b) of ESD Cell based on TSMC180nm technology.

interfered by the digital one. In addition, it can be seen that the guarding rings are utilised to isolate modules in terms of the input and cascode pair, the current generation unit, the programming circuit, the digital DLC circuit and the substantial load capacitors. The area of the pre-amplifier is $71.280\mu m \times 32.255\mu m$.

The complete layout, including the pads, is presented in Figure 7.5. In this tape-out, we fit the pre-amplifier into 26 in-line pads, whose total area is $2080.000\mu m \times 283.670\mu m$. In this design, we utilise the pads provided by the foundry for the power, analogue and digital signals and also use in-house electrodes directly connected to the memristive devices. It is because that programming memristor requires a higher current compared to other pins.

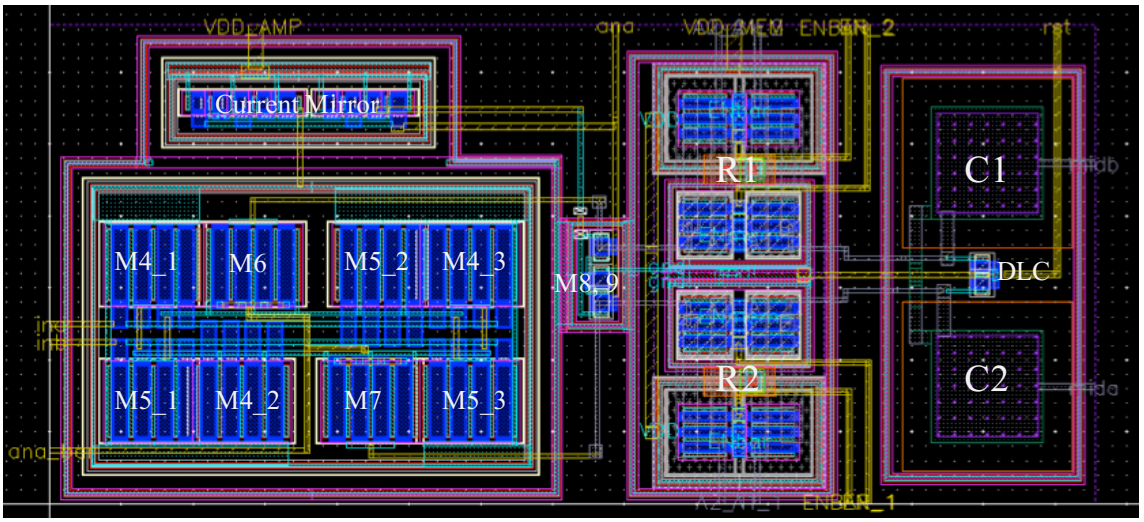


FIGURE 7.4: Layout of the core pre-amplifier circuit with the input and output on the boundary to connect the ESD cell or the pad directly.

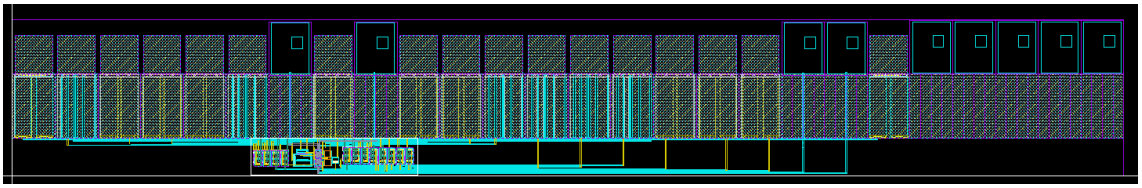


FIGURE 7.5: Complete layout design of the pre-amplifier. This design has two types of pads: i) the commercial pad from the foundry and ii) the in-house pad for programming the memristive devices.

TABLE 7.2: The pins of the pre-amplifier with the classification of type. There are some abbreviations: ana_in/out represents analogue input/output signal, ana_in_HV means high voltage analogue input which is $5V$ with the supply of VDD_MEM. Besides, dig_in/out is the digital input/output signal. pwr_x, where $x = 1.8/5$ is the supply voltage.

No.	Pin Name	Type	No.	Pin Name	Type
1	GND_AMPDLC	pwr	12	clk_rst	dig_in
2	clk_anabar	dig_in	13	clk	dig_in
3	inb	ana_in	14	outb	dig_out
4	ina	ana_in	15	outa	dig_out
5	VDD_AMP	pwr_1.8	16	VDD_DLC	pwr_1.8
6	clk_ana	dig_in	17	ENMEM_1	dig_in
7	A2_2	ana_in_HV	18	ENBAR_1	dig_in
8	VDD_MEM	pwr_5	19	A1_1	ana_in_HV
9	A1_2	ana_in_HV	20	A2_1	ana_in_HV
10	ENBAR_2	dig_in	21	GND_MEM	power
11	ENMEM_2	dig_in			

7.2 Post-Layout Simulation

After designing the layout, we extract the layout for the post-layout simulation and make the performance comparison between the pre-layout and post-layout pre-amplifier. For the post-layout circuit, the first thing is to determine the inherent offset and utilise the memristive device to compensate the offset voltage to zero. In order to obtain the inherent offset voltage, we apply the input voltage ramp in the range of $[-200\mu V, 200\mu V]$ into the circuit and capture the digital output. The result presents in Figure 7.6. It presents that the inherent offset voltage is $120\mu V$ and we tune the RS of the memristor to compensate for the offset voltage by R1 = $1k\Omega$ and R2 = $23.8k\Omega$.

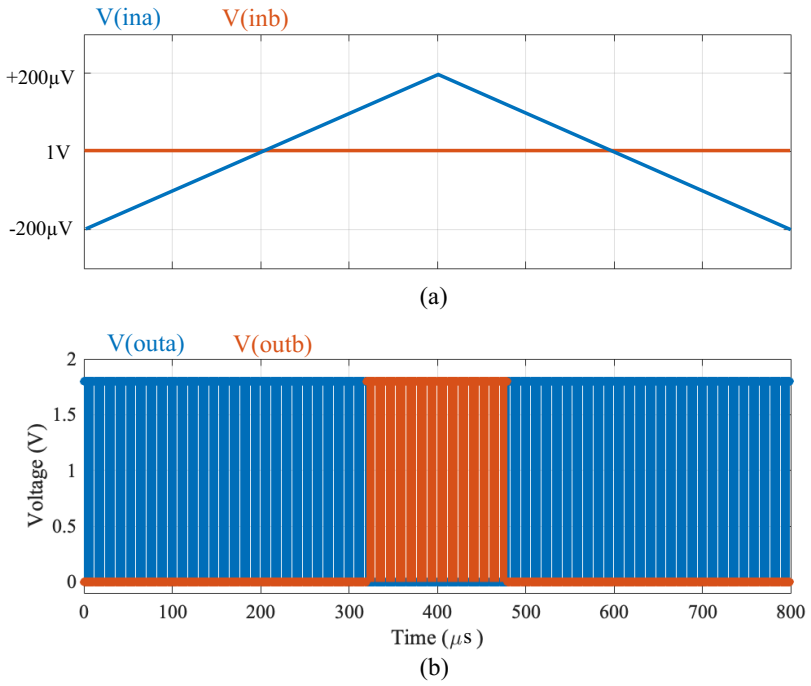


FIGURE 7.6: The input voltage ramp is applied to the circuit for the post-layout circuit to determine the inherent offset voltage. The figure shows that the inherent offset voltage is $V_{os} = V_{ina} - V_{inb} = 120\mu V$.

After calibrating the inherent offset, we run the transient simulation on this circuit in Figure 7.7. It can be seen from Figure 7.7(b)-(c) that the output of the core amplifier can be settled down within phase (ii). Moreover, the afterwards simulation can be based on this transient simulation setting.

After calibrating the circuit, we apply the same simulation methodology to this circuit to obtain a full performance metric. The comparison table between pre-layout and post-layout results presents in Table 7.3.

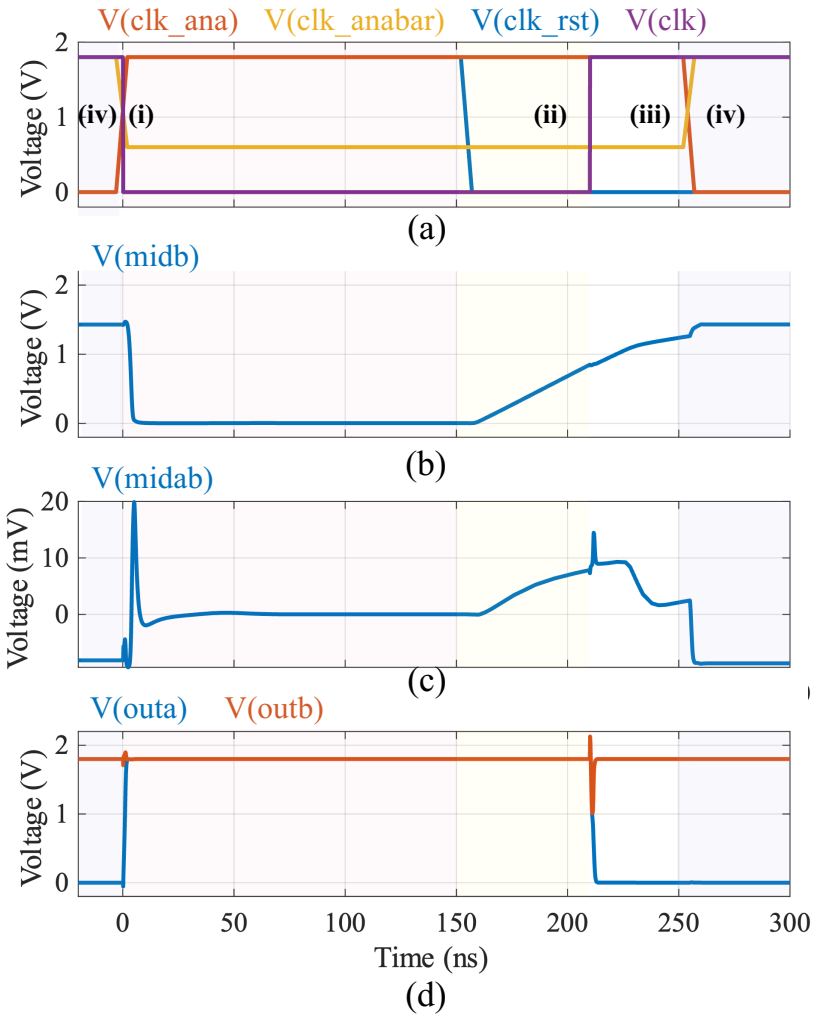


FIGURE 7.7: Timing diagram for post-layout simulation.

7.3 Strategy of Chip Testing

In this stage, we can access the function testing in terms of effective gain, input range, PSRR and CMRR, noise, tunable range and sensitivity by feeding continuous signals. In this stage, we only have access to the functional check of the signal channel pre-amplifier.

7.3.1 Programming Memristive Device

Before utilising the memristor, we need to initialise the device by electroforming to a usable resistance range (approximate $25k\Omega$ to $200k\Omega$, depending on the stack) [27]. Shown in Figure 7.2(b), the terminals with pass transistors are connected to the core amplifier. Thus, electroforming must be conducted by isolating the memristor with the core amplifier. The operation steps will be shown as follows.

1. The pass transistor needs to be turned off to ensure that the core amplifier has been

TABLE 7.3: Performance metrics comparison between pre-layout and post-layout simulation.

Performance	Pre-layout	Post-layout
CMOS Tech.	$0.18\mu m$	
Power Supply	$1.8V$	
Bias Current	$3\mu A$	
Duration of one detection	(i)150 + (ii)70+(iii)30=250ns	
Gain (midband)	$26.0dB$	$23.2dB$
Bandwidth	$8Hz - 16kHz$	$3Hz - 17kHz$
CMRR	$103.0dB$	$95.2dB$
PSRR	$119.0dB$	$78.6dB$
InputNoise	$9.84\mu V_{rms}$ @ input stage $54\mu V$ with latching	$19.3\mu V_{rms}$ @ input stage $59\mu V$ with latching
Memristor tuning range	$250k\Omega$ covers $\pm 700\mu V$	$280k\Omega$ covers $\pm 700\mu V$
Memristor tuning sensivity	$2\mu V/k\Omega$	
Energy ($250ns$)	$1.44pJ$	$1.60pJ$
Power ($250ns$)	$5.77\mu W$	$6.39\mu W$

isolated and will not be burned. In this step, $ENMEM = 0$ and $ENBAR = 5V$; meanwhile, the core amplifier is off.

2. Initialisation of memristive devices. The consecutive $1\mu s$ pulses from $8V$ to $12V$ will be applied on the device. It will induce that resistance drops from $10^6\Omega$ to a more stable range of $[10^4\Omega, 10^5\Omega]$.

3. Programming the device to a specific resistance. After initialisation, we can provoke the device by a sequence of $100ns$ $2V$ pulses to program the devices to $10k\Omega$. If the resistance is read and greater than $10k\Omega$, we can apply reversed $2V$ pulses to reduce the resistance.

Note that the pulse will be followed by the $0.5V$ triangular wave with a duration of $50ms$. The demonstration waveform has been presented in Figure 5.3. Moreover, we expect to programme the memristor to $1k - 10k\Omega$. The above setup and connection are visible in Table 7.4.

7.3.2 Measurement of the Bias Current

As shown in Figure 5.2, the duration of each phase depends on the change of voltage/current controlled by the external voltage sources. Thus, we are supposed to measure branch currents and integrate voltages at the nodes of $mida/b$ to adjust the ideal control signals from the circuit simulation.

TABLE 7.4: The pin connection for programming memristive devices.

No.	Pin Name	Connection	No.	Pin Name	Connection
1	GND_AMPDLC	gnd	12	clk_rst	gnd
2	clk_anabar	1.8V	13	clk	gnd
3	inb	1.8V	14	outb	float
4	ina	1.8V	15	outa	float
5	VDD_AMP	gnd	16	VDD_DLC	gnd
6	clk_ana	gnd	17	ENMEM_1	gnd
7	B2_2	gnd	18	ENBAR_1	5V
8	VDD_MEM	5V	19	B1_1	programming pulses
9	B1_2	gnd	20	B2_1	programming pulses
10	ENBAR_2	5V	21	GND_MEM	gnd
11	ENMEM_2	gnd			

Presented in Figure 7.2(a), terminals $B1/2$ allow us to measure branch currents and voltages of $mida/b$ when the pass transistors are turned on. In order to get the current through the memristor, we need to connect the pads of $B1/2$ to the external current-sensing device, which acquires the voltage drop across a current-sense resistor. To measure the branch current, the core amplifier will be biased to a constant state:

- $clk\_ana = 1.8V$ to allow the bias current to drain into the core amplifier.
- $clk\_anabar = 0.6V$ to bias the cascode pair (M6&M7) to saturation region.
- $ENMEM = 5V$ and $ENBAR = 0$ to connect memrisors to current branches of amplifier.
- $clk\_rst = 1.8V$ to turn on the reset transistors (M8&M9) to ensure the core amplifier is in the constant status.
- $clk = 0$ since we do not need access to DLC.
- $ina/b = 1V$ to ensure the input signal will not incline the branch currents.

The above settings are summarised in Table 7.5. The 'large' signal' of $mida/b$ is required to calculate the integrating time in phase (ii) in Figure 2.14. In this case, we will keep the control signals except for the clk_rst. The clk_rst transitions from $1.8V$ to 0 to record the rough duration. When the voltage of $mida/b$ increases from zero to one volt, we record the time as t_1. This helps determine the approximate phase duration, and a more accurate time setting will be conducted in the next step. By accessing both branch currents through the memristor and the rough integrating time, we can not determine the phase setting but verify the function of capacitor integration and the initial error between two branches through $V = \frac{i \cdot t}{C}$.

TABLE 7.5: The pin connection for measuring branch current.

No.	Pin Name	Connection	No.	Pin Name	Connection
1	GND_AMPDLC	gnd	12	clk_rst	1.8V
2	clk_anabar	0.6V	13	clk	gnd
3	inb	1V	14	outb	float
4	ina	1V	15	outa	float
5	VDD_AMP	1.8V	16	VDD_DLC	1.8V
6	clk_ana	1.8V	17	ENMEM_1	5V
7	B2_2	current sensing	18	ENBAR_1	gnd
8	VDD_MEM	5V	19	B1_1	current sensing
9	B1_2	current sensing	20	B2_1	current sensing
10	ENBAR_2	gnd	21	GND_MEM	gnd
11	ENMEM_2	5V			

7.3.3 Determining the Operation Phases

After calibrating the chip, we need to determine three operation phases: (i) reset, (ii) integration and (iii) digitisation phase by monitoring the analogue (B2_1 &B2_2) or digital (outa & outb) outputs.

The setup of measuring the reset phase is given in Table 7.6. In this case, we should monitor the differential voltage of analogue outputs ΔV_{mid} (from pads B2_1 &B2_2) by using embed LPC1768. We record the duration that ΔV_{mid} needs to settle down. In this measurement, we can try sweeping the V_{clk_anabar} and $V_{ina/b}$ to determine their effects on the reset time.

TABLE 7.6: The pin connection for measuring the reset phase.

No.	Pin Name	Connection	No.	Pin Name	Connection
1	GND_AMPDLC	gnd	12	clk_rst	1.8V
2	clk_anabar	0.6V	13	clk	gnd
3	inb	1V	14	outb	float
4	ina	1V+1mV	15	outa	float
5	VDD_AMP	1.8V	16	VDD_DLC	1.8V
6	clk_ana	1.8V	17	ENMEM_1	5V
7	B2_2	mbed sensing	18	ENBAR_1	gnd
8	VDD_MEM	5V	19	B1_1	floating
9	B1_2	floating	20	B2_1	mbed sensing
10	ENBAR_2	gnd	21	GND_MEM	gnd
11	ENMEM_2	5V			

Table 7.7 shows the setup of measuring the integration phase. In this measurement, we need to switch off the bottom transistors to allow integration on the load capacitors and recording ΔV_{mid}. In this case, we need to record the below variables: (i) sweep V_{clk_anabar} and record V_{mid} versus integration time. (ii) keep ΔV_{in} and sweep V_{inb} to measure the

V_{mid} so that we can obtain the result similar to Figure 5.4 and the common mode input range. (iii) keep V_{inb} and sweep V_{ina} to determine the differential voltage gain (similar to Figure 5.5).

TABLE 7.7: The pin connection for measuring the integrating phase.

No.	Pin Name	Connection	No.	Pin Name	Connection
1	GND_AMPDLC	gnd	12	clk_rst	1.8V → 0V
2	clk_anabar	0.6V	13	clk	gnd
3	inb	1V	14	outb	float
4	ina	1V+1mV	15	outa	float
5	VDD_AMP	1.8V	16	VDD_DLC	1.8V
6	clk_ana	1.8V	17	ENMEM_1	5V
7	B2_2	mbed sensing	18	ENBAR_1	gnd
8	VDD_MEM	5V	19	B1_1	floating
9	B1_2	floating	20	B2_1	mbed sensing
10	ENBAR_2	gnd	21	GND_MEM	gnd
11	ENMEM_2	5V			

7.3.4 Voltage Gain

Since the gain depends mainly on the integration time, we need to identify the exact proper duration for the $mida/b$ to reach the voltage to trigger DLC and acceptable gain with a time margin. Thus, we designed the process of assessing the gain based on the fabricated device, taking the ideal simulations as a reference. The process accesses the exact duration of each phase and gains as follows.

1. From capturing the voltage, we can obtain the reference time t_1 and gain g_1. In the test, the reset and digitisation time is based on the test result above. At first, we set the integration phase as t_1 for $ina = 1V + 1mV$ and $inb = 1V$.

2. Test the differential gain g_x with the integration time of $t_x = t_1 - 10ns$ and find if the peak of gain voltage appear. If the maximum gain appears (corresponding peak in Figure 5.5), we record the peak as g_{max} and time as t_{max}. If the gain decreases progressively to zero, we can move to the next step.

3. Record the gain g_x with the integration time of $t_x = t_1 + 10ns$ and find out the peak gain g_{max} and t_{max}.

4. With the consideration of time margin, we need to set the actual integrating time as $t = t_{max} - 20ns$ and acceptable gain as g with the reason in Section 4.2.3. Eventually, we will access each phase's gain and accurate testing time, which can be utilised in the following testing.

7.3.5 Noise, CMRR and PSRR

The input-referred noise can be determined by feeding fixed input voltage to the chip for 1000 cycles and recording the digital output, where we can obtain the detection accuracy directly. Then, we can sweep the differential input voltage to obtain a full map of detection accuracy versus differential input voltage, followed by the Gaussian fitting to derive the standard derivation (σ) as the desired input-referred noise. The methodology is similar to the transient noise simulation in Section 5.2.2.

The testing of CMRR is similar to the differential mode in terms of the setting of control signals. The control signals are pulses from the wave generator with different duty cycles within $1\mu s$ detection period. As for the input voltage, we keep the input difference to $1mV$ for higher tolerance to the noise. Meanwhile, change the common voltage V_{cm} with $5mV$ per step from 0 to $1.8V$. There will be five detection cycles for each V_{cm}, so we can average them to obtain a more accurate gain. After acquiring a train of gain from $V_{cm} \in [0, 1.8V]$, we can process the data using Matlab. After the analysis, it is possible to find another common mode voltage $V_{cm=x}$ that induces more stability to the circuit. A similar setup can be utilised to measure the PSRR.

7.3.6 Tunable Range and Sensitivity

In this case, we can test the offset tuning range and tuning sensitivity on the resistive state of memristive devices. Considering the duty cycle in the waveform generator, we set one detection cycle to $1\mu s$ in the tunable testing range. The testing follows the same principle in Figure 2.15, sweeping input across a range of $2mV$ with $-1mV$ to $1mV$ ascending and $1mV$ to $-1mV$ descending at $1V$ common voltage. The testing is within $400\mu s$, yielding an offset estimate resolution of $10\mu V$. In this case, we can only assess $outa/b$ for the tuneable range.

As for the signal setting:

- Import the DC $V_{cm=x}$ to inb, while the ina will be the triangular waveform from $V_{cm=x} - 1mV$ to $V_{cm=x} + 1mV$.
- The control signals are the same as above.
- Program the memristive devices to record the switching boundary of $outa/b$ which corresponds to the voltage of inb in the tuning range.

7.4 Precise Signal Generation Board for Chip Test

As mentioned above, to test the pre-amplifier chip, we need to generate precise analogue signals for *ina/b* and record the analogue signals from *mida/b* and digital signals from *outa/b*. The embed LPC1768 can record analogue and digital signals with 96MHz, indicating the device can record the signals around every 10ns. Besides, the LPC1768 can convert the analogue signals to 16-bit digital and operates at 3.3V, which indicates it can record analogue signal minimum to $50\mu V$. As for the one-bit digital signals can be directly recorded by embed LPC1768 to present as ones or zeros. In this measurement, we plan to input the differential voltage down to $50\mu V$ to check the circuit functionality and linearity.

As for the precise analogue input (to *ina/b*), we design a signal generator that utilises a 16-bit ADC (DAC8532) and ultra-low offset amplifier (OPA191). The schematic is presented in Figure 7.8.

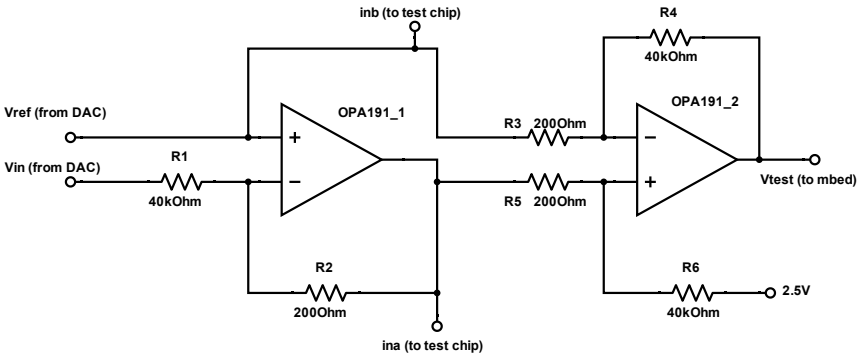


FIGURE 7.8: The schematic of utilising two OPA191 to generate and test precise analogue signals.

The schematic presents that the outputs of DAC (V_{ref} & V_{in}) feed into OPA191_1. V_{ref} is directly fed into *inb* of the chip, while the voltage difference $\Delta V_{DAC} = V_{in} - V_{ref}$ is reduced by 200 times to micron-volt level by OPA191_1, resulting $V_{ina} = V_{inb} + \Delta V_{DAC}/200$. However, due to the device offset and variation, the result feeding in *ina/b* might be shifted. To measure the input signal of the chip, we utilise the OPA191_2 to amplify the ΔV_{DAC} by 200 times and record it in embed LPC1768 as V_{test}. The principle relies on the below derivation.

$$\Delta V_{in} = (\Delta V_{DAC} + V_{os,1}) \times \frac{R_2}{R_1} \tag{7.1}$$

where, ΔV_{in} is the differential input to the chip, ΔV_{DAC} is the differential output from DAC, and $V_{os,1}$ is the offset from OPA191_1. Equation 7.1 can be converted as below.

$$\Delta V_{DAC} = \Delta V_{in} \times \frac{R_2}{R_1} - V_{os,1} \tag{7.2}$$

The mbed LPC1768 record the amplified ΔV_{in} as V_{test}. By considering the device offset,

V_{test} is

$$\Delta V_{test} - 2.5 = \frac{R_6}{R_5} \times (\Delta V_{in} + V_{os,2}) \tag{7.3}$$

where $V_{os,2}$ is the offset of OPA191_2. There exist an error between ΔV_{DAC} and $\Delta V_{test} - 2.5$, and we set it as a parameter:

$$\Delta V_{error} = \Delta V_{DAC} - (\Delta V_{test} - 2.5) \tag{7.4}$$

In this case, ΔV_{DAC} is set by the embed device and V_{test} is recorded directly by the embed device. Thus, ΔV_{error} can be derived directly. Combining the Equation 7.2, 7.3 and 7.4, we obtain a function of V_{error}

$$\Delta V_{error} = (\frac{R_6}{R_5} - \frac{R_1}{R_2}) \times \Delta V_{in} + (\frac{R_6}{R_5} \times V_{os,2} + V_{os,1}) \tag{7.5}$$

Therefore, the differential input to the chip V_{in} can be reflected by the recorded signal ΔV_{error}:

$$\Delta V_{in} = \frac{V_{error} - \frac{R_6}{R_5} \times V_{os,2} - V_{os,1}}{\frac{R_6}{R_5} - \frac{R_1}{R_2}} \tag{7.6}$$

7.5 Pre-amplifier Chip Test

In this tape-out, we got the chip that stops at the Metal 4 layer from the foundry, which allows us to assemble the memristor and memristor wiring on the top. The testing is conducted on both chips (with and without memristors).

7.5.1 Testing on the Chip without Memristor

The full view of the chip without memristor and memristor pad is presented in Figure 7.9. It can be seen that the memristor pad is missing, and we can only access the DLC with floating DLC input. The available circuit which can be accessed, is presented in Figure 7.10.

To test the DLC module of the chip, we set the pad connection as shown in Table 7.8. However, the inputs of DLC are floating that we fail to measure the full function of the DLC.

The measurement results are present in Figure 7.11. In this measurement, we import the pulse train as the *clk* into the pre-amplifier chip, where the pulse duration is 10ms with different pulse widths $\in \{2ms, 4ms, 6ms, 8ms\}$. It can be seen that the DLC can output the digital signals based on different pulse width cycles. It indicates that the pre-amplifier can be powered up adequately. However, the result shows a severe delay in the output, which the probe card and jumper wires between the chip and the test board might cause.

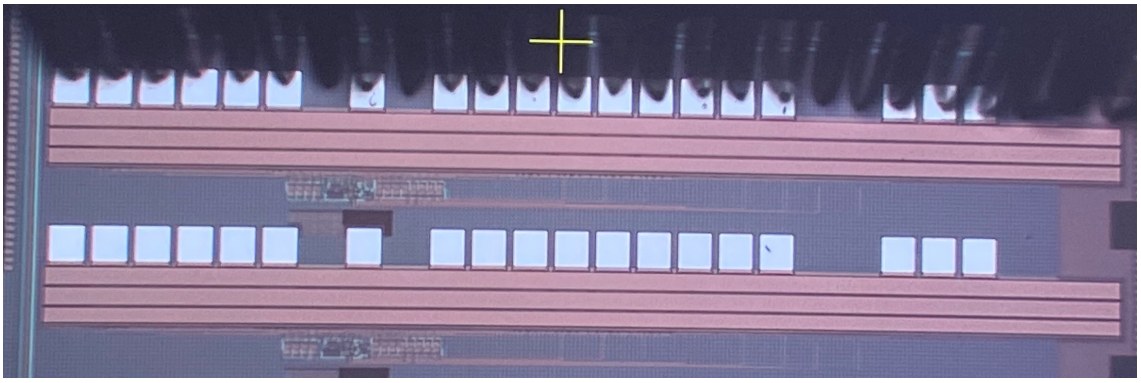


FIGURE 7.9: Full view of the chip without the memristor electrode.

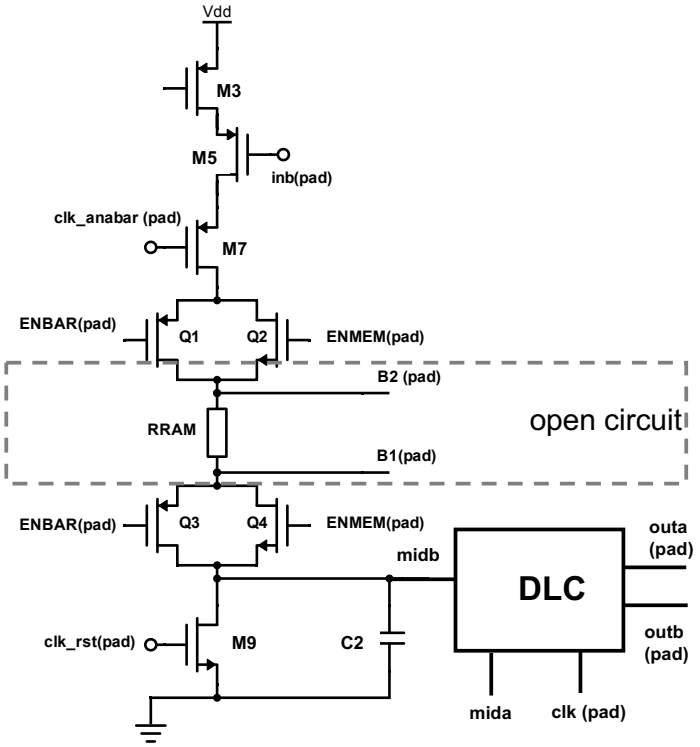


FIGURE 7.10: The schematic view presents the connection inside the chip where there exist open circuit due to missing memristor.

In addition, V_{outb} swings between 0.9V and 1.1V, which the floating input might cause. However, the reason is still unclear until we can access the nodes of $mida/b$.

7.5.2 Testing on the Chip with Memristor

The pre-amplifier chip with memristor and memristor electrodes (B2_2 & B2_1) on the top is presented in Figure 7.12. However, fabricating the memristor directly on the top shorts all Metal 4 layer, including all the power signals across the padring. It has been tested that all the power is shorted in this chip.

TABLE 7.8: The pin connection for measuring the DLC only.

No.	Pin Name	Connection	No.	Pin Name	Connection
1	GND_AMPDLC	gnd	12	clk_rst	gnd
2	clk_anabar	1.8V	13	clk	pulses
3	inb	1.8V	14	outb	mbed monitoring
4	ina	1.8V	15	outa	mbed monitoring
5	VDD_AMP	1.8V	16	VDD_DLC	1.8V
6	clk_ana	gnd	17	ENMEM_1	gnd
7	B2_2	/	18	ENBAR_1	5V
8	VDD_MEM	5V	19	B1_1	/
9	B1_2	/	20	B2_1	/
10	ENBAR_2	5V	21	GND_MEM	gnd
11	ENMEM_2	gnd			

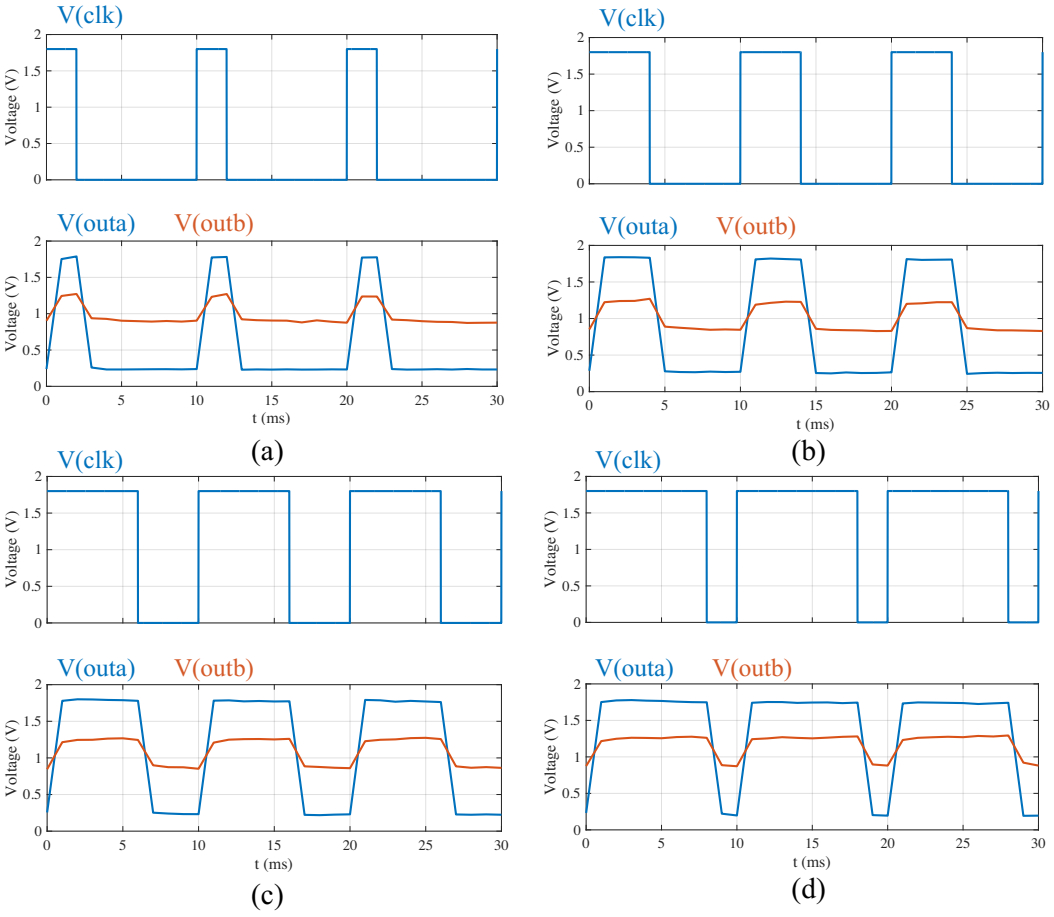


FIGURE 7.11: Testing Result of the DLC with floating inputs with different pulse widths. In this test, the pulse in the duration of $10ms$ with four pulse widths $\in \{2ms, 4ms, 6ms, 8ms\}$ are fed into the chip (presented at the top trace of sub-figures). The digital outputs are presented in the bottom trace of each sub-figure.

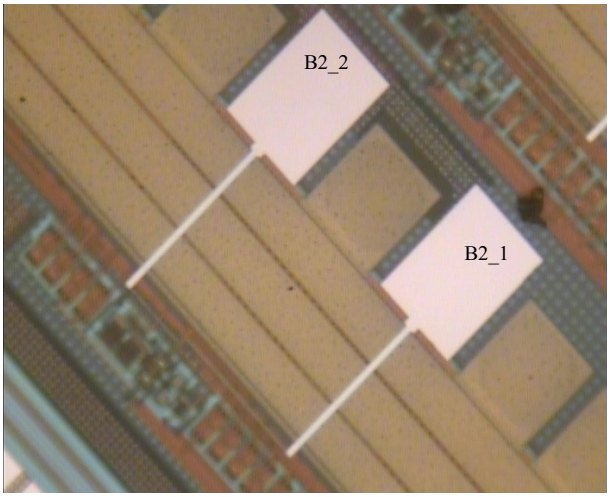


FIGURE 7.12: The pre-amplifier chip with the memristor pads and wiring on the top (in white). But the wirings across the padring short the power.

7.5.3 Circuitry Improvements

To summarise, the chip test cannot be completed in the current stage and we can only access the DLC with floating input. Nevertheless, we can still gain insights from this chip testing, improve our layout, and debug the design.

1. Choose the square padring that reduces wiring between the core circuit and pads. It can be seen that there exists a long wiring route for the in-line padring. In order to reduce the wiring and potential parasitics, we can apply a square padring for future tape-out.

2. Wiring for the memristors and pads. There are two options to overcome the short circuit in the chip. The first option is to deposit an isolation layer to cover the padring before having direct wiring as Figure 7.12 shown. The second option is to develop a new wiring route that avoids the Metal 4 layer to connect the memristors and pads, presented in Figure 7.13. Future testing will be conducted on this chip.

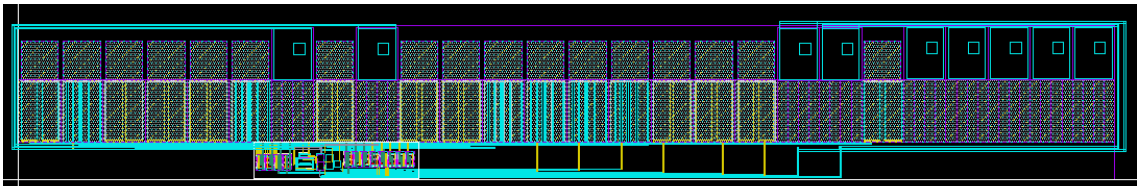


FIGURE 7.13: The full layout with new wiring route for memristors and electrodes that prevents shorting circuit.

3. Reduce jumper wires in chip testing. A severe delay on the DLC might be caused by the long wiring between the chip and the test board. To solve this, we can assemble the pre-amplifier chip in the package PLCC68. Then, the PLCC68 can be assembled on the test board and we can directly conduct the testing on the packed chip. The wire bonding in PLCC68 presents in Figure 7.14.

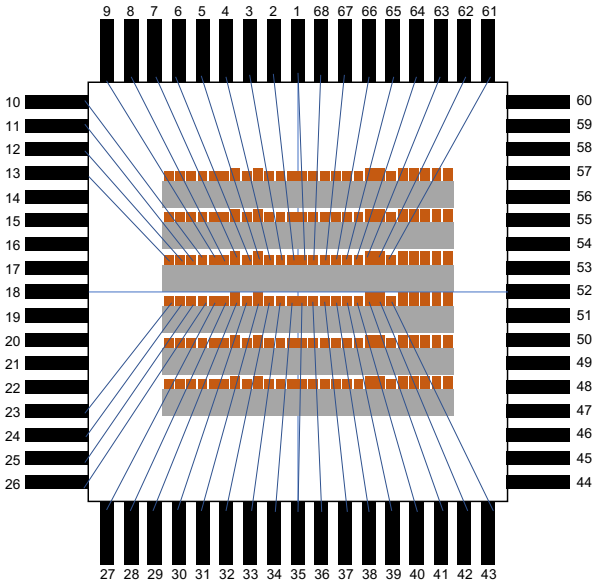


FIGURE 7.14: The full layout with new wiring route that prevents shorting circuit.

7.6 Summary

This chapter covers (i) the layout design and post-layout simulation, (ii) preparation for the chip test and (iii) test results. However, we cannot test the full functionality of the preamplifier at this point due to the chip's open circuit (missing memristors). However, we can still gain insights from testing on DLC only and develop strategies for improving the layout design and test strategy going forward. The improvements of the circuit should be based on the testing results afterwards.

Chapter 8

A Memristor-Based ΔΣADC

In this chapter, we introduce the ΔΣADC structure based on the initial design of the front-end to reduce the input-referred noise through over-sampling and noise shaping. The proposed ΔΣADC utilises the initial design as the Gm-C integrator and induces a current-steering DAC (IDAC) as an extra module. The architecture of the ΔΣADC is presented in Figure 8.1. The front-end consists of two loops: (i) a DC cancellation loop with memristive devices and (ii) a delta-sigma modulator loop filter to shape the noise to high frequencies.

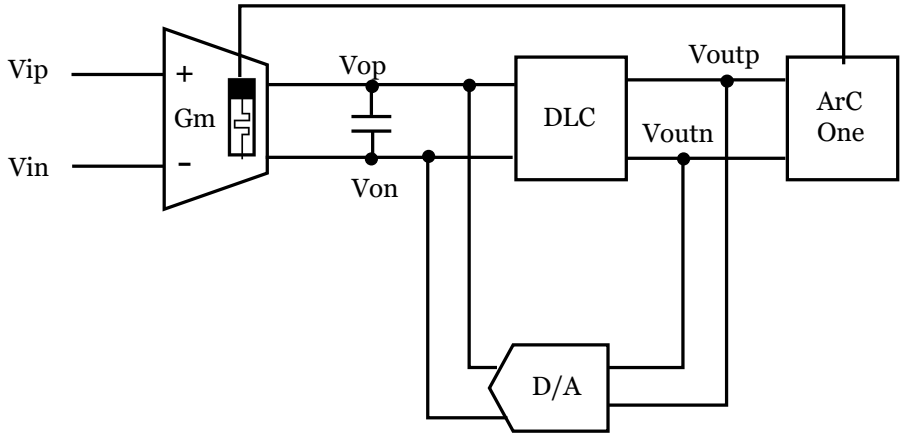


FIGURE 8.1: The block diagram of proposed ΔΣADC.

The proposed circuit utilises the ArC One as the external device to program the memristor. ArC One is the platform to validate and calibrate the memristor [98]. The operation of the ΔΣADC can be divided into two stages. Stage one is to block the delta-sigma loop, followed by the ArCOne instrument programs the memristor and cancels the offset based on the output of DLC. The ArCOne stops programming when the status of V_{outp} and V_{outn} switch, which indicates the DC offset is compensated. The second stage is to block the Offset modulation loop and turn on the delta-sigma loop to detect the neural spike with low noise.

8.1 Architecture of ΔΣADC

In this thesis, we proposed Gm-C based one-bit continuous-time delta-sigma (CT-ΔΣ) ADC modulator. The CT techniques feature an inherent anti-aliasing filter, which reduces system components and energy consumption. However, the first-stage integrator occupies a large percentage of overall power. As for the integrator, the active-RC integrator produces high linearity at the cost of more power required to run the feedback network with enough loop gain and phase margin. On the other hand, the Gm-C integrator utilises less power than the active-RC integrator while driving an open capacitive load loop. The most challenging aspect of developing a Gm-C-based modulator is the limited linear input range caused by the nonlinear Gm amplifier, which results in a low SNDR. Nevertheless, this system is applied for the neural signal that input changes are in the millivolt range. It enables the development of a single-stage OTA without compromising power to improve linearity.

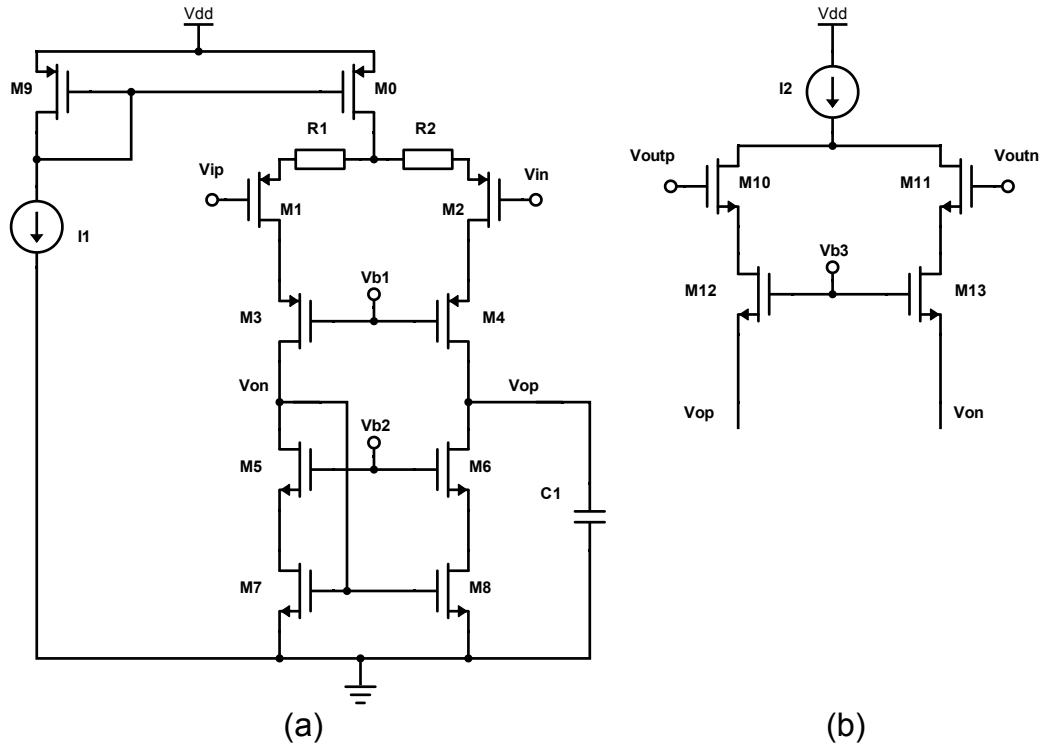


FIGURE 8.2: The schematic of OTA of ΔΣADC. (a) The OTA is the input and the integrator interface in the ΔΣ modulator. (b) The IDAC shown on the right creates a current feedback loop.

The core of ΔΣ modulator is built with a Gm-C integrator, current-steering DAC (IDAC), and a DLC (shown in Figure 8.1 and Figure 8.2). The sizes of OTA are presented in Table 8.1 while the sizes of DLC are the same as the previous design in Chapter 5. In this case, we utilise the oversampling ratio (OSR) of 512 to achieve low quantisation error with the signal bandwidth of 10 kHz. The OTA operates as the input stage of the channel and

the integrator interfaces. The OTA consists of a current source (M0), input differential pair (M1&M2), the active load (M7&M8), and the cascode pairs (M3-M6). It operates under the bias current of 3 μA. Its output is coupled to the load capacitor (C1), IDAC and DLC. A high output impedance is essential to preserve the noise transfer function of the modulator. The output is equipped with cascode pairs (M3-M7) to increase the output impedance.

TABLE 8.1: Sizes of devices of OTA in the proposed ΔΣ ADC.

Devices	W/L ($\mu m/\mu m$)	Devices	W/L ($\mu m/\mu m$)
M1, M2	8.5/6	M10, M11	1/5
M3, M4	2/2	M12, M13	1/5
M5, M6	2/5	C1	200 fF
M7, M8	1/18	I1	$3\mu A$
M0, M9	3/1	I2	$240nA$

In this design, we utilise the feedback path constructed in the current domain to eliminate the need for an additional pair of input transistors. In order to minimise the area even further, asymmetric current feedback is achieved by utilising the current source solely. The input of the IDAC is the bitstream output by the quantiser (V_{outp}&V_{outn}). A common logic block generates the control signals for these two blocks from the quantisation bitstream. The IDAC feeds the regulated current of 240 nA to the output of the OTA (V_{op}&V_{on}) in a full-scale mode of 10 mV.

Similar to the above front-end designs, the proposed ΔΣADC induces a pair of memristive devices (R1&R2) to compensate for the DC offset. The programming module is the same as above. In this case, the memristors are put above the input transistors, operating as source degeneration. Meanwhile, it can compensate the same DC offset (50 mV) with a lower resistive state. The offset compensation is summarised in Table 8.2.

TABLE 8.2: The compensated DC offset versus the resistive state of memristors under the condition of R1 = 1 $k\Omega$.

R2 ($k\Omega$)	Offset Compensation (mV)
5	8.24
10	18.70
15	28.87
20	39.00
25	49.00

8.2 Performance ΔΣADC

The operation of the ΔΣADC is given in Figure 8.3 under the sine 10 mV_{pp} input. The feedback modulation on the nodes of V_{op} and V_{on} is given by the IDAC. After collecting the output data, we conduct the Fast Fourier Transform (FFT) to obtain the output spectrum

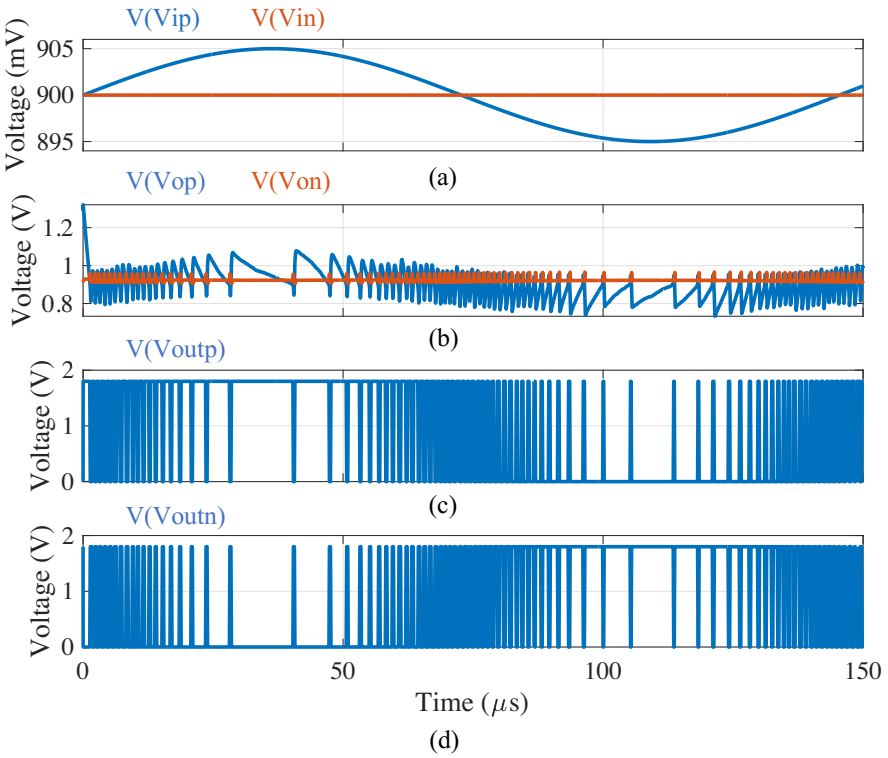


FIGURE 8.3: The timing diagram of ΔΣADC. (a) the input of the system. (b) the output of the OTA. The one-bit IDAC provides the modulation. (c) and (d) are the outputs of the DLC, respectively.

in Figure 8.4(a) with the noise bandwidth of [0.1Hz,100MHz]. The spectrum presents that the noise is shaped to a higher frequency range so that the front-end can keep the low noise level in the signal band. It can achieve 60.22 dBc Spurious-Free Dynamic Range (SFDR) and 53dB Signal-to-Noise And Distortion Ratio (SINAD), which indicates 8.51 Effective Number Of Bits (ENOB). In addition, Figure 8.4(b) presents the SNR versus the input signal in dB, which indicates the 55 dB Dynamic Range (DR) of the system (AC input range). As for the DC input range, it is measured at the input differential pair that the input voltage keeps the input transistor in saturation/sub-saturation region. The accepted DC input range in this design is 200 mV from 0.85V to 1.05V. The performance is summarised in Table 8.3.

The transconductance amplifier acts as the input stage for the whole system, making its noise performance crucial. The input-referred noise of the OTA is 9.56 μV_{rms}, mainly given by the input differential and the load pairs of the OTA, which occupies the most area of ΔΣADC as well. The transistor dimensions and noise contribution are given in Table 8.4.

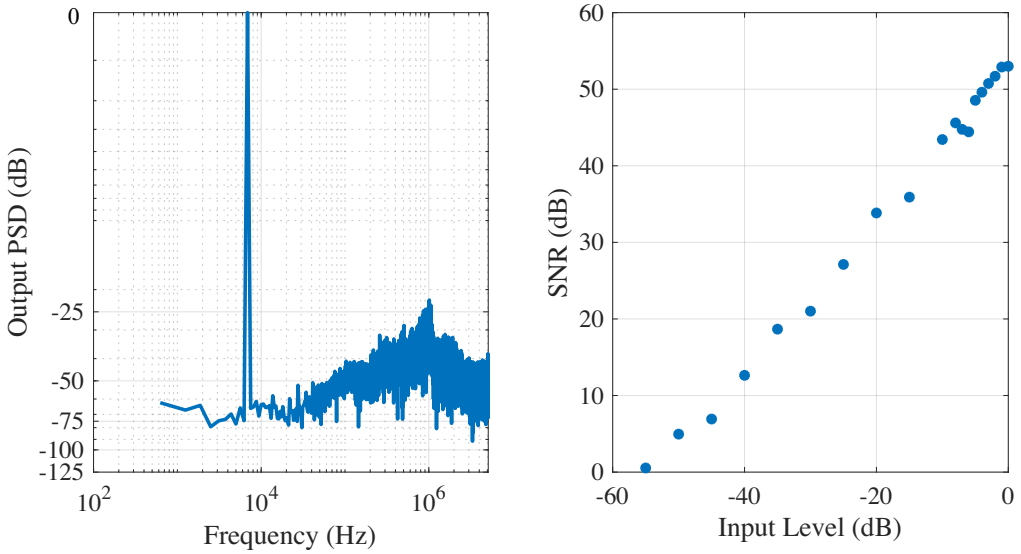


FIGURE 8.4: Performance of ΔΣADC. (a) the output spectrum for the input sine wave of 10 mV_{pp} under the noise simulation. (b) SNR versus input level where indicates 55dB DR.

TABLE 8.3: Performance of the proposed ΔΣADC under the condition of transient noise.

Parameter	**Value**
F_s	10.24 MHz
R_{out} @F_s	3.10 $M\Omega$
SINAD	53.00 dB
ENOB	8.51 bits
SFDR	60.22 dBc
Input-referred noise of OTA	9.56 μV_{rms}
DC Input Range	0.85-1.05 V
AC Input Range	10 mV
Power	5.50 μW

TABLE 8.4: The critical transistor dimensions and noise contribution of the OTA.

Transistor	W/L ($\mu m/\mu m$)	Noise Contribution	
M7&M8	18/1	50.2%	4.86 μV_{rms}
M1&M2	8.5/6	48.78%	4.66 μV_{rms}

8.3 Simulation with Spike Train

In this section, we investigate the performance of ΔΣ ADC on a spike train with 15.77mV DC offset by conducting a transient simulation with a spike train. The timing diagram presents in Figure 8.5, where (a) shows the spike train in *inp* and *inn* is set to 0.9V. The OTA is operated as a conventional amplifier in this design without switching status. As a result, the input signal is free of kickback noise (compared with the inputs in both AC-coupled and DC-coupled front-end). In this simulation, we set $R1 = 1k\Omega$ and $R2 = 8.5k\Omega$

to compensate $15.77mV$ DC offset. In addition, we get a spectrum from *outn* presented in Figure 8.5. In this simulation, the input frequency is $10.07kHz$ and its amplitude is $500\mu V$ at its maximum. ENOB = 1.34 indicates the resolution is $198\mu V$, significantly lower than the simulation of pure sine waves (ENOB = 5.33). The DC offset ($15.77mV$) needs to be compensated by high precision memristors of $0.01mV$. Nevertheless, the memristors in this circuit are placed at the source of input transistors to achieve a wider tuneable offset range, albeit at the expense of precision. Due to the DC input, the SINAD and ENOB may experience attenuation. Thus, improving the ΔΣ ADC in terms of eliminating common mode input effect and practical detection remains a future work.

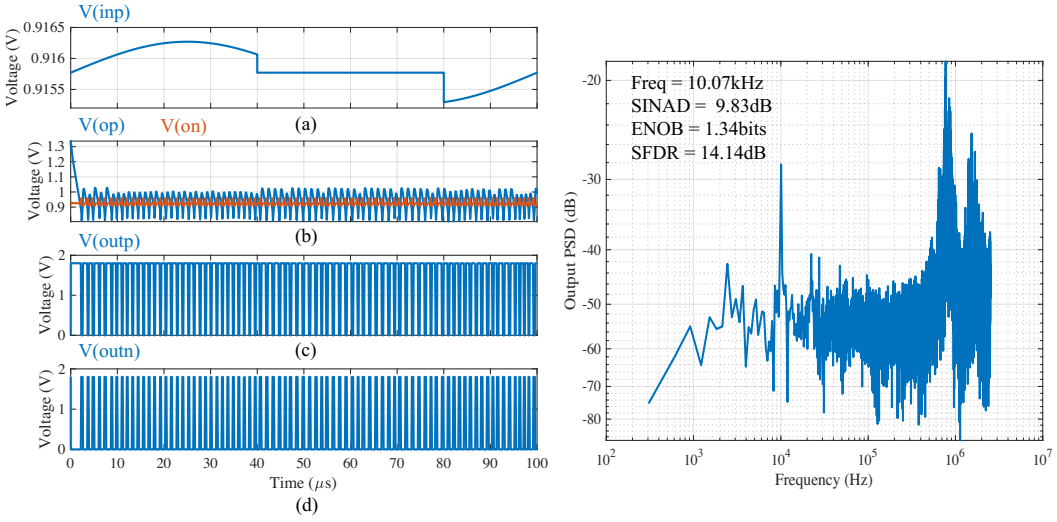


FIGURE 8.5: Left: the timing diagram of ΔΣADC with (a) the spike train input with $15.77mV$ DC offset. (b) the output of the OTA. (c) and (d) are the outputs of the DLC, respectively. Right: output spectrum for spike train with the maximum amplitude of $500\mu V$ and $15.77mV$ DC offset.

8.4 Summary

This chapter proposes the DC-coupled continuous-time ΔΣADC based on our previous designs. Even though the design of Gm-C based CT-ΔΣADC meets the challenge of relative low linearity and SNDR, this limitation can be lifted since it is applied for low input range bio-signal ($10mV$). Meanwhile, memristive devices can compensate for the large DC offset at the input. Under transient noise in the range of [$1Hz$,$100MHz$], it operates as an eight-bit ADC at the front-end. The OTA occupied the most significant proportion of the total power consumption, contributing to a low input-referred noise level achieved by the modulator's loop filter. Compared with the front-end in Chapter 4 - Chapter 6, the OTA operates in a static mode that consumes more power than the dynamic mode, which uses the start-stop scheme to reduce the power down to microvolt levels. In summary, the memristor-based amplifier can be applied to the dynamic pre-amplifier and the static one in ΔΣ ADC. This contributes to rejecting the DC offset with low power consumption and area efficiency.

Chapter 9

Conclusion and Future Work

9.1 Conclusion

This dissertation performs a theoretical analysis of the core functionality of memristive integrating amplifiers. It uses industrial CAD-level simulations to provide a specific example of an integrating amplifier design targeting electrophysiological applications. Throughout our analysis, the performance enhancement over traditional, continuous-mode amplifiers could be most intuitively understood as a gain-boosting effect arising from the integration process. Moreover, we explained how standard amplifier performance metrics, such as gain and input common mode range, and new metrics, such as offset voltage tuneability range, can be described by governing equations for use by designers. Finally, we implemented an exemplar design in commercially available $180nm$ CMOS and demonstrated typical values for all studied performance parameters that can be expected from a $0.18\mu m$ node technology.

The memristive device was utilised as a tunable resistor along the differential current path to compensate for the offset voltage. We also studied the impact of memristor IV non-linearity on the system, especially for the integration process. We checked whether transitioning from ideal resistor-based analysis to realistic memristor-based presents a fundamental conceptual challenge towards the viability of low-power, ultra-low offset memristive amplifiers. The result was that it still needs to remove the last fundamental/conceptual roadblock we have identified towards scale-up and, ultimately, commercial exploitation; the unique challenges are of an engineering nature.

This work is a stepping stone towards de-risking and documenting the memristor-based integrating amplifier. With a nano-scaled memristor implanted, the circuit does not need a large module to eliminate the offset voltage after fabrication. However, the memristive device still has some limitations. The in-house memristor is connected to a switch unit which needs $8V$ to $12V$ to switch the resistance state. In this case, the switch unit induces noise risk to this sensitive system and increases power consumption. The trade-off induced

by the integration process, combined with the offset trimming enabled by the memristor, has the potential to add a robust circuit topology to the arsenal of the analogue designer.

After defining and measuring the key performance metric for integrating amplifier, we applied the memristor-based pre-amplifier into three types of application circuits: (i) open-loop network with input capacitors and MOS-bipolar pseudo-resistor as an AC-coupled solution to reject DC offset, (ii) directly operates in the DC-coupled solution for DC offset up to $50mV$, and (iii) memristor-based $\Delta\Sigma$ ADC. Through measurement, the AC-coupled and DC-coupled memristor-based pre-amplifier with a 'start-stop' scheme can achieve above 95% detection accuracy while maintaining the power consumption at hundreds of nanowatt per channel. Moreover, the $\Delta\Sigma$ ADC with static pre-amplifier operates at $10.24MHz$, consuming $5.5\mu W$.

In the tape-out of single channel pre-amplifier, we used in-house memristor electrodes/-pads that directly connect to the memristor. It is because electroforming the device requires $8-12V$ that exceeds the maximum tolerance voltage of the chosen commercial pad. However, the latest version of the hybrid CMOS/memristor chip encounters the situation that the memristor wiring shorts the power path that cannot be tested. Under this situation, the chip test can only be conducted on the chip with missing memristors and memristive pads, where an open circuit exists inside the core amplifier. We can only access the DLC module with floating inputs. The test results indicate that the chip can be powered up, but the DLC operates with long delays, which might be due to the long wiring (i) inside the chip, and (ii) jumper wires between the probe card to the test board. We can still gain some insights from this chip test regarding the test strategy and layout design concluded in Chapter 7. Furthermore, the improvement of the circuit design based on chip tests remains future work.

9.2 Recommendation for Future Work

The research work in this thesis brings new application prospects for the memristive device, utilising it as the trimming component to reject the DC offset at the input in the application of neural signal detection. The presented work can be developed in the following ways.

1. Development of the memristive Verilog-A model. Currently, the memristive Verilog-A model in Cadence can only trace the RS in transient simulation. Moreover, it operates as the ideal static resistor in DC/AC simulation (without noise), which restricts us from exploring the frequency impact on the memristor. Therefore, developing the AC and noise model for the memristor is recommended, which contributes to higher accuracy of the performance of the memristor in CMOS circuits. In addition, the neural front-end should be operated with ultra-low power consumption, especially for the electroforming and the programming of the memristive devices. In this project, we utilised the device

that requires $8 - 12V$ for electroforming/activation. It induces the risk of destroying the circuit in $0.18\mu m$ even though we utilised a memristor electrode and separated it from other signal pads. For future neural front-end development, developing the Verilog-A model for low-voltage activation is recommended, which contributes to power efficiency.

2. Develop the automated programming interface of the memristor. Currently, we utilise the memristor for trimming the offset voltage by monitoring the digital output and manually programming using an external device (ArCOne Instrument). It is recommended to develop an automated programming interface that incorporates a detection/comparison circuit at the output of the OTA to control the programming module until the DC offset is completely covered. Note that this can be realised when the memristor can be electroformed and programmed with a low voltage. In this case, the circuit can be calibrated at the first stage, followed by the actual operation.

3. Full system development. Current research focuses on the design of a single-channel pre-amplifier. The design of multiple channels and implementation of the full system remain future work. The focus should be placed on parasitic analysis in the multi-channel design. Moreover, a power module is required to generate stable bias current to different pre-amplifier channels. In addition, the full system requires a controlling module that generates clocking signals to control the operation of the pre-amplifier with the consideration of clock skew and jitter.

Bibliography

[1] J. Kao, S. Stavisky, D. Sussillo, P. Nuyujukian, and K. Shenoy, "Information systems opportunities in brain–machine interface decoders," *Proceedings of the IEEE*, vol. 102, pp. 666–682, 2014.

[2] 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.

[3] S. Nag and N. V. Thakor, "Implantable neurotechnologies: electrical stimulation and applications," *Medical & biological engineering & computing*, vol. 54, no. 1, pp. 63–76, 2016.

[4] D. Loi, C. Carboni, G. Angius, G. N. Angotzi, M. Barbaro, L. Raffo, S. Raspopovic, and X. Navarro, "Peripheral neural activity recording and stimulation system," *IEEE transactions on biomedical circuits and systems*, vol. 5, no. 4, pp. 368–379, 2011.

[5] D. Tsai, D. Sawyer, A. Bradd, R. Yuste, and K. L. Shepard, "A very large-scale microelectrode array for cellular-resolution electrophysiology," *Nature communications*, vol. 8, no. 1, pp. 1–11, 2017.

[6] X. Yuan, M. Schröter, M. E. J. Obien, M. Fiscella, W. Gong, T. Kikuchi, A. Odawara, S. Noji, I. Suzuki, J. Takahashi *et al.*, "Versatile live-cell activity analysis platform for characterization of neuronal dynamics at single-cell and network level," *Nature communications*, vol. 11, no. 1, pp. 1–14, 2020.

[7] S. Wang, S. K. Garakoui, H. Chun, D. G. Salinas, W. Sijbers, J. Putzeys, E. Martens, J. Craninckx, N. Van Helleputte, and C. M. Lopez, "A compact quad-shank cmos neural probe with 5,120 addressable recording sites and 384 fully differential parallel channels," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 6, pp. 1625–1634, 2019.

[8] R. Fiáth, B. C. Raducanu, S. Musa, A. Andrei, C. M. Lopez, M. Welkenhuysen, P. Ruther, A. Aarts, and I. Ulbert, "Fine-scale mapping of cortical laminar activity during sleep slow oscillations using high-density linear silicon probes," *Journal of neuroscience methods*, vol. 316, pp. 58–70, 2019.

[9] G. N. Angotzi, F. Boi, A. Lecomte, E. Miele, M. Malerba, S. Zucca, A. Casile, and L. Berdondini, "Sinaps: An implantable active pixel sensor cmos-probe for simultaneous large-scale neural recordings," *Biosensors and Bioelectronics*, vol. 126, pp. 355–364, 2019.

[10] E. Musk and Neuralink, "An integrated brain-machine interface platform with thousands of channels (submitted)," 2019.

[11] X. Zou, L. Liu, J. H. Cheong, L. Yao, P. Li, M.-Y. Cheng, W. L. Goh, R. Rajkumar, G. S. Dawe, K.-W. Cheng *et al.*, "A 100-channel 1-mw implantable neural recording ic," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 10, pp. 2584–2596, 2013.

[12] 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.

[13] 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.

[14] H. Gao, R. M. Walker, P. Nuyujukian, K. A. Makinwa, K. V. Shenoy, B. Murmann, and T. H. Meng, "Hermese: A 96-channel full data rate direct neural interface in $0.13\mu m$ cmos," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 4, pp. 1043–1055, 2012.

[15] M. S. Chae, Z. Yang, M. R. Yuce, L. Hoang, and W. Liu, "A 128-channel 6 mw wireless neural recording ic with spike feature extraction and uwb transmitter," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 17, no. 4, pp. 312–321, 2009.

[16] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, 2006.

[17] A. D. Rush and P. R. Troyk, "A power and data link for a wireless-implanted neural recording system," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 11, pp. 3255–3262, 2012.

[18] R. Sharpeshkar, "Ultra low power bioelectronics: Fundamentals, biomedical applications, and bio-inspired system," 2010.

[19] I. Karageorgos, K. Sriram, J. Veselỳ, M. Wu, M. Powell, D. Borton, R. Manohar, and A. Bhattacharjee, "Hardware-software co-design for brain-computer interfaces," in *2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 2020, pp. 391–404.

[20] N. Even-Chen, D. G. Muratore, S. D. Stavisky, L. R. Hochberg, J. M. Henderson, B. Murmann, and K. V. Shenoy, "Power-saving design opportunities for wireless intracortical brain–computer interfaces," *Nature biomedical engineering*, vol. 4, no. 10, pp. 984–996, 2020.

[21] J. T. Robinson, E. Pohlmeyer, M. C. Gather, C. Kemere, J. E. Kitching, G. G. Malliaras, A. Marblestone, K. L. Shepard, T. Stieglitz, and C. Xie, "Developing next-generation brain sensing technologies—a review," *IEEE sensors journal*, vol. 19, no. 22, pp. 10 163–10 175, 2019.

[22] I. Gupta, A. Serb, A. Khiat, R. Zeitler, S. Vassanelli, and T. Prodromakis, "Real-time encoding and compression of neuronal spikes by metal-oxide memristors," *Nature communications*, vol. 7, no. 1, pp. 1–9, 2016.

[23] L. Fassio, L. Lin, R. De Rose, M. Lanuzza, F. Crupi, and M. Alioto, "Trimming-less voltage reference for highly uncertain harvesting down to 0.25 v, 5.4 pw," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 10, pp. 3134–3144, 2021.

[24] S. Lv, J. Liu, and Z. Geng, "Application of memristors in hardware security: A current state-of-the-art technology," *Advanced Intelligent Systems*, vol. 3, no. 1, p. 2000127, 2021.

[25] Q. Xia, W. Robinett, M. W. Cumbie, N. Banerjee, T. J. Cardinali, J. J. Yang, W. Wu, X. Li, W. M. Tong, D. B. Strukov *et al.*, "Memristor- cmos hybrid integrated circuits for reconfigurable logic," *Nano letters*, vol. 9, no. 10, pp. 3640–3645, 2009.

[26] F. Cai, J. M. Correll, S. H. Lee, Y. Lim, V. Bothra, Z. Zhang, M. P. Flynn, and W. D. Lu, "A fully integrated reprogrammable memristor–cmos system for efficient multiply–accumulate operations," *Nature Electronics*, vol. 2, no. 7, pp. 290–299, 2019.

[27] S. Stathopoulos, A. Khiat, M. Trapatseli, S. Cortese, A. Serb, I. Valov, and T. Prodromakis, "Multibit memory operation of metal-oxide bi-layer memristors," *Scientific reports*, vol. 7, no. 1, pp. 1–7, 2017.

[28] P. Mazumder, S.-M. Kang, and R. Waser, "Memristors: devices, models, and applications," *Proceedings of the IEEE*, vol. 100, no. 6, pp. 1911–1919, 2012.

[29] K.-H. Kim, S. Gaba, D. Wheeler, J. M. Cruz-Albrecht, T. Hussain, N. Srinivasa, and W. Lu, "A functional hybrid memristor crossbar-array/cmos system for data storage and neuromorphic applications," *Nano letters*, vol. 12, no. 1, pp. 389–395, 2012.

[30] A. Serb and T. Prodromakis, "High-sensitivity memristor-based threshold detection," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–4.

[31] J. Zha, H. Huang, T. Huang, J. Cao, A. Alsaedi, and F. E. Alsaadi, "A general memristor model and its applications in programmable analog circuits," *Neurocomputing*, vol. 267, pp. 134–140, 2017.

[32] Q. Xia, "Memristive nanodevices: Cmos compatibility and novel applications," in *2016 18th Mediterranean Electrotechnical Conference (MELECON)*. IEEE, 2016, pp. 1–4.

[33] I. Bhati, M.-T. Chang, Z. Chishti, S.-L. Lu, and B. Jacob, "Dram refresh mechanisms, penalties, and trade-offs," *IEEE Transactions on Computers*, vol. 65, no. 1, pp. 108–121, 2015.

[34] K. R. GENDEPUJARI, "Cantilever based nonvolatile memories: A survey," Ph.D. dissertation, SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY, SURAT, 2020.

[35] S. Maheshwari, S. Stathopoulos, J. Wang, A. Serb, Y. Pan, A. Mifsud, L. B. Leene, J. Shen, C. Papavassiliou, T. G. Constandinou *et al.*, "Design flow for hybrid cmos/memristor systems—part i: Modeling and verification steps," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 12, pp. 4862–4875, 2021.

[36] I. Messaris, S. Nikolaidis, A. Serb, S. Stathopoulos, I. Gupta, A. Khiat, and T. Prodromakis, "A tio2 reram parameter extraction method," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2017, pp. 1–4.

[37] I. Messaris, A. Serb, S. Stathopoulos, A. Khiat, S. Nikolaidis, and T. Prodromakis, "A data-driven verilog-a reram model," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 12, pp. 3151–3162, 2018.

[38] J. Wang, A. Serb, C. Papavassiliou, S. Maheshwari, and T. Prodromakis, "Analysing and measuring the performance ofmemristive integrating amplifiers," *arXiv preprint arXiv:2009.09909*, 2020.

[39] J. Wang, A. Serb, S. Wang, and T. Prodromakis, "Hybrid cmos/memristor front-end for multiunit activity processing," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2022, pp. 1–5.

[40] J. Wang, A. Serb, C. Papavassiliou, and T. Prodromakis, "Accounting for memristor iv non-linearity in low power memristive amplifiers," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2021, pp. 1–5.

[41] J. Wang, A. Serb, S. Wang, and T. Prodromakis, "Offset rejection in a dc-coupled hybrid cmos/memristor neural front-end," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2022, pp. 1–5.

[42] B. P. Bean, "The action potential in mammalian central neurons," *Nature Reviews Neuroscience*, vol. 8, no. 6, pp. 451–465, 2007.

[43] M. Dipalo, H. Amin, L. Lovato, F. Moia, V. Caprettini, G. C. Messina, F. Tantussi, L. Berdondini, and F. De Angelis, "Intracellular and extracellular recording of spontaneous action potentials in mammalian neurons and cardiac cells with 3d plasmonic nanoelectrodes," *Nano letters*, vol. 17, no. 6, pp. 3932–3939, 2017.

[44] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, A. J. Hudspeth, S. Mack *et al.*, *Principles of neural science*. McGraw-hill New York, 2000, vol. 4.

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

[46] G. Dornhege, J. d. R. Millán, T. Hinterberger, D. J. McFarland, K.-r. Muller *et al.*, *Toward brain-computer interfacing*. Citeseer, 2007, vol. 63.

[47] A. Molleman, *Patch clamping: an introductory guide to patch clamp electrophysiology*. John Wiley & Sons, 2003.

[48] J. G. Webster, *Medical instrumentation: application and design*. John Wiley & Sons, 2009.

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

[50] Y.-S. Choi, M. A. Koenig, X. Jia, and N. V. Thakor, "Quantifying time-varying multiunit neural activity using entropy-based measures," *IEEE transactions on biomedical engineering*, vol. 57, no. 11, pp. 2771–2777, 2010.

[51] E. Stark and M. Abeles, "Predicting movement from multiunit activity," *Journal of Neuroscience*, vol. 27, no. 31, pp. 8387–8394, 2007.

[52] I. Stolerman, *Encyclopedia of psychopharmacology*. Springer Science & Business Media, 2010.

[53] G. Hong and C. M. Lieber, "Novel electrode technologies for neural recordings," *Nature Reviews Neuroscience*, vol. 20, no. 6, pp. 330–345, 2019.

[54] T. Torfs, A. A. Aarts, M. A. Erismis, J. Aslam, R. F. Yazicioglu, K. Seidl, S. Herwik, I. Ulbert, B. Dombovari, R. Fiath *et al.*, "Two-dimensional multi-channel neural probes with electronic depth control," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, no. 5, pp. 403–412, 2011.

[55] A. C. Paulk, Y. Kfir, A. R. Khanna, M. L. Mustroph, E. M. Trautmann, D. J. Soper, S. D. Stavisky, M. Welkenhuysen, B. Dutta, K. V. Shenoy *et al.*, "Large-scale neural recordings with single neuron resolution using neuropixels probes in human cortex," *Nature Neuroscience*, vol. 25, no. 2, pp. 252–263, 2022.

[56] N. A. Steinmetz, C. Aydin, A. Lebedeva, M. Okun, M. Pachitariu, M. Bauza, M. Beau, J. Bhagat, C. Böhm, M. Broux *et al.*, "Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings," *Science*, vol. 372, no. 6539, p. eabf4588, 2021.

[57] S. Wurth, M. Capogrosso, S. Raspopovic, J. Gandar, G. Federici, N. Kinany, A. Cutrone, A. Piersigilli, N. Pavlova, R. Guiet *et al.*, "Long-term usability and bio-integration of polyimide-based intra-neural stimulating electrodes," *Biomaterials*, vol. 122, pp. 114–129, 2017.

[58] G. Hong, X. Yang, T. Zhou, and C. M. Lieber, "Mesh electronics: a new paradigm for tissue-like brain probes," *Current opinion in neurobiology*, vol. 50, pp. 33–41, 2018.

[59] X. Yang, T. Zhou, T. J. Zwang, G. Hong, Y. Zhao, R. D. Viveros, T.-M. Fu, T. Gao, and C. M. Lieber, "Bioinspired neuron-like electronics," *Nature materials*, vol. 18, no. 5, pp. 510–517, 2019.

[60] T.-M. Fu, G. Hong, R. D. Viveros, T. Zhou, and C. M. Lieber, "Highly scalable multichannel mesh electronics for stable chronic brain electrophysiology," *Proceedings of the National Academy of Sciences*, vol. 114, no. 47, pp. E10 046–E10 055, 2017.

[61] J. K. Krauss, N. Lipsman, T. Aziz, A. Boutet, P. Brown, J. W. Chang, B. Davidson, W. M. Grill, M. I. Hariz, A. Horn *et al.*, "Technology of deep brain stimulation: current status and future directions," *Nature Reviews Neurology*, vol. 17, no. 2, pp. 75–87, 2021.

[62] H. Shin, Y. Son, U. Chae, J. Kim, N. Choi, H. J. Lee, J. Woo, Y. Cho, S. H. Yang, C. J. Lee *et al.*, "Multifunctional multi-shank neural probe for investigating and modulating long-range neural circuits in vivo," *Nature communications*, vol. 10, no. 1, pp. 1–11, 2019.

[63] M. Zhang, Z. Tang, X. Liu, and J. Van der Spiegel, "Electronic neural interfaces," *Nature Electronics*, vol. 3, no. 4, pp. 191–200, 2020.

[64] 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.

[65] D. Y. Barsakcioglu, Y. Liu, P. Bhunjun, J. Navajas, A. Eftekhar, A. Jackson, R. Q. Quiroga, and T. G. Constandinou, "An analogue front-end model for developing neural spike sorting systems," *IEEE transactions on biomedical circuits and systems*, vol. 8, no. 2, pp. 216–227, 2014.

[66] H. G. Rey, C. Pedreira, and R. Q. Quiroga, "Past, present and future of spike sorting techniques," *Brain research bulletin*, vol. 119, pp. 106–117, 2015.

[67] C. D. Ferris, *Introduction to bioinstrumentation.* Humana Press, 1978.

[68] D. De Dorigo, C. Moranz, H. Graf, M. Marx, D. Wendler, B. Shui, A. S. Herbawi, M. Kuhl, P. Ruther, O. Paul *et al.*, "Fully immersible subcortical neural probes with modular architecture and a delta-sigma adc integrated under each electrode for parallel readout of 144 recording sites," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3111–3125, 2018.

[69] D.-Y. Yoon, S. Pinto, S. Chung, P. Merolla, T.-W. Koh, and D. Seo, "A 1024-channel simultaneous recording neural soc with stimulation and real-time spike detection," in *2021 Symposium on VLSI Circuits.* IEEE, 2021, pp. 1–2.

[70] X. T. Pham, N. T. Nguyen, V. T. Nguyen, and J.-W. Lee, "A 0.6-μw chopper amplifier using a noise-efficient dc servo loop and squeezed-inverter stage for power-efficient biopotential sensing," *Sensors*, vol. 20, no. 7, p. 2059, 2020.

[71] 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.

[72] N. Laskar, K. Guha, S. Nath, S. Chanda, K. Baishnab, P. Paul, and K. Rao, "Design of high gain, high bandwidth neural amplifier ic considering noise-power trade-off," *Microsystem technologies*, vol. 27, no. 2, pp. 585–599, 2021.

[73] H. S. Bindra, C. E. Lokin, D. Schinkel, A.-J. Annema, and B. Nauta, "A 1.2-v dynamic bias latch-type comparator in 65-nm cmos with 0.4-mv input noise," *IEEE journal of solid-state circuits*, vol. 53, no. 7, pp. 1902–1912, 2018.

[74] R. Wu, J. H. Huijsing, and K. A. Makinwa, *Precision instrumentation amplifiers and read-out integrated circuits.* Springer Science & Business Media, 2012.

[75] 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 (ISCAS).* IEEE, 2012, pp. 1075–1078.

[76] A. Samiei and H. Hashemi, "A chopper stabilized, current feedback, neural recording amplifier," *IEEE solid-state circuits letters*, vol. 2, no. 3, pp. 17–20, 2019.

[77] J. Wu, M.-K. Law, P.-I. Mak, and R. P. Martins, "A $2\mu w$ $45nv/\sqrt{Hz}$ readout front end with multiple-chopping active-high-pass ripple reduction loop and pseudofeedback dc servo loop," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 4, pp. 351–355, 2015.

[78] A. Nikas, S. Jambunathan, L. Klein, M. Voelker, and M. Ortmanns, "A continuous-time delta-sigma modulator using a modified instrumentation amplifier and current reuse dac for neural recording," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 10, pp. 2879–2891, 2019.

[79] S. Wang, M. Ballini, X. Yang, C. Sawigun, J.-W. Weijers, D. Biswas, N. Van Helleputte, and C. M. Lopez, "A compact chopper stabilized δ-δ σ neural readout ic with input impedance boosting," *IEEE Open Journal of the Solid-State Circuits Society*, vol. 1, pp. 67–78, 2021.

[80] R. C. Nongpiur, D. J. Shpak, and A. Antoniou, "Improved design method for nearly linear-phase iir filters using constrained optimization," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 895–906, 2012.

[81] C. Gold, D. A. Henze, C. Koch, and G. Buzsaki, "On the origin of the extracellular action potential waveform: a modeling study," *Journal of neurophysiology*, vol. 95, no. 5, pp. 3113–3128, 2006.

[82] R. Quian Quiroga and S. Panzeri, "Extracting information from neuronal populations: information theory and decoding approaches," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 173–185, 2009.

[83] 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, p. R53, 1998.

[84] E. M. Schmidt, "Computer separation of multi-unit neuroelectric data: a review," *Journal of neuroscience methods*, vol. 12, no. 2, pp. 95–111, 1984.

[85] F. Kloosterman, S. P. Layton, Z. Chen, and M. A. Wilson, "Bayesian decoding using unsorted spikes in the rat hippocampus," *Journal of neurophysiology*, 2014.

[86] 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.

[87] D. Wendler, D. De Dorigo, M. Amayreh, A. Bleitner, M. Marx, R. Willaredt, and Y. Manoli, "A 0.0046-mm^2 two-step incremental delta–sigma analog-to-digital converter neuronal recording front end with 120-mvpp offset compensation," *IEEE Journal of Solid-State Circuits*, 2022.

[88] C. Lee, T. Jeon, M. Jang, S. Park, J. Kim, J. Lim, J.-H. Ahn, Y. Huh, and Y. Chae, "A 6.5-μw 10-khz bw 80.4-db sndr g m-c-based ct modulator with a feedback-assisted g m linearization for artifact-tolerant neural recording," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 11, pp. 2889–2901, 2020.

[89] 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.

[90] G. Foffani, M. Morales-Botello, and J. Aguilar, "Spike timing, spike count, and temporal information for the discrimination of tactile stimuli in the rat ventrobasal complex," *Journal of Neuroscience*, vol. 29, no. 18, pp. 5964–5973, 2009.

[91] R. Rieger, J. Taylor, A. Demosthenous, N. Donaldson, and P. J. Langlois, "Design of a low-noise preamplifier for nerve cuff electrode recording," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 8, pp. 1373–1379, 2003.

[92] S. M. Sharroush, Y. S. Abdalla, A. A. Dessouki, and E. A. El-Badawy, "Subthreshold mosfet transistor amplifier operation," in *2009 4th International Design and Test Workshop (IDT)*, 2009, pp. 1–6.

[93] K. G. Lamb, S. J. Sanchez, and W. T. Holman, "A low noise operational amplifier design using subthreshold operation," in *Proceedings of 40th Midwest Symposium on Circuits and Systems. Dedicated to the Memory of Professor Mac Van Valkenburg*, vol. 1. IEEE, 1997, pp. 35–38.

[94] R. Harrison, "A low-power integrated circuit for adaptive detection of action potentials in noisy signals," *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, vol. 4, p. 3325–3328, 2003.

[95] K. Kundert, "Simulating switched-capacitor filters with spectrerf," *The Designer's Guide Community*, pp. 1–25, 2006.

[96] T. Sepke, P. Holloway, C. G. Sodini, and H.-S. Lee, "Noise analysis for comparator-based circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 3, pp. 541–553, 2008.

[97] 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.

[98] A. Serb, J. Bill, A. Khiat, R. Berdan, R. Legenstein, and T. Prodromakis, "Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses," *Nature communications*, vol. 7, no. 1, p. 12611, 2016.